Trabajo de integración Curricular previa a la obtención de título de Master en Ingeniería Automotriz con mención en procesos y calidad de servicio automotriz.

AUTORES:

Cajas López Adrián Santiago

Meneses Chalcualan Ruth Irene

TUTOR:

Cristopher Israel Fuertes Iturralde

Análisis del proceso de servicio mecánico en el flujo de alistamiento de vehículos seminuevos

Quito, julio 2025

CERTIFICACION

Nosotros, Adrian Santiago Cajas López y Ruth Irene Meneses Chalcualan, declaramos bajo juramento, que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, reglamento y leyes.

Ing. Adrian Santiago Cajas López

CI: 1717095747

Ing. Ruth Irene Meneses Chalcualan

CI: 1725516734

CERTIFICACION DEL DIRECTOR

Yo, Msc. Cristopher Isarel Fuertes Iturralde, certifico que conozco a los autores del presente trabajo. Siendo los responsables exclusivos tanto de su originalidad y autenticidad, como de su contenido.

Firmado electrónicamente por:
CRISTOPHER ISRAEL
FUERTES ITURRALDE
Validar únicamente con FirmaEC

Firma del director de Trabajo de Grado

Msc. Cristopher Isarel Fuertes Iturralde

DEDICATORIA

Dedico este proyecto a mi familia y mis hermanos que me han acompañado durante todo esto desafío apoyándome en cada momento y decisión que he tomado, por sus enseñanzas de nunca darme por vencido y seguir siempre hacia adelante.

A Ruth por el gran esfuerzo y compromiso que demostramos para afrontar cada una de los obstáculos que se presentaron y el apoyo moral que siempre se demostro.

Adrián Santiago Cajas López

AGRADECIMIENTOS

Agradezco a Dios por haber dado la fuerza y la sabiduría para poder seguir en cada una de los desafíos que se presentaron y poder sobrellevarlos de la manera más correcta.

A mi familia, por su paciencia, comprensión y por saber aceptar mi ausencia en momentos importantes.

A los profesores presentes durante este periodo que han compartido sus conocimientos y experiencia.

Adrián Santiago Cajas López

DEDICATORIA

A Dios, a pesar de mis errores me acepta tal como soy.

A mis padres, por su amor incondicional, por enseñarme con su ejemplo que el esfuerzo y la humildad abren cualquier puerta, y por ser siempre mi mayor apoyo, incluso en silencio.

A mis amigos, por acompañarme con fe, paciencia y palabras que fortalecieron mi espíritu cuando más lo necesitaba.

Ruth Irene Meneses Chalcualan

AGRADECIMIENTOS

Agradezco a Dios por haber permitido cumplir un objetivo muy importante en mi vida como es culminar esta maestría.

A mis padres quienes han apoyado en cada momento y han sabido comprender cada pasa que he dado.

Agradezco a mis compañeros de trabajo por la ayuda prestada en este proceso.

Ruth Irene Meneses Chalcualan

INDICE DE CONTENIDO

INDICE DE TABLAS

INDICE DE FIGURAS

ANÁLISIS DEL PROCESO DE SERVICIO MECANICO EN EL FLUJO DE ALISTAMIENTO DE VEHÍCULOS SEMINUEVOS

Ing. Adrian Cajas[1], Ing. Ruth Meneses[2]

Ingeniero Automotriz – Universidad de las Fuerzas Armadas ESPE,
adcajaslo@uide.com.ec, Latacunga - Ecuador
Ingeniero Automotriz – Universidad de las Fuerzas Armadas ESPE,
ruchalcualanch@uide.edu.ec, Latacunga – Ecuador

RESUMEN

Introducción: El proceso de alistamiento en talleres de vehículos seminuevos presenta desafíos operativos significativos, se propone un diagnóstico mediante el uso de métodos para el análisis de procesos, con el objetivo de identificar cuellos de botella y proponer mejoras para la optimización de recursos. **Metodología:** Se empleó un enfoque combinado utilizando técnicas cualitativas y cuantitativas por medio de la observación directa y toma de tiempos, se diagnosticaron fallas operativas, con la ayuda de la aplicación de los indicadores de gestión como el TAT, Ishikawa y mapa de cadena de valor. **Resultados:** Las principales causas de la ineficiencia son la falta de comunicación y demoras en repuestos, de igual manera con el VSM se evidencio que el 57% del tiempo se pierde en la solicitud de repuestos y que existe un alto desperdicio por espera y defectos, finalmente solo el 5% de vehículos fue entregado dentro del tiempo objetivo. **Conclusión:** Con los datos obtenidos se evidencia una baja eficiencia operativa y la necesidad urgente de mejoras en los procesos, brindando mejor capacitación al personal y evaluando la carga laboral que cada uno realiza.

Palabras clave: Vehículos seminuevos, VSM, Ishikawa, KPI TAT

ABSTRAC

Introduction: The pre-owned vehicle setup process in workshops presents significant operational challenges. A diagnostic approach is proposed using process analysis methods to identify bottlenecks and propose improvements for resource optimization. **Methodology**: A combined approach using qualitative and quantitative techniques was employed through direct observation and time recording. Operational failures were diagnosed with the help of management indicators such as TAT, Ishikawa, and value chain mapping. **Results**: The main causes of inefficiency are lack of communication and delays in spare parts. Similarly, the VSM showed that 57% of time is lost requesting spare parts and that there is high waste due to waiting and defects. Ultimately, only 5% of vehicles were delivered within the target time. **Conclusion**: The data obtained demonstrates low operational efficiency and an urgent need for process improvements, providing better staff training and evaluating the workload each performs.

Keywords: Used vehicles, VSM, Ishikawa, KPI TAT

1. INTRODUCCIÓN

En los patios de compra y venta de vehículos seminuevos, la eficiencia en el proceso de alistamiento mecánico se ha convertido en un factor determinante para cumplir con los tiempos de entrega y mantener altos niveles de satisfacción del cliente. No obstante, en muchos de estos entornos, especialmente en talleres de menor escala, la gestión operativa se desarrolla de forma empírica, sustentada principalmente en la experiencia del personal, sin una estructura metodológica definida. Como lo señala (Bernal, 2012), esta administración basada en la práctica y no en el análisis técnico puede ocasionar múltiples problemas operativos, como retrasos, reprocesos y costos adicionales innecesarios.

Frente a esta problemática, es necesario realizar un estudio detallado del proceso actual que permite identificar las causas que originan ineficiencias y proponer soluciones concretas. Según (Galo & Pascua, 2023), al analizar los puntos del proceso donde se concentra el mayor impacto negativo, da paso a establecer acciones orientadas a transformar la gestión y a mejorar los resultados operativos. Este tipo de intervención permite una administración más efectiva, basada en datos y evidencias, en lugar de suposiciones o hábitos adquiridos.

El análisis de procesos es una herramienta esencial para comprender cómo interactúan las distintas actividades dentro de un sistema de trabajo. De acuerdo con (Cheung & Bal, 1998), su aplicación identifica las tareas innecesarias, retrasos frecuentes y fallas en la secuencia de actividades. Para ello se debe identificar las razones de las variaciones entre los tiempos objetivos y los tiempos reales; así como establecer o el tiempo promedio en que un vehículo es atendido en el taller. El abordar estos elementos, favorece a un funcionamiento más fluido del proceso, fortalece los mecanismos de control y mejora la asignación de recursos, lo que contribuye a elevar los estándares del servicio entregado.

A partir de una metodología basada en la observación directa y el análisis de datos históricos, se emplean herramientas de análisis de procesos tales como el Mapa de Flujo de Valor (VSM), el indicador TAT (Turnaround Time) y técnicas de diagnóstico estructurado como el Diagrama de Ishikawa, que permiten identificar variables críticas que generan desperdicio dentro del proceso. Asimismo, el empleo de entrevistas semiestructuradas al personal técnico, favorecen a la comprensión integral de los factores que influyen en el desempeño operativo y en la gestión de recursos.

Con un enfoque que integra análisis cuantitativo y cualitativo, se proporciona una base sólida para formular estrategias orientadas al incremento la eficiencia operativa, como reducir tiempos improductivos y optimizar la utilización de recursos humanos y materiales. El análisis cuantitativo, apoyado en indicadores de desempeño y datos históricos, permite identificar tendencias, cuantificar los tiempos de intervención y evidenciar posibles cuellos de botella que afectan el flujo de trabajo. Por otro lado, el análisis cualitativo, basado en la percepción y experiencia del personal técnico, ofrece una visión complementaria que ayuda a comprender las causas subyacentes de las ineficiencias y las oportunidades de mejora desde una perspectiva práctica y operativa. (Draheim, 2010)

La combinación de ambos enfoques posibilita una comprensión integral y sistémica del proceso, orientada no solo a resolver problemas inmediatos, sino también a establecer mecanismos de control y mejora continua a largo plazo. De esta manera, se promueve la estandarización de las actividades, se fortalece la comunicación interna y se fomenta una cultura organizacional centrada en la calidad y la satisfacción del cliente, aspectos fundamentales para consolidar la competitividad y la sostenibilidad en el mercado automotriz actual. (Ranđelović, S., & Radosavljević, 2018)

2. MARCO TEÓRIO

2.1. Antecedentes

El crecimiento del mercado de vehículos seminuevos en el Ecuador ha sido impulsado por la constante búsqueda de precios accesibles y opciones con tecnologías modernas. Durante el 2024 las unidades comercializadas en el país tuvieron un incremento del 10% (Astudillo, 2024), lo que ha generado un aumento significativo en la demanda de servicios de mecánica para su alistamiento. Esta situación ha provocado cuellos de botella en los talleres, derivando en retrasos en los tiempos de entrega de las unidades a los clientes.

2.2. Alistamiento de Vehículos

(TOYOTA - Technical Service Bulletin, 2023) lo define como el proceso fundamental para satisfacer la expectativa del cliente con respecto al vehículo. Las áreas principales dentro del proceso de alistamiento son:

- Funcionamiento correcto de los sistemas mecánicos

- Funcionamiento correcto de los accesorios eléctricos (incluida la luz interior, el reloj y el reinicio de la radio).
- Inspección cuidadosa para detectar rayones o desconchones de pintura y abolladuras o golpes en la carrocería.
- Limpieza interior.

Un tiempo de reacondiciona rápido sería de 72 horas, pero algunos vehículos usados pueden tomas hasta 12 días para estar listo para la venta. Cuanto más tiempo se tarde en terminar el reacondicionamiento del vehículo usado, la rentabilidad se reduce (Brunson, 2020).

2.3. Proceso de Servicio Automotriz

(Rezin, 2009) define este proceso como la serie de pasos o tareas necesarias para resolver una solicitud de servicio técnico automotriz, se asegura que las reparaciones se hayan realizado correctamente y el cliente esté satisfecho. Consta de los siguientes pasos:

- Recepción
- Distribución del trabajo
- Rendimiento laboral
- Control de calidad
- Entrega

2.4. Análisis de procesos

A nivel conceptual, es una poderosa herramienta visual para identificar brechas sistémicas dentro de una organización. Puede incentivar a la alta dirección a definir prioridades en la administración del proceso. El análisis de procesos de negocio a nivel táctico es una forma de minimizar costos, estandarizar la ejecución del trabajo y contribuir a mejorar la productividad del trabajo diario (Барсегян, Куприянов, Холод, Тесс, & Елизаров, 2009).

(ТЮКАВКИН & МИРОНОВА, 2022) establece que la documentación generada por el análisis incluye los siguientes elementos:

- Propósito del proceso
- Mapeo del proceso (qué se hace, entradas y salidas)
- Métricas del proceso

- Causas de la ineficiencia del proceso
- Identificación de desperdicios
- Potencial para mejorar
- Soluciones recomendadas

La documentación debe describir claramente el estado actual del proceso y contener la información necesaria para planificar cambios.

2.5. Propósito del proceso

De acuerdo con (Dumas, Rosa, Mendling, & Reijers, 2017), el objetivo de establecer el propósito del proceso es que una organización debe centrarse en aquellos procesos que crean valor de relevancia estratégica o que presentan problemas sustanciales (o ambos). Esta tarea se basa en la observación de procesos de negocio en base a términos de importancia y madurez.

2.6. Mapeo del proceso (VSM – Mapa Flujo de Valor)

El mapeo de procesos es una herramienta esencial en la gestión organizacional, ya que permite visualizar de forma estructurada la secuencia de actividades, los flujos de información y los puntos críticos de un proceso. Esta técnica brinda una comprensión integral de cómo se ejecutan las tareas, cómo interactúan las distintas áreas y qué oportunidades de mejora existen para optimizar la eficiencia y la productividad (Coimbra, 2024).

2.7. Ineficiencia del proceso

Un proceso es ineficiente cuando presenta desequilibrios entre los recursos utilizados (tiempo, dinero, esfuerzo humano, tecnología) y el valor generado para el cliente o la organización. Esta ineficiencia puede deberse a múltiples factores que se manifiestan a lo largo de todo el proceso (Барсегян, Куприянов, Холод, Тесс, & Елизаров, 2009).

2.8. Identificación de desperdicios

(Voehl, Harrington, Mignosa, & Charron, 2014) exponen que los desperdicios se componen generalmente de actividades innecesarias que pueden describirse cualitativa o cuantitativamente. La identificación y reducción de desperdicios es una forma eficaz de aumentar la rentabilidad. Las nueve categorías son:

1. Sobreproducción

2. Exceso de inventario
3. Defectos
4. Procesos adicionales
5. Espera
6. Movimiento
7. Transporte
8. Personal subutilizado
9. Comportamiento de los empleados

2.9. Potencial para mejorar

La optimización de procesos es fundamental para mejorar la eficiencia y reducir la burocracia en las organizaciones. Para lograr este objetivo, es necesario analizar los datos de desempeño e identificar oportunidades de mejora. Algunas técnicas para realizar este análisis incluyen:

- Análisis de Indicadores de Desempeño
- Análisis de mapeo de flujo de valor
- Análisis de Causa y Efecto (Diagrama de Ishikawa):
- Análisis de datos.

2.10. Diagrama de Ishikawa

El Diagrama de Ishikawa, también conocido como Diagrama de Causa y Efecto o Diagrama de Espina de Pescado, es una valiosa herramienta para analizar problemas e ineficiencias en los procesos. De acuerdo con, (Coimbra, 2024), el diagrama de Ishikawa está estructurado como un gráfico de espina de pescado, donde el problema en cuestión se coloca en la parte superior y las principales categorías de posibles causas se representan como granos. Las categorías comunes incluyen "Mano de obra", "Materiales", "Maquinaria", "Métodos", "Mediciones" y "Medio ambiente" (o "Materia prima" y "Equipo", según el contexto).

2.11. Indicadores de Gestión

Un indicador de desempeño es una métrica utilizada para medir, monitorear y evaluar el comportamiento o la efectividad de un proceso, actividad o área dentro de una organización. Sirve como una brújula que señala si se están cumpliendo los objetivos establecidos y en qué medida. (Barrera, Barrientos, Santiago, & Canepa, 2018)

2.12. TAT - Turnaround Time

Es un indicador clave que mide el tiempo total que transcurre desde que se inicia un proceso hasta que finaliza. En contextos operativos o de servicio, se refiere al tiempo que toma completar una tarea o entregar un producto desde su solicitud hasta su finalización. (Barrera, Barrientos, Santiago, & Canepa, 2018)

3. MATERIALES Y METODOS

3.1. Materiales

3.1.1. Población

La población de investigación está conformada por 6 técnicos que están constantemente en el proceso de alistamiento de los vehículos semi nuevos a diario, lo que nos va a permitir obtener los datos con mayor relevancia (Rummler & Brache, 2013). Para poder seleccionar los participantes de la investigación se establecieron criterios de inclusión, exclusión y eliminación los que determinaran que técnicos son idóneos para ser seleccionados.

3.1.2. Lugar

La investigación se realizó en el concesionario ubicado en el sector norte de la ciudad de Quito, bajo las coordenadas VGFC+4R Quito, debido a que este concesionario tiene como necesidad la estandarización del proceso adecuado para el alistamiento de los vehículos semi nuevos. Con la información obtenida en el establecimiento indicado, se realizó un análisis del proceso de servicio mecánico dentro del flujo de alistamiento en el que se proponen mejoras continuas a largo plazo.

Figura 1

Ubicación del taller automotriz

Fuente: Google Maps

3.1.3. Tipo de estudio

La presente investigación se desarrolló a través de un estudio etnográfico, el cual permitió observar y comprender de forma directa el comportamiento del personal técnico involucrado en el proceso de servicio mecánico de vehículos seminuevos, dentro de su entorno operativo cotidiano, ya que los colaboradores del área técnica representan los principales actores dentro del proceso de alistamiento, y su conocimiento aporta una visión clave para el análisis (Weinzierl, Zilker, Dunzer, & Matzner, 2024). Bajo esta metodología, se utilizaron herramientas como la observación participante, entrevistas y revisión de datos internos, lo cual permitió obtener muestras de datos reales y contextualizados dentro del taller. Los instrumentos de medición empleados fueron diseñados con base en criterios de validez y confiabilidad, por lo que se procuró asegurar tanto la validez como la confiabilidad en cada instancia de recolección y análisis de datos (Draheim, 2010).

3.1.4. Procedimiento

Con el objetivo de establecer un proceso de alistamiento estandarizado y eficiente para vehículos seminuevos, se propone la aplicación del ciclo PHVA (Planificar, Hacer, Verificar, Actuar), el cual permite abordar la mejora continua bajo un enfoque estructurado (Parmenter, 2015).

4. METODOLOGÍA

4.1. Propósito del proceso

(ТЮКАВКИН & МИРОНОВА, 2022) establece que el primer paso es definir el contexto empresarial, para comprender por qué es importante el proceso, para esto es necesario responder las siguientes preguntas (tabla 1), esto se realiza mediante la observación:

Tabla 1

Bloque de preguntas para definir el propósito del proceso

PREGUNTA	RESPUESTA
¿Qué debe hacer el proceso?	Realizar mecánica correctiva y de mantenimiento en vehículos seminuevos.
¿Por qué apareció?	Previamente se tercerizaba el servicio de mecánica, ante el alto costo, la gerencia decide implementar el taller dentro de la cadena de valor.

¿Cuál es la necesidad del análisis?	Estandarizar proceso con búsqueda de la mejora continua, debido al continuo retraso de entrega del vehículo al proceso de pintura y lavado
¿Cómo encaja el proceso en la cadena de valor en la organización?	Se encarga de ofrecer confiabilidad en el FUNCIONAMIENTO del producto (vehículos seminuevos).
¿Cuál de los líderes de la organización es responsable de garantizar que el proceso logre los resultados requeridos?	Jefe de taller

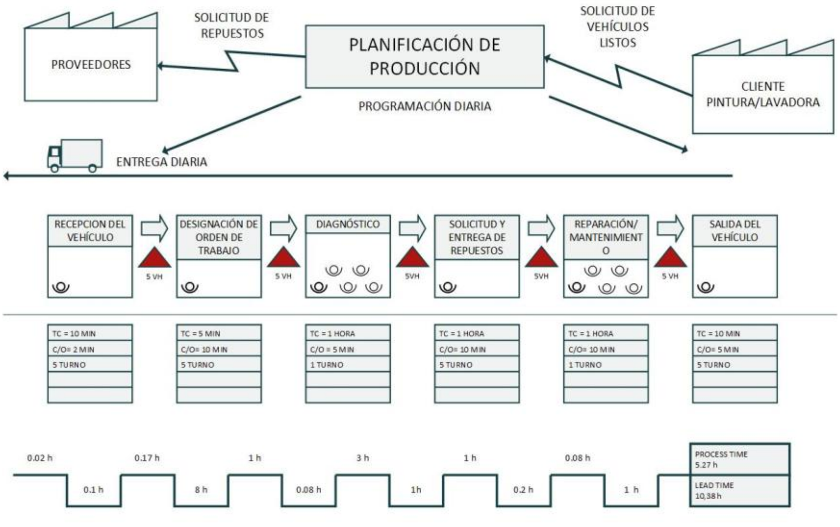
Fuente: (ТЮКАВКИН & МИРОНОВА, Análisis y gestión de procesos de empresas y organizaciones, 2022)

4.2. Mapeo del proceso

Se recopila información mediante la observación directa y se determina el proceso a través del Mapa del Flujo de Valor – VSM representado en la figura 2, de acuerdo con (Dumser, 2017), ofrece una visión global simple de todo el proceso integrando todas las informaciones necesarias para comprender de forma visual la materia e información, identificando desperdicios y sus causas.

Figura 2

Mapa de Cadena de Valor - VSM del proceso de servicio mecánico



Fuente: autores

Acorde al VSM observado, tenemos un tiempo de trabajo activo en el vehículo de 5,27 h y un total de tiempo de espera de 10,38 h para que los vehículos sean entregados al siguiente proceso.

4.3. Métricas de proceso

Para realizar el análisis de los tiempos de permanencia de los vehículos en el taller y detectar posibles cuellos de botella en el proceso, se empleará el indicador de calidad TAT (Turnaround Time), de los 360 vehículos registrados en la base de datos del taller, la cual incluye las fechas y horas de ingreso y salida de cada vehículo en el periodo entre enero y marzo del 2025.

Para facilitar un análisis más preciso, los mantenimientos han sido clasificados según su nivel de dificultad como se detalla a continuación en la tabla 2.

Tabla 2:

Categorización de la dificultad de trabajo en base a las horas invertidas

DIFICULTAD	DESCRIPCION	OPTIMO (horas)	ACEPTABLE (horas)	LIMITE MAXIMO (horas)	OBJETIVO (horas)
BASICA	Mantenimiento preventivo	0 – 2	2.1 – 3.5	3.6 – 4	4
MEDIA	Cambio de elementos de desgaste y eléctricos	4.1 – 5.5	5.6 – 7	7.1 – 8	8
ALTA	Diagnósticos y reparaciones de sistemas completos	8.1 – 15	15.1 – 30	30.1 – 40	40

Fuente: autores

Los casos cuyo TAT exceda el límite de 40 horas serán clasificados como “Excede tiempo objetivo”, lo que permitirá identificar las causas de demora y establecer parámetros de mejora adecuados.

La tabla 3 presenta el promedio de ingreso de 360 vehículos registrados en la base de datos, dentro del proceso de MECANICA correspondiente al flujo de alistamiento y clasificados según los niveles previamente definidos.

$$TAT = HORA\ DE\ ENTREGA - HORA\ DE\ INGRESO$$

Tabla 3

Promedio de horas utilizadas por cada dificultad en comparación al tiempo objetivo

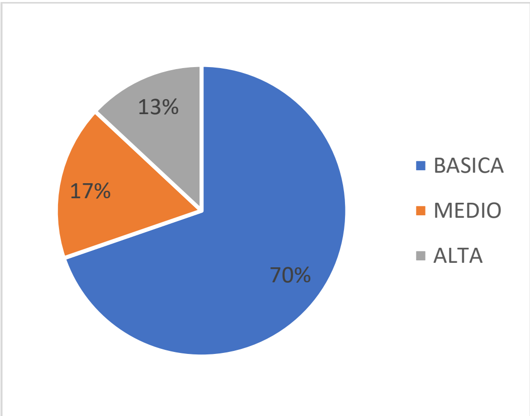
DIFICULTAD	ENERO	FEBRERO	MARZO	OBJETIVO
BASICA	8.86	11.19	10.76	4
MEDIA	38.82	44.73	42.64	8
ALTA	119.65	109.88	90.61	40

Fuente: autores

En la figura 3 determinar que el 70% del total de vehículos ingresados al taller dentro del periodo definido, entro bajo la categorización de dificultad BASICA, el 17% como MEDIA y el 13% restante como ALTA.

Figura 3

Porcentaje de vehículos ingresados



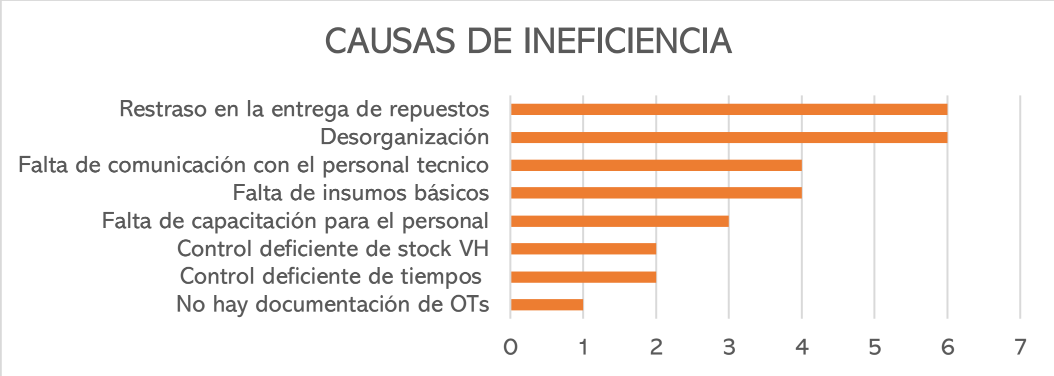
Fuente: autores

4.4. Causas de la ineficiencia del proceso

Por medio de la entrevista a 6 técnicos de planta, se observa la perspectiva de los técnicos acerca de las causas de ineficiencia del proceso mediante la pregunta: *"¿Qué cree que genera retrasos en la entrega de vehículos?"*, las respuestas y el número de técnicos que consideran la misma idea se representan en la figura 4:

Figura 4

Causas de ineficiencia conforme a la perspectiva del personal técnico



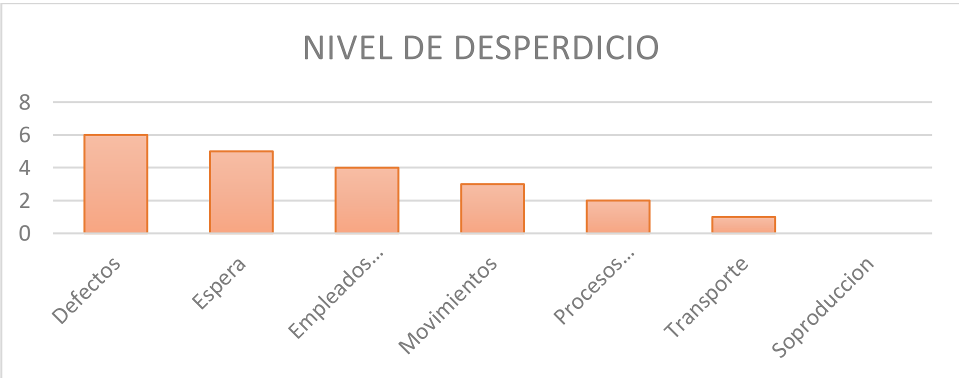
Fuente: autores

4.5. Identificación de desperdicios

(Voehl, Harrington, Mignosa, & Charron, 2014) presentan una "check list" para cada categoría de desperdicio contestadas mediante observación del proceso (descarta exceso de inventario, debido a que ese análisis pertenece al área comercial) – véase Anexo 1. La figura refleja un resumen de la cantidad de desperdicio existente por categoría.

Figura 5

Resumen de nivel de desperdicio en el proceso



Fuente: autores

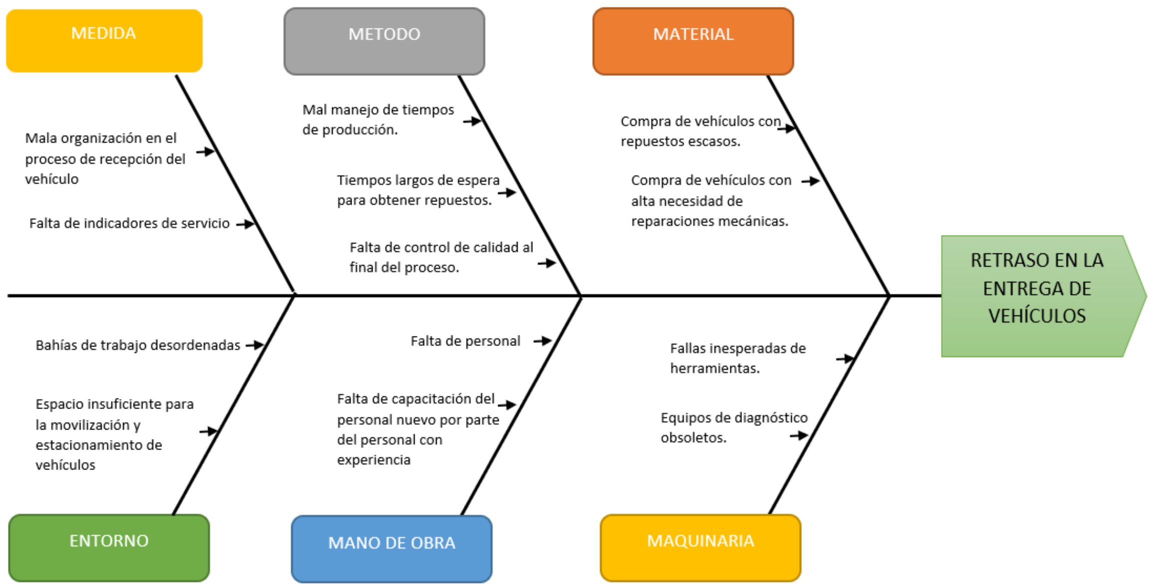
Se observa que los principales desperdicios que genera el proceso son por defectos, espera y empleados subutilizados.

4.6. Potencial de mejora

El Diagrama de Ishikawa es una herramienta valiosa para la gestión de operaciones, ya que permite un análisis estructurado de los problemas y ayuda a evitar soluciones superficiales que no aborden las verdaderas causas (Coimbra, 2024). El diagrama (figura 6) determina que el retraso en las entregas no se debe a causas aisladas, sino a un conjunto de factores interrelacionados, desde problemas operativos y de gestión, hasta deficiencias en infraestructura, personal y tecnología.

Figura 6

Diagrama Ishikawa



Fuente: autores

4.7. Soluciones recomendadas

- No hay documentación de OTs
  - Es necesario estandarizar los procesos operativos y aplicar los tiempos establecidos como objetivos para cada tipo de reparación, los cuales servirán como referencia para programar tareas y controlar desviaciones.
- Retraso en la entrega de repuestos
  - Se recomienda realizar un análisis de la carga laborar para poder justificar la contratación de una persona que se encargue específicamente de la búsqueda y compra de repuestos.
- Desorganización
  - Reorganizar físicamente las bahías de trabajo aplicando metodologías 5S y KAIZEN para permitir una mejor movilidad y eficiencia en el taller.
- Falta de comunicación con el personal técnico

- Realizar reuniones diarias y de corta duración para la asignación de tareas y seguimiento de ordenes de trabajo.

- Falta de insumos básicos
    - Integrar un sistema de requisición de insumos con el de planificación de ordenes de trabajo, conjuntamente con un responsable para el abastecimiento semanal de estos.
- Falta de capacitación para el personal
    - Diseñar un plan de capacitación continuo que incluya procesos, herramientas y buenas prácticas de trabajo, se debe implementar este plan de capacitación en los procesos estandarizados.
- Control deficiente de stock de vehículos
    - Implementar un mejor control de registro del ingreso, salida, ubicación y procedimiento de los vehículos, con códigos por estado (espera, siendo atendido, finalizado).
- Control deficiente de tiempos
    - Se recomienda implementar el KPI TAT como medida de análisis mensual para mejorar el tiempo promedio de entrega, el número de reproceses y la satisfacción al cliente.

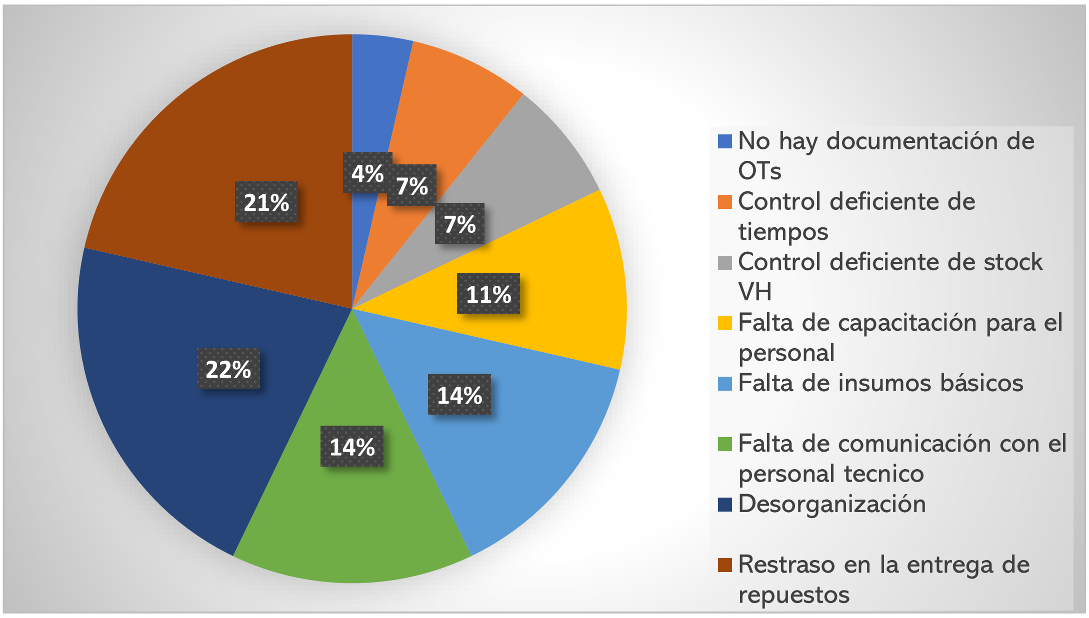
5. RESULTADOS

5.1. Causas de ineficiencia

La figura 7 se visualiza el porcentaje de participación que tiene las respectivas causas desde su perspectiva, teniendo como la falta de comunicación con el personal, desorganización, retraso en la entrega de repuestos y falta de insumos básicos, como observaciones frecuentes observadas por la mayoría de técnicos.

Figura 7

Porcentaje de participación de las causas de ineficiencia en base a la perspectiva de los técnicos.



Fuente: Autores

Esto se complementa con el análisis del VSM donde se ocupa la mayor cantidad de tiempo en la solicitud y obtención de repuestos, que representa el 51% del tiempo efectivo durante el proceso, de igual manera, converge con el análisis de desperdicios, donde se refleja el desperdicio por espera en 5 de 6 preguntas.

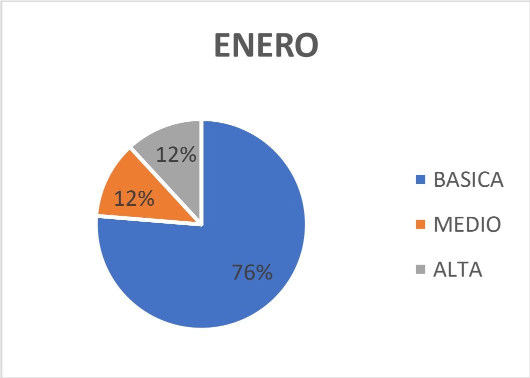
Otro punto a considerar es el desperdicio por defectos, ya que representa el más alto de desperdicios, comparando con la información de la entrevista, la mitad del personal requiere de capacitaciones para abordar problemas mecánicos más complejos.

5.2.Porcentaje de cumplimiento de TAT

En las siguientes figuras 8, 9 y 10 se puede determinar el porcentaje de vehículos que fueron categorizados cada mes en las dificultades ya establecidas.

Figura 8

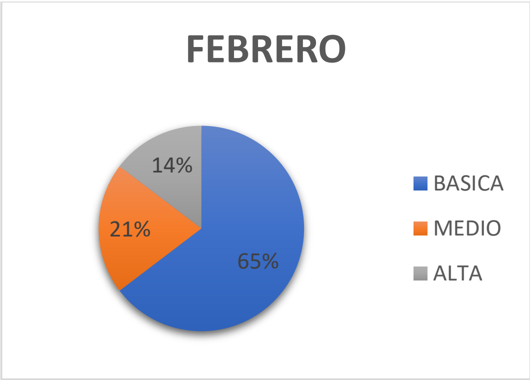
Porcentaje de vehículos ingresados en el mes de enero por cada categoría



Fuente autores

Figura 9

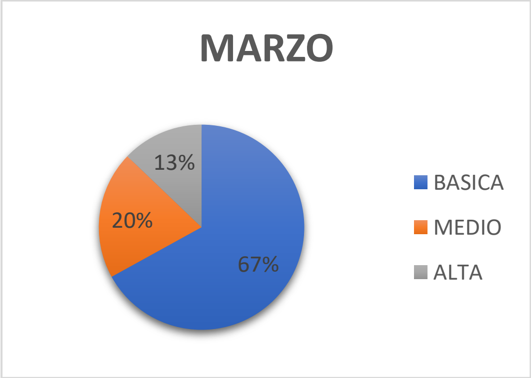
Porcentaje de vehículos ingresados en el mes de febrero por cada categoría



Fuente autores

Figura 10

Porcentaje de vehículos ingresados en el mes de marzo por cada categoría



Fuente: autores

Con la información obtenida podemos determinar el porcentaje de cumplimiento del TAT

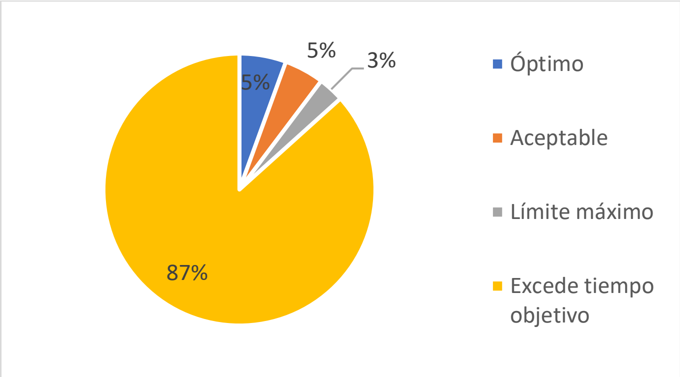
Para evaluar el desempeño en relación con el tiempo objetivo establecido, con la ayuda del TAT, el cual se calcula como el porcentaje de vehículos cuyo tiempo de atención se encuentra dentro del rango permitido.

$$\%\, Cumplimiento = \frac{Ordenes\ entregadas\ dentro\ del\ TAT\ objetivo}{Total\ de\ ordenes}$$

Este indicador permite cuantificar la eficiencia operativa del taller en función del tiempo de entrega comprometido.

Figura 11

Porcentaje de cumplimiento de vehículos atendidos



Fuente: autores

Del total de vehículos atendidos solo el 5% se encuentra dentro de un rango óptimo de entrega, mientras que el 87% excede el tiempo objetivo establecido para la entrega, dejando un 5% como aceptable y un 3% dentro del límite máximo para la entrega.

6. CONCLUSIONES

El análisis basado en el registro de 360 vehículos evidencia variaciones entre los tiempos objetivo y los tiempos reales de entrega. En trabajos básicos, cuyo objetivo es de 4 horas, se registraron promedios de hasta 11,19 horas; los trabajos catalogados como dificultad media, con un objetivo de 8 horas, se superaron las 44 horas; y en trabajos de alta dificultad, con un objetivo de 40 horas, se alcanzaron promedios de hasta 119,65 horas, por lo que solo 5% de los vehículos cumplieron con el rango óptimo de entrega, mientras que el 87% superaron el tiempo establecido, reflejando una eficiencia operativa significativamente baja.

El Mapa de Flujo de Valor identificó un tiempo activo promedio de 5,27 horas frente a 10,38 horas de espera, evidenciando que más del 49% del tiempo total corresponde a inactividad. Los resultados destacan la urgencia de implementar estrategias para reducir los tiempos de permanencia en taller, optimizar la gestión de recursos e insumos y fortalecer el control operativo mediante indicadores de desempeño y herramientas de mejora continua.

Se evidencia que el funcionamiento del taller no solo depende de la ejecución técnica de las reparaciones, sino también de elementos de gestión como la organización del espacio físico, la comunicación interna, la disponibilidad de recursos y la estandarización de procedimientos. La percepción del personal técnico refuerza esta visión, al identificar múltiples factores interrelacionados que inciden en el rendimiento del servicio, especialmente los relacionados a la obtención de repuestos, la documentación de órdenes de trabajo y la planificación operativa. Asimismo, la observación directa ha permitido identificar patrones de desperdicio en el uso del tiempo, los recursos humanos y las tareas operativas, los cuales afectan negativamente la eficiencia global del proceso. Estos resultados demuestran la necesidad de adoptar una cultura organizacional orientada a la mejora continua, respaldada por metodologías como 5S y KAIZEN que promueven un ambiente de trabajo más ordenado y eficiente, mientras que las reuniones diarias fortalecerán la comunicación y coordinación del equipo técnico. La integración de un sistema de solicitud de insumos ligado a la planificación y el diseño de un plan de capacitación continuo garantizarán un abastecimiento oportuno y una mejora constante en las competencias del personal. Finalmente, la implementación de un control riguroso de stock de vehículos y el seguimiento mediante indicadores clave de desempeño (KPI), como el TAT, facilitarán una mejor gestión del tiempo y recursos, incrementando la productividad general del taller.

7. BIBLIOGRAFÍA

Africano Rivera, S. A., & Cañón López, A. N. (2022). *Propuesta de optimización de tiempos y procesos en el taller automotriz KIA 224.* Bogota.

Aguirre, H. S. (2016). *Minería de procesos: Fundamentos y metodología de aplicación.* Editorial Pontificia Universidad Javeriana.

Astudillo, G. (marzo de 2024). ¿Cómo está la demanda de vehículos usados en Ecuador? *El Comercio*.

Barrera, R., Barrientos, V., Santiago, J., & Canepa, A. (2018). Gestión de procesos de negocio. *INVENTIO La génesis de la cultuta universitaria en Morelos*, 43-48.

Bernal, Á. A. (2012). Manejo y Optimización de las Operaciones de Mantenimiento. En Á. A. Bernal, *Manejo y Optimización de las Operaciones de Mantenimiento* (págs. 16-17). Guayaquil, Ecuador: ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

Brand, W. (2015). *Business Process Management for Dummies.* John Wiley & Sons, Inc.

Cheung, Y., & Bal, J. (1998). Process analysis techniques and tools for business improvements. *Business Process Management Journal*, 274-290.

Conger, S. (2011). *Process mapping and management.* Business expert press.

Draheim, D. (2010). *Business Process Technology.* Austria: Springer.

Dumas, M., Rosa, M. L., Mendling, J., & Reijers, H. A. (2017). *Fundamentals of Business Process Management.* Berlin: Springer.

Gadatsch, A. (2023). Analysis, Modelling, Optimisation and Controlling of Processes. Springer Nature. *Business Process Management*, Capitulo 1.

Gadatsch, A. (2023). *Business Process Management: Analysis, Modelling, Optimisation and Controlling of Processes.* Sankt Augustin, Germany: Springer.

Gadatsch, A. (2023). *Business Process Management: Analysis, Modelling, Optimisation and Controlling of Processes.* Springer.

Gaiardelli, P., Saccani, N., & Songini, L. (2007). Performance measurement systems in after-sales service: an integrated framework. *International Journal of Business Performance Management,*, 145-171.

Galo, G. N., & Pascua, P. M. (2023). Productivity Improvement in an Automotive Workshop through Lean. *Proceedings of the 9th World Congress on Mechanical, Chemical, and Material Engineering (MCM'23)Brunel*, ICMIE 145: 1-10.

Gómez, K., Gálvez, D., & Ferreira, G. (2019). Procesos de Negocio en la Gestión Empresarial. *Revista Metropolitana de Ciencias Aplicadas*, 99-111.

Guzzo, A., Joaristi, M., & Antonio Rullo, E. S. (2021). A multi-perspective approach for the analysis of complex business processes behavior . *Expert Systems With Applications* .

Hung, R. Y.-Y. (2011). Business Process Management as Competitive Advantage: a Review and Empirical Study. *Total Quality Management & Business Excellence*, 21-40.

Hunt, V. (1996). *Process mapping: how to reengineer your business processes.* John Wiley & Sons.

Lindsay, A., Downs, D., & Lunn, K. (2003). Business processes—attempts to find a definition. *INFORMATION AND SOFTWARE TECHNOLOGY*, 1015-1019.

Oña Tipán, A. G. (2022). *Mejoramiento del proceso de servicio de mantenimiento y reparación de vehículos del taller automotriz “Auto Solution del Valle”.* Quito.

Parmenter, D. (2015). *Key performance indicators: developing, implementing, and using winning KPIs.* John Wiley & Sons.

Ranđelović, A., S., M. R., & Radosavljević, M. (2018). Identification and analysis of key business process management factors. *Economic Themes*, 57-78.

Rummler, G. A., & Brache, A. P. (2013). *How to Manage the White Space on the.* San Francisco: Jossey-Bass.

Villagra Villanueva, J. A. (2016). *Indicadores de gestión: un enfoque práctico.* México, DF: Cengage.

Weinzierl, S., Zilker, S., Dunzer, S., & Matzner, M. (2024). Machine learning in business process management: A systematic literature. *Expert Systems With Applications*.

Weinzierl, S., Zilker, S., Dunzer, S., & Matzner, M. (2024). Machine learning in business process management: A systematic literature review. *Expert Systems With Applications*.

Варзунов, А. В., Торосян, Е. К., & Сажнева, Л. П. (2016). *Análisis y Gestión de procesos de Negocio.* San Petesburgo: MINISTERIO DE EDUCACION Y CIENCIA DE LA FEDERACIÓN RUSA.

ТЮКАВКИН, Н., & МИРОНОВА, Е. (2022). *Análisis y gestión de procesos de empresas y organizaciones.* Samara: Издательство Самарского университета.

ТЮКАВКИН, Н., & МИРОНОВА, Е. (2022). *АНАЛИТИКА И УПРАВЛЕНИЕ БИЗНЕС-ПРОЦЕССАМИ ПРЕДПРИЯТИЙ И ОРГАНИЗАЦИЙ.* САМАРА: Издательство Самарского университета.

8. ANEXOS

Anexo 1: Base de datos de entrada y salida de vehículos

PBU6105	BLANCO	3/2/2025 17:13	5/2/2025 11:51	14,63	BASICA	1 días 18 horas	Excede tiempo objetivo
PBL2845	PLATEADO	6/2/2025 12:38	7/2/2025 17:46	15,13	BASICA	1 días 5 horas	Excede tiempo objetivo
PDE2166	ROJO	8/1/2025 10:27	9/1/2025 15:39	15,20	BASICA	1 días 5 horas	Excede tiempo objetivo
PDV8113	PLATA	21/1/2025 11:42	22/1/2025 16:54	15,20	BASICA	1 días 5 horas	Excede tiempo objetivo
PCD1219	PLATEADA	25/2/2025 17:50	27/2/2025 13:03	15,22	BASICA	1 días 19 horas	Excede tiempo objetivo
PBK7826	PLATA	28/2/2025 11:22	1/3/2025 16:35	15,22	BASICA	1 días 5 horas	Excede tiempo objetivo
PCI5802	BLANCO	21/1/2025 8:31	22/1/2025 13:46	15,25	BASICA	1 días 5 horas	Excede tiempo objetivo
PDC4089	NEGRO	11/3/2025 8:59	12/3/2025 14:30	15,52	BASICA	1 días 5 horas	Excede tiempo objetivo
PCO2448	PLOMO	28/1/2025 11:08	29/1/2025 16:42	15,57	BASICA	1 días 5 horas	Excede tiempo objetivo
PDN4431	ROJO	21/1/2025 12:45	23/1/2025 8:22	15,62	BASICA	1 días 19 horas	Excede tiempo objetivo
PDQ8580	NEGRO	28/2/2025 9:55	1/3/2025 15:33	15,63	BASICA	1 días 5 horas	Excede tiempo objetivo
GTA2392	PLATEADO	1/2/2025 16:56	3/2/2025 12:49	15,88	BASICA	1 días 19 horas	Excede tiempo objetivo
GTI7703	PLOMO	21/1/2025 12:09	23/1/2025 8:08	15,98	BASICA	1 días 19 horas	Excede tiempo objetivo
TBC5854	DORADO	11/3/2025 8:15	12/3/2025 14:30	16,25	BASICA	1 días 6 horas	Excede tiempo objetivo
PCT3160	BLANCO	13/3/2025 9:56	14/3/2025 16:24	16,47	BASICA	1 días 6 horas	Excede tiempo objetivo
PDD8865	PLATEADO	11/3/2025 15:00	13/3/2025 11:30	16,50	BASICA	1 días 20 horas	Excede tiempo objetivo
PCY5873	PLOMO	27/3/2025 16:00	29/3/2025 12:54	16,90	BASICA	1 días 20 horas	Excede tiempo objetivo
PDH7640	PLATA	15/3/2025 17:32	17/3/2025 14:45	17,22	BASICA	1 días 21 horas	Excede tiempo objetivo
PBU6106	BLANCO	3/2/2025 17:13	5/2/2025 14:39	17,43	BASICA	1 días 21 horas	Excede tiempo objetivo
PDU5413	ROJO	10/2/2025 12:02	12/2/2025 9:34	17,53	BASICA	1 días 21 horas	Excede tiempo objetivo
PBU6107	BLANCO	3/2/2025 17:04	5/2/2025 14:39	17,58	BASICA	1 días 21 horas	Excede tiempo objetivo
PDM5981	NEGRO	10/3/2025 9:05	11/3/2025 16:43	17,63	BASICA	1 días 7 horas	Excede tiempo objetivo
PDO8256	VINO	25/1/2025 13:00	27/1/2025 10:52	17,87	BASICA	1 días 21 horas	Excede tiempo objetivo
PCO8087	PLATEADO	22/2/2025 10:25	24/2/2025 8:57	18,53	BASICA	1 días 22 horas	Excede tiempo objetivo
PCO7294	BLANCO	15/3/2025 15:54	17/3/2025 14:45	18,85	BASICA	1 días 22 horas	Excede tiempo objetivo
GTD1099	PLOMO	21/1/2025 16:00	23/1/2025 14:54	18,90	BASICA	1 días 22 horas	Excede tiempo objetivo
PDO9849	CREMA	15/1/2025 12:23	17/1/2025 11:41	19,30	BASICA	1 días 23 horas	Excede tiempo objetivo
PCU7410	PLOMO	1/2/2025 16:55	3/2/2025 16:18	19,38	BASICA	1 días 23 horas	Excede tiempo objetivo
PDK1009	BLANCO	29/3/2025 16:23	31/3/2025 15:47	19,40	BASICA	1 días 23 horas	Excede tiempo objetivo
PBV3386	DORADO	15/2/2025 15:31	17/2/2025 15:01	19,50	BASICA	1 días 23 horas	Excede tiempo objetivo
PCL6220	BLANCO	15/3/2025 15:05	17/3/2025 14:45	19,67	BASICA	1 días 23 horas	Excede tiempo objetivo
PDP6400	PLOMO	7/1/2025 15:45	9/1/2025 15:39	19,90	BASICA	1 días 23 horas	Excede tiempo objetivo
PCU3401	BLANCO	13/2/2025 13:55	15/2/2025 13:55	20,00	BASICA	2 días 0 horas	Excede tiempo objetivo
PDK9186	AZUL	22/2/2025 13:28	24/2/2025 13:47	20,32	BASICA	2 días 0 horas	Excede tiempo objetivo
PDL4877	PLOMO	1/2/2025 8:19	3/2/2025 8:49	20,50	BASICA	2 días 0 horas	Excede tiempo objetivo
PDH4034	NEGRO	11/1/2025 11:09	13/1/2025 11:39	20,50	BASICA	2 días 0 horas	Excede tiempo objetivo
PDI7040	AZUL	1/2/2025 11:41	3/2/2025 12:30	20,82	BASICA	2 días 0 horas	Excede tiempo objetivo
PDJ8763	BLANCO	24/2/2025 15:12	26/2/2025 16:04	20,87	BASICA	2 días 0 horas	Excede tiempo objetivo
PBZ6156	PLATA	15/3/2025 13:04	17/3/2025 14:45	21,68	BASICA	2 días 1 horas	Excede tiempo objetivo
PDF2719	VINO	22/2/2025 14:37	24/2/2025 16:35	21,97	BASICA	2 días 1 horas	Excede tiempo objetivo
PDL5554	NEGRO	28/3/2025 17:14	31/3/2025 9:35	22,35	BASICA	2 días 16 horas	Excede tiempo objetivo
PBW2452	BLANCO	4/2/2025 11:43	6/2/2025 14:05	22,37	BASICA	2 días 2 horas	Excede tiempo objetivo
PCK8715	PLOMO	3/2/2025 15:41	6/2/2025 8:20	22,65	BASICA	2 días 16 horas	Excede tiempo objetivo
PCN3902	NEGRO	26/3/2025 12:51	28/3/2025 15:45	22,90	BASICA	2 días 2 horas	Excede tiempo objetivo
PBY6278	PLATEADO	5/3/2025 17:16	8/3/2025 10:54	23,63	BASICA	2 días 17 horas	Excede tiempo objetivo
PCS6535		28/1/2025 12:02	30/1/2025 15:55	23,88	BASICA	2 días 3 horas	Excede tiempo objetivo
PBH8964	BLANCO	25/1/2025 8:54	27/1/2025 13:04	24,17	BASICA	2 días 4 horas	Excede tiempo objetivo
PCZ3906	PLOMO	25/1/2025 10:04	27/1/2025 14:26	24,37	BASICA	2 días 4 horas	Excede tiempo objetivo
PCZ4641	AZUL	21/1/2025 10:28	23/1/2025 14:54	24,43	BASICA	2 días 4 horas	Excede tiempo objetivo
PDS2505	DORADO	22/2/2025 12:30	24/2/2025 17:09	24,65	BASICA	2 días 4 horas	Excede tiempo objetivo
PCI7612	BLANCO	10/3/2025 9:24	12/3/2025 14:24	25,00	BASICA	2 días 5 horas	Excede tiempo objetivo
PDJ3918	ROJO	8/2/2025 13:34	11/2/2025 8:46	25,20	BASICA	2 días 19 horas	Excede tiempo objetivo
TBJ9655	GRIS	21/2/2025 15:36	24/2/2025 12:50	27,23	BASICA	2 días 21 horas	Excede tiempo objetivo
PCD4107	BLANCO	24/1/2025 11:14	27/1/2025 8:51	27,62	BASICA	2 días 21 horas	Excede tiempo objetivo
PCO4470	PLATEADO	19/3/2025 10:07	22/3/2025 4:38	27,88	BASICA	2 días 18 horas	Excede tiempo objetivo
PDR5993	BLANCO	5/2/2025 12:13	8/2/2025 10:31	28,30	BASICA	2 días 22 horas	Excede tiempo objetivo
PDR5864	BLANCO	5/2/2025 12:13	8/2/2025 10:31	28,30	BASICA	2 días 22 horas	Excede tiempo objetivo
PBN7084	DORADO	7/2/2025 17:46	10/2/2025 16:28	28,70	BASICA	2 días 22 horas	Excede tiempo objetivo
PCJ8318	PLOMO	28/1/2025 8:50	30/1/2025 17:32	28,70	BASICA	2 días 8 horas	Excede tiempo objetivo
PCY7238	VERDE	14/3/2025 10:51	17/3/2025 9:34	28,72	BASICA	2 días 22 horas	Excede tiempo objetivo
PDC5946	NEGRO	19/3/2025 9:05	22/3/2025 1:09	28,92	BASICA	2 días 16 horas	Excede tiempo objetivo
PCW8038	NEGRO	22/3/2025 16:05	25/3/2025 15:04	28,98	BASICA	2 días 22 horas	Excede tiempo objetivo
PCC7913	AZUL	21/3/2025 16:23	24/3/2025 15:38	29,25	BASICA	2 días 23 horas	Excede tiempo objetivo

PDD6408	PLOMO	24/1/2025 15:21	27/1/2025 15:21	30,00	MEDIO	3 días 0 horas	Excede tiempo objetivo
PCJ3596	BLANCO	25/2/2025 12:26	28/2/2025 12:27	30,02	MEDIO	3 días 0 horas	Excede tiempo objetivo
PBR7486	PLOMO	24/1/2025 8:20	27/1/2025 8:23	30,05	MEDIO	3 días 0 horas	Excede tiempo objetivo
PCL6220	BLANCO	26/2/2025 10:02	1/3/2025 10:39	30,62	MEDIO	3 días 0 horas	Excede tiempo objetivo
PCV5345	CAFE	14/2/2025 11:52	17/2/2025 12:38	30,77	MEDIO	3 días 0 horas	Excede tiempo objetivo
PDM8503	PLATEADO	24/1/2025 8:22	27/1/2025 9:45	31,38	MEDIO	3 días 1 horas	Excede tiempo objetivo
PDJ3481	AZUL	18/3/2025 16:36	22/3/2025 4:38	31,40	MEDIO	3 días 12 horas	Excede tiempo objetivo
HPB1218	PLATEADO	11/3/2025 11:38	14/3/2025 13:37	31,98	MEDIO	3 días 1 horas	Excede tiempo objetivo
CD2690	PLATEADO	21/2/2025 13:58	24/2/2025 16:05	32,12	MEDIO	3 días 2 horas	Excede tiempo objetivo
PCO6550	NEGRO	23/1/2025 16:08	27/1/2025 8:23	32,25	MEDIO	3 días 16 horas	Excede tiempo objetivo
CD2319	AZUL	24/3/2025 15:24	27/3/2025 17:40	32,27	MEDIO	3 días 2 horas	Excede tiempo objetivo
IBB6819	VINO	21/2/2025 15:41	25/2/2025 8:36	32,92	MEDIO	3 días 16 horas	Excede tiempo objetivo
PCK7676	PLATA	25/1/2025 10:40	28/1/2025 13:39	32,98	MEDIO	3 días 2 horas	Excede tiempo objetivo
PCY8381	PLATEADO	24/3/2025 11:59	27/3/2025 15:06	33,12	MEDIO	3 días 3 horas	Excede tiempo objetivo
PCY2093	AZUL	1/3/2025 16:26	5/3/2025 10:01	33,58	MEDIO	3 días 17 horas	Excede tiempo objetivo
TBG5830	DORADO	6/3/2025 15:56	10/3/2025 10:00	34,07	MEDIO	3 días 18 horas	Excede tiempo objetivo
PCE5059	PLATEADO	18/1/2025 12:01	21/1/2025 16:14	34,22	MEDIO	3 días 4 horas	Excede tiempo objetivo
PCJ6936	BLANCO	1/2/2025 14:49	5/2/2025 9:27	34,63	MEDIO	3 días 18 horas	Excede tiempo objetivo
PDS6662	BLANCO	25/1/2025 11:22	28/1/2025 16:41	35,32	MEDIO	3 días 5 horas	Excede tiempo objetivo
PDN5651	PLATA	20/2/2025 15:06	24/2/2025 10:54	35,80	MEDIO	3 días 19 horas	Excede tiempo objetivo
PCU9802	PLOMO	20/3/2025 10:04	24/3/2025 2:06	37,93	MEDIO	3 días 16 horas	Excede tiempo objetivo
TBG5965	PLOMO	18/1/2025 14:54	22/1/2025 13:17	38,38	MEDIO	3 días 22 horas	Excede tiempo objetivo
PDL2474	BLANCO	17/1/2025 8:54	20/1/2025 17:56	39,03	MEDIO	3 días 9 horas	Excede tiempo objetivo
PCU9366	BLANCO	13/2/2025 10:26	17/2/2025 10:49	40,38	MEDIO	4 días 0 horas	Excede tiempo objetivo
PDQ8508	ROJO	21/3/2025 14:24	25/3/2025 15:04	40,67	MEDIO	4 días 0 horas	Excede tiempo objetivo
PBH8964	BLANCO	10/1/2025 8:55	14/1/2025 9:56	41,02	MEDIO	4 días 1 horas	Excede tiempo objetivo
PBZ7497	PLATEADO	21/3/2025 14:03	25/3/2025 15:04	41,02	MEDIO	4 días 1 horas	Excede tiempo objetivo
GCA1096	PLOMO	24/3/2025 14:29	28/3/2025 15:45	41,27	MEDIO	4 días 1 horas	Excede tiempo objetivo
HBD7866	PLATEADO	13/3/2025 9:35	17/3/2025 11:30	41,92	MEDIO	4 días 1 horas	Excede tiempo objetivo
PBJ5192	NEGRO	15/1/2025 17:52	20/1/2025 9:54	42,03	MEDIO	4 días 16 horas	Excede tiempo objetivo
PDB8483	BLANCO	14/3/2025 9:45	18/3/2025 11:55	42,17	MEDIO	4 días 2 horas	Excede tiempo objetivo
PDQ9202	NEGRO	10/2/2025 14:23	14/2/2025 16:40	42,28	MEDIO	4 días 2 horas	Excede tiempo objetivo
PCY3517	NEGRO	25/2/2025 9:38	1/3/2025 11:59	42,35	MEDIO	4 días 2 horas	Excede tiempo objetivo
PCS3537	PLOMO	22/1/2025 15:40	27/1/2025 8:51	43,18	MEDIO	4 días 17 horas	Excede tiempo objetivo
GTL8715	NEGRO	17/1/2025 13:02	21/1/2025 16:14	43,20	MEDIO	4 días 3 horas	Excede tiempo objetivo
PDE6677	PLATEADO	13/3/2025 11:30	17/3/2025 14:45	43,25	MEDIO	4 días 3 horas	Excede tiempo objetivo
PBY9254	DORADO	18/1/2025 14:58	23/1/2025 8:27	43,48	MEDIO	4 días 17 horas	Excede tiempo objetivo
PFJ3839	BLANCO	23/3/2025 13:06	27/3/2025 17:40	44,57	MEDIO	4 días 4 horas	Excede tiempo objetivo
HCM0385	DORADO	14/2/2025 12:47	18/2/2025 17:22	44,58	MEDIO	4 días 4 horas	Excede tiempo objetivo
ABF9223	ROJO	12/3/2025 16:54	17/3/2025 11:30	44,60	MEDIO	4 días 18 horas	Excede tiempo objetivo
PCY5847	BLANCO	14/2/2025 10:36	18/2/2025 15:22	44,77	MEDIO	4 días 4 horas	Excede tiempo objetivo
PCV8951	PLOMO	21/2/2025 8:24	25/2/2025 14:06	45,70	MEDIO	4 días 5 horas	Excede tiempo objetivo
IBB3759	DORADO	24/3/2025 15:51	29/3/2025 12:18	46,45	MEDIO	4 días 20 horas	Excede tiempo objetivo
PBN4637	CAFE	1/3/2025 16:12	6/3/2025 12:54	46,70	MEDIO	4 días 20 horas	Excede tiempo objetivo
PDC9893	NEGRO	1/2/2025 8:32	5/2/2025 15:58	47,43	MEDIO	4 días 7 horas	Excede tiempo objetivo
HBC6042	NEGRO	19/2/2025 13:45	24/2/2025 11:35	47,83	MEDIO	4 días 21 horas	Excede tiempo objetivo
PCU7557	NEGRO	26/3/2025 13:22	31/3/2025 12:18	48,93	MEDIO	4 días 22 horas	Excede tiempo objetivo
PCD9080	BLANCO	28/2/2025 16:30	5/3/2025 15:40	49,17	MEDIO	4 días 23 horas	Excede tiempo objetivo
PBB9327	PLATEADO	20/2/2025 13:50	25/2/2025 13:35	49,75	MEDIO	4 días 23 horas	Excede tiempo objetivo
PBY5818	PLATEADO	12/2/2025 9:38	17/2/2025 10:00	50,37	MEDIO	5 días 0 horas	Excede tiempo objetivo
RBA3957	PLATEADO	12/2/2025 14:05	17/2/2025 15:01	50,93	MEDIO	5 días 0 horas	Excede tiempo objetivo
PCA4578	NEGRO	29/1/2025 12:21	3/2/2025 13:28	51,12	MEDIO	5 días 1 horas	Excede tiempo objetivo
PCB6734	AZUL	22/3/2025 16:17	27/3/2025 17:40	51,38	MEDIO	5 días 1 horas	Excede tiempo objetivo
PCO6567	PLATEADO	26/3/2025 11:21	31/3/2025 12:46	51,42	MEDIO	5 días 1 horas	Excede tiempo objetivo
PBM5242	NEGRO	24/1/2025 13:13	29/1/2025 16:42	53,48	MEDIO	5 días 3 horas	Excede tiempo objetivo
PDR5998	BLANCO	5/2/2025 12:13	11/2/2025 8:46	56,55	MEDIO	5 días 20 horas	Excede tiempo objetivo
HBC2305	AZUL	12/2/2025 11:40	18/2/2025 9:33	57,88	MEDIO	5 días 21 horas	Excede tiempo objetivo
PCL1827	NEGRO	28/2/2025 17:58	6/3/2025 16:00	58,03	MEDIO	5 días 22 horas	Excede tiempo objetivo
PFD5763	PLATEADO	8/2/2025 8:17	13/2/2025 17:19	59,03	MEDIO	5 días 9 horas	Excede tiempo objetivo
PIH0628	PLOMO	22/3/2025 16:10	28/3/2025 15:45	59,58	MEDIO	5 días 23 horas	Excede tiempo objetivo
PDC1646	PLOMO	11/2/2025 12:21	17/2/2025 12:02	59,68	MEDIO	5 días 23 horas	Excede tiempo objetivo
PBU1554	BLANCO	11/3/2025 11:42	17/3/2025 11:30	59,80	MEDIO	5 días 23 horas	Excede tiempo objetivo
PCG9749	PLOMO	29/1/2025 11:22	4/2/2025 11:32	60,17	ALTA	6 días 0 horas	Excede tiempo objetivo

PCU7428	BLANCO	9/1/2025 9:08	9/1/2025 16:45	7,62	BASICA	0 días 7 horas	Excede tiempo objetivo
PDP3002	BLANCO	28/1/2025 9:02	28/1/2025 16:42	7,67	BASICA	0 días 7 horas	Excede tiempo objetivo
PFI1941	ROJO	21/3/2025 14:18	22/3/2025 12:05	7,78	BASICA	0 días 21 horas	Excede tiempo objetivo
PBO3657	PLOMO	26/2/2025 15:00	27/2/2025 13:03	8,05	BASICA	0 días 22 horas	Excede tiempo objetivo
PCP3728	BLANCO	1/3/2025 8:50	1/3/2025 16:56	8,10	BASICA	0 días 8 horas	Excede tiempo objetivo
PDR5864	BLANCO	31/3/2025 9:54	31/3/2025 18:22	8,10	BASICA	0 días 8 horas	Excede tiempo objetivo
PCW1321	AZUL	30/1/2025 16:48	31/1/2025 14:56	8,13	BASICA	0 días 22 horas	Excede tiempo objetivo
PDU3269	PLATEADO	29/1/2025 14:27	30/1/2025 12:36	8,15	BASICA	0 días 22 horas	Excede tiempo objetivo
TBG4759	BLANCO	28/1/2025 10:08	29/1/2025 8:18	8,17	BASICA	0 días 22 horas	Excede tiempo objetivo
PPA7744	PLOMO	29/1/2025 8:47	29/1/2025 16:58	8,18	BASICA	0 días 8 horas	Excede tiempo objetivo
PDT6679	CREMA	1/3/2025 8:45	1/3/2025 17:01	8,27	BASICA	0 días 8 horas	Excede tiempo objetivo
PDM6193	PLATEADO	4/2/2025 15:17	5/2/2025 13:35	8,30	BASICA	0 días 22 horas	Excede tiempo objetivo
PBI8263	ROJO	15/1/2025 16:46	16/1/2025 15:06	8,33	BASICA	0 días 22 horas	Excede tiempo objetivo
PDO2732	BLANCO	10/1/2025 8:00	10/1/2025 16:21	8,35	BASICA	0 días 8 horas	Excede tiempo objetivo
PCO5974	PLATEADO	18/3/2025 8:06	18/3/2025 16:38	8,53	BASICA	0 días 8 horas	Excede tiempo objetivo
HBB8609	PLATEADO	15/1/2025 16:27	16/1/2025 15:06	8,65	BASICA	0 días 22 horas	Excede tiempo objetivo
PDR8228	VINO	10/2/2025 17:55	11/2/2025 16:34	8,65	BASICA	0 días 22 horas	Excede tiempo objetivo
PCV9199	AZUL	29/1/2025 17:13	30/1/2025 15:54	8,68	BASICA	0 días 22 horas	Excede tiempo objetivo
PDC7953	ROJO	13/2/2025 15:55	14/2/2025 14:36	8,68	BASICA	0 días 22 horas	Excede tiempo objetivo
PDW4301	NEGRO	21/3/2025 13:16	22/3/2025 12:05	8,82	BASICA	0 días 22 horas	Excede tiempo objetivo
GSO3481	PLATA	9/1/2025 11:10	10/1/2025 10:00	8,83	BASICA	0 días 22 horas	Excede tiempo objetivo
TBG9753	VINO	27/1/2025 12:08	28/1/2025 11:05	8,95	BASICA	0 días 22 horas	Excede tiempo objetivo
PDF3603	PLOMO	22/1/2025 10:54	23/1/2025 9:52	8,97	BASICA	0 días 22 horas	Excede tiempo objetivo
PCU9530	PLATEADO	7/2/2025 10:33	8/2/2025 9:35	9,03	BASICA	0 días 23 horas	Excede tiempo objetivo
PDP7131	PLATEADO	29/1/2025 16:52	30/1/2025 16:15	9,38	BASICA	0 días 23 horas	Excede tiempo objetivo
PCU9530	PLATEADO	7/2/2025 10:33	8/2/2025 10:12	9,65	BASICA	0 días 23 horas	Excede tiempo objetivo
PCR2500	PLATEADO	24/1/2025 15:27	25/1/2025 15:26	9,98	BASICA	0 días 23 horas	Excede tiempo objetivo
PFC6384	CAFE	6/2/2025 13:30	7/2/2025 13:30	10,00	BASICA	1 días 0 horas	Excede tiempo objetivo
PBR4426	VERDE	24/2/2025 15:36	25/2/2025 15:36	10,00	BASICA	1 días 0 horas	Excede tiempo objetivo
PCL3689	BLANCO	19/3/2025 14:06	20/3/2025 14:06	10,00	BASICA	1 días 0 horas	Excede tiempo objetivo
PDN4779	PLATEADO	13/1/2025 17:39	14/1/2025 17:40	10,02	BASICA	1 días 0 horas	Excede tiempo objetivo
PDD5574	BLANCO	22/1/2025 10:19	23/1/2025 10:20	10,02	BASICA	1 días 0 horas	Excede tiempo objetivo
PCJ6473	PLOMO	14/1/2025 14:47	15/1/2025 14:54	10,12	BASICA	1 días 0 horas	Excede tiempo objetivo
PDF3456	PLOMO	24/2/2025 12:54	25/2/2025 13:06	10,20	BASICA	1 días 0 horas	Excede tiempo objetivo
PFF4306	PLOMO	13/3/2025 9:56	14/3/2025 10:08	10,20	BASICA	1 días 0 horas	Excede tiempo objetivo
PCJ6936	BLANCO	10/2/2025 14:45	11/2/2025 15:00	10,25	BASICA	1 días 0 horas	Excede tiempo objetivo
ABI5191	PLATEADO	20/1/2025 15:43	21/1/2025 16:14	10,52	BASICA	1 días 0 horas	Excede tiempo objetivo
PDA8863	AZUL	7/3/2025 10:36	8/3/2025 11:10	10,57	BASICA	1 días 0 horas	Excede tiempo objetivo
PDG3002	PLOMO	20/3/2025 16:02	21/3/2025 16:45	10,72	BASICA	1 días 0 horas	Excede tiempo objetivo
PCW8429	BLANCO	8/1/2025 11:49	9/1/2025 12:47	10,97	BASICA	1 días 0 horas	Excede tiempo objetivo
PCQ6609	PLATEADO	19/2/2025 9:56	20/2/2025 11:00	11,07	BASICA	1 días 1 horas	Excede tiempo objetivo
PDF3603	PLOMO	22/1/2025 10:54	23/1/2025 12:03	11,15	BASICA	1 días 1 horas	Excede tiempo objetivo
PDV2698	BLANCO	20/1/2025 13:44	21/1/2025 14:54	11,17	BASICA	1 días 1 horas	Excede tiempo objetivo
JBA1251	BLANCO	16/1/2025 15:15	17/1/2025 16:28	11,22	BASICA	1 días 1 horas	Excede tiempo objetivo
PCN7805	PLATA	13/1/2025 14:20	14/1/2025 15:45	11,42	BASICA	1 días 1 horas	Excede tiempo objetivo
PCJ8902	NEGRO	17/1/2025 14:29	18/1/2025 15:55	11,43	BASICA	1 días 1 horas	Excede tiempo objetivo
PCJ8902	NEGRO	17/1/2025 14:29	18/1/2025 15:55	11,43	BASICA	1 días 1 horas	Excede tiempo objetivo
PDF2667	PLOMO	27/1/2025 14:30	28/1/2025 16:34	12,07	BASICA	1 días 2 horas	Excede tiempo objetivo
PQE0926	VERDE	15/3/2025 17:32	17/3/2025 9:48	12,27	BASICA	1 días 16 horas	Excede tiempo objetivo
PDO3604	PLOMO	21/1/2025 11:01	22/1/2025 13:17	12,27	BASICA	1 días 2 horas	Excede tiempo objetivo
GCT2929	BLANCO	14/1/2025 12:28	15/1/2025 14:54	12,43	BASICA	1 días 2 horas	Excede tiempo objetivo
TBG7878	BLANCO	28/3/2025 13:55	29/3/2025 16:28	12,55	BASICA	1 días 2 horas	Excede tiempo objetivo
PDS6637	PLATA	9/1/2025 10:00	10/1/2025 12:45	12,75	BASICA	1 días 2 horas	Excede tiempo objetivo
PCY8381	PLATEADO	24/3/2025 11:59	25/3/2025 15:04	13,08	BASICA	1 días 3 horas	Excede tiempo objetivo
PDM5068	ROJO	8/3/2025 16:43	10/3/2025 10:00	13,28	BASICA	1 días 17 horas	Excede tiempo objetivo
GPN0668	NEGRO	20/1/2025 13:22	21/1/2025 16:42	13,33	BASICA	1 días 3 horas	Excede tiempo objetivo
PDI3849	BLANCO	25/1/2025 15:02	27/1/2025 8:23	13,35	BASICA	1 días 17 horas	Excede tiempo objetivo
PCN8873	PLOMO	19/3/2025 10:45	20/3/2025 14:06	13,35	BASICA	1 días 3 horas	Excede tiempo objetivo
PBU2624	AZUL	11/1/2025 16:08	13/1/2025 9:42	13,57	BASICA	1 días 17 horas	Excede tiempo objetivo
PBO3669	NEGRO	10/2/2025 13:37	11/2/2025 17:20	13,72	BASICA	1 días 3 horas	Excede tiempo objetivo
PDH6378	NEGRO	8/2/2025 15:27	10/2/2025 9:21	13,90	BASICA	1 días 17 horas	Excede tiempo objetivo
PCW9455	BLANCO	22/2/2025 15:57	24/2/2025 9:56	13,98	BASICA	1 días 17 horas	Excede tiempo objetivo
PCL5130	BLANCO	24/3/2025 10:44	25/3/2025 15:04	14,33	BASICA	1 días 4 horas	Excede tiempo objetivo

PLACA	COLOR	INGRESO	SALIDA	DIFERENCIA TOTAL	DIFICULTAD	TIEMPO	OBSERVACION
PFB4294	ROJO	24/1/2025 9:15	24/1/2025 9:32	0,28	BASICA	0 días 0 horas	Optimo
PFB1428	BLANCO	24/2/2025 10:45	24/2/2025 11:13	0,47	BASICA	0 días 0 horas	Optimo
PCL3158	PLOMO	27/1/2025 17:09	27/1/2025 17:49	0,67	BASICA	0 días 0 horas	Optimo
PCY2231	BLANCO	10/1/2025 13:08	10/1/2025 14:05	0,95	BASICA	0 días 0 horas	Optimo
TBI5760	CAFE	10/2/2025 12:12	10/2/2025 13:10	0,97	BASICA	0 días 0 horas	Optimo
PDL3426	BLANCO	8/2/2025 13:33	8/2/2025 14:33	1,00	BASICA	0 días 1 horas	Optimo
PCQ5511	ROJO	4/2/2025 9:00	4/2/2025 10:02	1,03	BASICA	0 días 1 horas	Optimo
PFH3559	PLOMO	27/1/2025 16:44	27/1/2025 17:48	1,07	BASICA	0 días 1 horas	Optimo
PCO8499	PLATEADO	14/1/2025 8:45	14/1/2025 9:56	1,18	BASICA	0 días 1 horas	Optimo
PFE6038	BLANCO	20/1/2025 16:45	20/1/2025 17:56	1,18	BASICA	0 días 1 horas	Optimo
PDR7030	PLOMO	22/1/2025 18:11	23/1/2025 9:14	1,23	BASICA	0 días 15 horas	Optimo
PDT5521	PLOMO	19/2/2025 18:34	20/2/2025 9:26	1,43	BASICA	0 días 14 horas	Optimo
PDK6132	ROJO	25/1/2025 9:36	25/1/2025 11:07	1,52	BASICA	0 días 1 horas	Optimo
PFF5627	PLATEADO	3/2/2025 9:24	3/2/2025 10:56	1,53	BASICA	0 días 1 horas	Optimo
PDZ6654	AZUL	23/1/2025 10:27	23/1/2025 12:03	1,60	BASICA	0 días 1 horas	Optimo
PDJ7618	PLOMO	18/3/2025 14:59	18/3/2025 16:38	1,65	BASICA	0 días 1 horas	Optimo
PDT2022	AZUL	21/1/2025 14:25	21/1/2025 16:14	1,82	BASICA	0 días 1 horas	Optimo
PDL5201	BLANCO	5/2/2025 11:31	5/2/2025 13:21	1,83	BASICA	0 días 1 horas	Optimo
PCT4146	PLOMO	20/1/2025 17:33	21/1/2025 9:26	1,88	BASICA	0 días 15 horas	Optimo
PDJ2772	DORADA	6/2/2025 10:17	6/2/2025 12:11	1,90	BASICA	0 días 1 horas	Optimo
PDE9864	BLANCO	26/2/2025 15:00	26/2/2025 17:02	2,03	BASICA	0 días 2 horas	Aceptable
PCZ7619	PLATEADO	22/3/2025 10:57	22/3/2025 13:04	2,12	BASICA	0 días 2 horas	Aceptable
PDP2073	PLOMO	16/1/2025 17:41	17/1/2025 9:52	2,18	BASICA	0 días 16 horas	Aceptable
HBC6503	BLANCO	29/3/2025 9:57	29/3/2025 12:18	2,35	BASICA	0 días 2 horas	Aceptable
PCO6524	BLANCO	17/1/2025 14:07	17/1/2025 16:28	2,35	BASICA	0 días 2 horas	Aceptable
TBH1328	ROJO	20/3/2025 12:08	20/3/2025 14:30	2,37	BASICA	0 días 2 horas	Aceptable
PDH2985	BLANCO	19/3/2025 10:45	19/3/2025 13:09	2,40	BASICA	0 días 2 horas	Aceptable
PDN5728	NEGRO	19/3/2025 10:34	19/3/2025 13:09	2,58	BASICA	0 días 2 horas	Aceptable
PDR4483	ROJO	28/1/2025 16:10	29/1/2025 8:46	2,60	BASICA	0 días 16 horas	Aceptable
PDP8060	PLATEADO	12/2/2025 9:25	12/2/2025 12:15	2,83	BASICA	0 días 2 horas	Aceptable
TBG1794	DORADO	7/1/2025 13:48	7/1/2025 16:47	2,98	BASICA	0 días 2 horas	Aceptable
IBA5413	PLATEADO	18/3/2025 8:45	18/3/2025 11:50	3,08	BASICA	0 días 3 horas	Aceptable
PDD5574	BLANCO	5/2/2025 10:24	5/2/2025 13:31	3,12	BASICA	0 días 3 horas	Aceptable
PCJ6222	PLOMO	1/2/2025 10:05	1/2/2025 13:23	3,30	BASICA	0 días 3 horas	Aceptable
MBE4774	VERDE	22/1/2025 16:01	23/1/2025 9:22	3,35	BASICA	0 días 17 horas	Aceptable
PDC7679	PLATEADO	28/3/2025 8:45	28/3/2025 12:08	3,38	BASICA	0 días 3 horas	Aceptable
PBJ9645	NEGRO	24/1/2025 14:26	24/1/2025 17:52	3,43	BASICA	0 días 3 horas	Aceptable
PPC2774	BLANCO	16/1/2025 17:27	17/1/2025 11:01	3,57	BASICA	0 días 17 horas	Limite maximo
TBI1486	BLANCO	31/3/2025 16:47	1/4/2025 10:23	3,60	BASICA	0 días 17 horas	Limite maximo
PDM3525	CAFE	23/1/2025 12:26	23/1/2025 16:02	3,60	BASICA	0 días 3 horas	Limite maximo
PBU9176	PLOMO	5/2/2025 11:11	5/2/2025 14:47	3,60	BASICA	0 días 3 horas	Limite maximo
PCG7713	BLANCO	12/3/2025 11:57	12/3/2025 15:34	3,62	BASICA	0 días 3 horas	Limite maximo
PDH3917	PLATEADO	15/1/2025 17:57	16/1/2025 11:38	3,68	BASICA	0 días 17 horas	Limite maximo
PCQ5320	BLANCO	15/1/2025 17:55	16/1/2025 11:38	3,72	BASICA	0 días 17 horas	Limite maximo
HBB3145	PLOMO	24/1/2025 8:20	24/1/2025 12:03	3,72	BASICA	0 días 3 horas	Limite maximo
PBV5415	DORADO	12/3/2025 10:12	12/3/2025 13:59	3,78	BASICA	0 días 3 horas	Limite maximo
PDC5946	NEGRO	16/3/2025 17:35	17/3/2025 11:30	3,92	BASICA	0 días 17 horas	Limite maximo
TBH1565	BLANCO	16/1/2025 17:41	17/1/2025 11:41	4,00	BASICA	0 días 18 horas	Limite maximo
PDL5565	BLANCO	17/3/2025 10:44	17/3/2025 14:45	4,02	BASICA	0 días 4 horas	Excede tiempo objetivo
PFM3151	PLOMO	12/3/2025 9:54	12/3/2025 13:59	4,08	BASICA	0 días 4 horas	Excede tiempo objetivo
PQO0700	DORADO	19/2/2025 17:49	20/2/2025 12:00	4,18	BASICA	0 días 18 horas	Excede tiempo objetivo
GSX3597	PLATEADO	5/3/2025 11:45	5/3/2025 16:00	4,25	BASICA	0 días 4 horas	Excede tiempo objetivo
PDY4622	BLANCO	24/1/2025 17:46	25/1/2025 12:04	4,30	BASICA	0 días 18 horas	Excede tiempo objetivo
PBB7923	ROJO	17/3/2025 14:45	18/3/2025 9:05	4,33	BASICA	0 días 18 horas	Excede tiempo objetivo
PCM7323	BLANCO	20/3/2025 9:06	20/3/2025 13:26	4,33	BASICA	0 días 4 horas	Excede tiempo objetivo
PDN9773	PLATEADO	22/1/2025 11:29	22/1/2025 15:54	4,42	BASICA	0 días 4 horas	Excede tiempo objetivo
PCZ8294	ROJO	3/2/2025 8:54	3/2/2025 13:21	4,45	BASICA	0 días 4 horas	Excede tiempo objetivo
PDX2953	BLANCO	31/1/2025 17:27	1/2/2025 12:00	4,55	BASICA	0 días 18 horas	Excede tiempo objetivo
PDL5165	PLATEADO	10/2/2025 11:17	10/2/2025 15:55	4,63	BASICA	0 días 4 horas	Excede tiempo objetivo
PCZ8410	NEGRO	26/2/2025 11:56	26/2/2025 16:34	4,63	BASICA	0 días 4 horas	Excede tiempo objetivo
PCX4642	ROJO	24/1/2025 13:05	24/1/2025 17:47	4,70	BASICA	0 días 4 horas	Excede tiempo objetivo
TBI3471	VINO	5/3/2025 8:00	5/3/2025 12:45	4,75	BASICA	0 días 4 horas	Excede tiempo objetivo

PDC3190	ROJO	22/1/2025 18:51	29/1/2025 8:18	60,30	ALTA	6 días 13 horas	Excede tiempo objetivo
PCF1644	BLANCO	1/3/2025 14:48	7/3/2025 16:50	62,03	ALTA	6 días 2 horas	Excede tiempo objetivo
LDB6754	BLANCO	19/3/2025 10:39	25/3/2025 15:04	64,42	ALTA	6 días 4 horas	Excede tiempo objetivo
PBH4566	NEGRO	24/1/2025 9:16	30/1/2025 15:06	65,83	ALTA	6 días 5 horas	Excede tiempo objetivo
PFC3880	PLOMO	11/2/2025 16:50	18/2/2025 15:37	68,78	ALTA	6 días 22 horas	Excede tiempo objetivo
PDK8557	BLANCO	11/2/2025 15:38	18/2/2025 14:41	69,05	ALTA	6 días 23 horas	Excede tiempo objetivo
PFF8929	PLOMO	1/3/2025 9:21	8/3/2025 9:48	70,45	ALTA	7 días 0 horas	Excede tiempo objetivo
PCF2610	CAFE	29/1/2025 11:28	5/2/2025 15:36	74,13	ALTA	7 días 4 horas	Excede tiempo objetivo
GTI7919	BLANCO	23/3/2025 13:02	31/3/2025 10:14	77,20	ALTA	7 días 21 horas	Excede tiempo objetivo
TDP0054	NEGRO	20/3/2025 9:48	27/3/2025 17:40	77,87	ALTA	7 días 7 horas	Excede tiempo objetivo
PBP2527	PLATA	9/1/2025 13:39	17/1/2025 11:41	78,03	ALTA	7 días 22 horas	Excede tiempo objetivo
GSF2942	VINO	12/2/2025 10:45	20/2/2025 9:38	78,88	ALTA	7 días 22 horas	Excede tiempo objetivo
PDA8412	BLANCO	9/1/2025 8:49	17/1/2025 8:55	80,10	ALTA	8 días 0 horas	Excede tiempo objetivo
PCR5773	PLATA	7/1/2025 14:36	15/1/2025 14:54	80,30	ALTA	8 días 0 horas	Excede tiempo objetivo
HBD4376	NEGRO	14/3/2025 9:05	22/3/2025 12:05	83,00	ALTA	8 días 3 horas	Excede tiempo objetivo
PCD4463	BLANCO	26/2/2025 9:50	6/3/2025 14:07	84,28	ALTA	8 días 4 horas	Excede tiempo objetivo
PCT6930	PLATA	26/2/2025 15:32	7/3/2025 11:47	86,25	ALTA	8 días 20 horas	Excede tiempo objetivo
PBI9420	PLATEADO	8/3/2025 17:15	17/3/2025 14:45	87,50	ALTA	8 días 21 horas	Excede tiempo objetivo
PBB4916	PLOMO	15/2/2025 11:07	24/2/2025 12:04	90,95	ALTA	9 días 0 horas	Excede tiempo objetivo
PDF3291	BLANCO	25/2/2025 10:41	6/3/2025 12:05	91,40	ALTA	9 días 1 horas	Excede tiempo objetivo
PBB8892	PLOMO	25/2/2025 10:23	6/3/2025 12:21	91,97	ALTA	9 días 1 horas	Excede tiempo objetivo
PCA9909	NEGRO	1/3/2025 11:14	10/3/2025 14:03	92,82	ALTA	9 días 2 horas	Excede tiempo objetivo
PCN3255	PLATEADO	1/3/2025 9:45	10/3/2025 14:33	94,80	ALTA	9 días 4 horas	Excede tiempo objetivo
PCQ9881	BLANCO	24/2/2025 16:56	6/3/2025 12:04	95,13	ALTA	9 días 19 horas	Excede tiempo objetivo
PDN5728	NEGRO	19/3/2025 10:34	28/3/2025 15:45	95,18	ALTA	9 días 5 horas	Excede tiempo objetivo
TBG1794	DORADO	7/1/2025 13:48	17/1/2025 15:04	101,27	ALTA	10 días 1 horas	Excede tiempo objetivo
PBI9420	PLATEADO	19/3/2025 10:15	29/3/2025 12:54	102,65	ALTA	10 días 2 horas	Excede tiempo objetivo
PDY7360	PLATEADO	15/2/2025 17:21	26/2/2025 12:34	105,22	ALTA	10 días 19 horas	Excede tiempo objetivo
PDP7631	NEGRO	14/3/2025 17:20	25/3/2025 15:04	107,73	ALTA	10 días 21 horas	Excede tiempo objetivo
PCM3259	VINO	27/2/2025 10:53	10/3/2025 12:41	111,80	ALTA	11 días 1 horas	Excede tiempo objetivo
PDJ8638	PLATA	15/3/2025 16:51	27/3/2025 17:40	120,82	ALTA	12 días 0 horas	Excede tiempo objetivo
PBI9443	PLATA	28/2/2025 11:16	12/3/2025 15:02	123,77	ALTA	12 días 3 horas	Excede tiempo objetivo
PCW3837	NEGRO	25/2/2025 15:04	10/3/2025 9:27	124,38	ALTA	12 días 18 horas	Excede tiempo objetivo
PDA0627	DORADO	10/1/2025 13:03	23/1/2025 12:25	129,37	ALTA	12 días 23 horas	Excede tiempo objetivo
TBE5001	DORADO	18/3/2025 10:45	31/3/2025 12:46	132,02	ALTA	13 días 2 horas	Excede tiempo objetivo
PBO9544	PLATEADO	13/1/2025 8:50	26/1/2025 10:53	132,05	ALTA	13 días 2 horas	Excede tiempo objetivo
PDO1398	ROJO	13/1/2025 15:11	27/1/2025 9:46	134,58	ALTA	13 días 18 horas	Excede tiempo objetivo
TBG5830	DORADO	20/2/2025 17:05	6/3/2025 15:56	138,85	ALTA	13 días 22 horas	Excede tiempo objetivo
HCM0385	DORADO	30/1/2025 17:20	13/2/2025 17:33	140,22	ALTA	14 días 0 horas	Excede tiempo objetivo
PDF3241	PLATEADO	2/1/2025 13:04	17/1/2025 11:41	148,62	ALTA	14 días 22 horas	Excede tiempo objetivo
PCL3689	BLANCO	22/2/2025 15:40	10/3/2025 9:29	153,82	ALTA	15 días 17 horas	Excede tiempo objetivo
PBB8892	PLOMO	25/2/2025 10:23	12/3/2025 15:34	155,18	ALTA	15 días 5 horas	Excede tiempo objetivo
TBE5001	CAFE	17/1/2025 8:35	3/2/2025 11:29	172,90	ALTA	17 días 2 horas	Excede tiempo objetivo
TBH7595	MORADO	14/2/2025 15:50	6/3/2025 14:07	198,28	ALTA	19 días 22 horas	Excede tiempo objetivo
PBL9946	PLATEADO	30/1/2025 9:36	20/2/2025 9:36	210,00	ALTA	21 días 0 horas	Excede tiempo objetivo
PBV8187	PLATEADO	25/1/2025 13:30	19/2/2025 10:02	246,53	ALTA	24 días 20 horas	Excede tiempo objetivo

PCF8225	BLANCO	30/1/2025 9:37	30/1/2025 14:23	4,77	BASICA	0 días 4 horas	Excede tiempo objetivo
PCQ6689	AZUL	24/1/2025 11:23	24/1/2025 16:11	4,80	BASICA	0 días 4 horas	Excede tiempo objetivo
PCZ9460	AZUL	18/1/2025 11:01	18/1/2025 15:55	4,90	BASICA	0 días 4 horas	Excede tiempo objetivo
PDE9377	PLATEADO	18/1/2025 11:01	18/1/2025 15:55	4,90	BASICA	0 días 4 horas	Excede tiempo objetivo
MBF1893	AZUL	12/3/2025 10:29	12/3/2025 15:32	5,05	BASICA	0 días 5 horas	Excede tiempo objetivo
PDL6767	CAFE	12/3/2025 10:04	12/3/2025 15:17	5,22	BASICA	0 días 5 horas	Excede tiempo objetivo
GSI5024	AZUL	21/1/2025 9:45	21/1/2025 15:00	5,25	BASICA	0 días 5 horas	Excede tiempo objetivo
PCP7957	NEGRO	5/2/2025 10:42	5/2/2025 15:58	5,27	BASICA	0 días 5 horas	Excede tiempo objetivo
PDF6267	BLANCO	18/1/2025 10:32	18/1/2025 15:55	5,38	BASICA	0 días 5 horas	Excede tiempo objetivo
PDF5460	PLOMO	24/1/2025 12:29	24/1/2025 17:54	5,42	BASICA	0 días 5 horas	Excede tiempo objetivo
PDF9308	AZUL	31/3/2025 9:54	31/3/2025 15:21	5,45	BASICA	0 días 5 horas	Excede tiempo objetivo
PDJ5358	PLOMO	11/3/2025 11:27	11/3/2025 16:54	5,45	BASICA	0 días 5 horas	Excede tiempo objetivo
PCZ5295	AZUL	12/3/2025 9:59	12/3/2025 15:34	5,58	BASICA	0 días 5 horas	Excede tiempo objetivo
PDK9886	PLOMO	20/3/2025 8:45	20/3/2025 14:20	5,58	BASICA	0 días 5 horas	Excede tiempo objetivo
PBZ4386	BLANCO	27/1/2025 11:09	27/1/2025 16:45	5,60	BASICA	0 días 5 horas	Excede tiempo objetivo
PDD7854	PLOMO	5/2/2025 10:54	5/2/2025 16:34	5,67	BASICA	0 días 5 horas	Excede tiempo objetivo
PCU7900	CAFE	7/3/2025 10:18	7/3/2025 15:59	5,68	BASICA	0 días 5 horas	Excede tiempo objetivo
PDS7109	PLATEADO	10/2/2025 11:50	10/2/2025 17:35	5,75	BASICA	0 días 5 horas	Excede tiempo objetivo
PDJ7441	VINO	11/3/2025 9:45	11/3/2025 15:30	5,75	BASICA	0 días 5 horas	Excede tiempo objetivo
PDK2835	PLATEADO	5/2/2025 14:37	6/2/2025 10:24	5,78	BASICA	0 días 19 horas	Excede tiempo objetivo
PBA3885	ROJO	1/2/2025 8:30	1/2/2025 14:33	6,05	BASICA	0 días 6 horas	Excede tiempo objetivo
PDI5593	PLATEADO	17/1/2025 16:20	18/1/2025 12:25	6,08	BASICA	0 días 20 horas	Excede tiempo objetivo
PDH1680	BLANCO	20/1/2025 14:51	21/1/2025 11:00	6,15	BASICA	0 días 20 horas	Excede tiempo objetivo
PCR9079	VINO	22/1/2025 16:21	23/1/2025 12:30	6,15	BASICA	0 días 20 horas	Excede tiempo objetivo
PDU7134	AZUL	17/1/2025 8:43	17/1/2025 14:54	6,18	BASICA	0 días 6 horas	Excede tiempo objetivo
PBB4194	PLATA	1/2/2025 9:40	1/2/2025 16:00	6,33	BASICA	0 días 6 horas	Excede tiempo objetivo
PDP7475	AZUL	23/1/2025 16:43	24/1/2025 13:04	6,35	BASICA	0 días 20 horas	Excede tiempo objetivo
PDI3817	AZUL	29/3/2025 10:07	29/3/2025 16:28	6,35	BASICA	0 días 6 horas	Excede tiempo objetivo
PDV2948	ROJO	29/1/2025 9:50	29/1/2025 16:15	6,42	BASICA	0 días 6 horas	Excede tiempo objetivo
PCY8135	PLOMO	25/2/2025 9:30	25/2/2025 16:00	6,50	BASICA	0 días 6 horas	Excede tiempo objetivo
PDN3936	PLOMA	8/1/2025 8:50	8/1/2025 15:21	6,52	BASICA	0 días 6 horas	Excede tiempo objetivo
PBK1680	BLANCO	18/2/2025 8:54	18/2/2025 15:26	6,53	BASICA	0 días 6 horas	Excede tiempo objetivo
PDN9467	PLOMO	5/2/2025 16:32	6/2/2025 13:08	6,60	BASICA	0 días 20 horas	Excede tiempo objetivo
PCK2925	PLATEADO	16/1/2025 12:17	17/1/2025 8:55	6,63	BASICA	0 días 20 horas	Excede tiempo objetivo
PCU8786	BLANCO	20/2/2025 9:24	20/2/2025 16:05	6,68	BASICA	0 días 6 horas	Excede tiempo objetivo
PDR5733	AZUL	14/1/2025 10:57	14/1/2025 17:40	6,72	BASICA	0 días 6 horas	Excede tiempo objetivo
PDN7587	PLATA	14/1/2025 10:57	14/1/2025 17:40	6,72	BASICA	0 días 6 horas	Excede tiempo objetivo
ABJ5161	ROJO	14/1/2025 10:57	14/1/2025 17:40	6,72	BASICA	0 días 6 horas	Excede tiempo objetivo
PBU8428	PLATEADO	14/1/2025 10:57	14/1/2025 17:40	6,72	BASICA	0 días 6 horas	Excede tiempo objetivo
PDF6087	PLOMO	7/3/2025 14:23	8/3/2025 11:10	6,78	BASICA	0 días 20 horas	Excede tiempo objetivo
ADX0306	DORADO	19/2/2025 12:01	20/2/2025 8:49	6,80	BASICA	0 días 20 horas	Excede tiempo objetivo
PSO0230	ROJO	10/3/2025 10:05	10/3/2025 16:54	6,82	BASICA	0 días 6 horas	Excede tiempo objetivo
PCS6435	PLATEADO	1/3/2025 9:45	1/3/2025 16:35	6,83	BASICA	0 días 6 horas	Excede tiempo objetivo
PDO3046	AZUL	14/3/2025 9:33	14/3/2025 16:24	6,85	BASICA	0 días 6 horas	Excede tiempo objetivo
PDC7216	NEGRO	20/2/2025 8:41	20/2/2025 15:33	6,87	BASICA	0 días 6 horas	Excede tiempo objetivo
PBB7923	ROJO	24/2/2025 9:10	24/2/2025 16:05	6,92	BASICA	0 días 6 horas	Excede tiempo objetivo
PCR4187	BLANCO	9/1/2025 15:57	10/1/2025 12:54	6,95	BASICA	0 días 20 horas	Excede tiempo objetivo
PDD6578	AZUL	10/2/2025 15:00	11/2/2025 12:00	7,00	BASICA	0 días 21 horas	Excede tiempo objetivo
PDT9829	PLOMO	8/2/2025 9:00	8/2/2025 16:00	7,00	BASICA	0 días 7 horas	Excede tiempo objetivo
GSQ4294	VERDE	23/3/2025 13:05	24/3/2025 10:06	7,02	BASICA	0 días 21 horas	Excede tiempo objetivo
PFH1595	DORADO	30/1/2025 17:51	31/1/2025 14:56	7,08	BASICA	0 días 21 horas	Excede tiempo objetivo
PTO0283	BLANCO	12/3/2025 9:38	12/3/2025 16:48	7,17	BASICA	0 días 7 horas	Excede tiempo objetivo
PCP9087	NEGRO	12/2/2025 8:47	12/2/2025 16:00	7,22	BASICA	0 días 7 horas	Excede tiempo objetivo
PDO4170	PLATEADO	7/3/2025 8:45	7/3/2025 15:59	7,23	BASICA	0 días 7 horas	Excede tiempo objetivo
IBE7448	VINO	10/2/2025 16:33	11/2/2025 13:51	7,30	BASICA	0 días 21 horas	Excede tiempo objetivo
ABK9259	PLATEADO	20/2/2025 8:45	20/2/2025 16:03	7,30	BASICA	0 días 7 horas	Excede tiempo objetivo
TBF4312	NEGRO	31/1/2025 17:27	1/2/2025 14:47	7,33	BASICA	0 días 21 horas	Excede tiempo objetivo
PCB1848	BLANCO	14/1/2025 17:31	15/1/2025 14:54	7,38	BASICA	0 días 21 horas	Excede tiempo objetivo
GCB4683	VINO	21/1/2025 15:54	22/1/2025 13:17	7,38	BASICA	0 días 21 horas	Excede tiempo objetivo
PBE2181	PLATEADO	10/3/2025 8:45	10/3/2025 16:14	7,48	BASICA	0 días 7 horas	Excede tiempo objetivo
PCR5481	PLATEADO	18/3/2025 9:05	18/3/2025 16:36	7,52	BASICA	0 días 7 horas	Excede tiempo objetivo
PQE0383	PLOMO	8/1/2025 9:10	8/1/2025 16:45	7,58	BASICA	0 días 7 horas	Excede tiempo objetivo
MBC5161	NEGRO	21/3/2025 14:29	22/3/2025 12:05	7,60	BASICA	0 días 21 horas	Excede tiempo objetivo

Anexo 2: Banco de preguntas para cada categoría de desperdicio

DESPERDICIO POR SOBREPRODUCCIÓN

DESCRIPCIÓN	SI	NO	CAUSA APARENTE
¿Se genera más producto del que se requiere para el siguiente paso del proceso?		X	No existe compras excesivas de vehículos
¿Fabricamos más producto del que requiere el cliente?		X	No, el cliente interno debe esperar por el producto
¿Mantenemos la maquinaria en funcionamiento incluso cuando no hay demanda?		X	
¿Generamos “trabajo innecesario” para los empleados cuando la demanda disminuye?		X	
¿Se produce más informes de los necesarios?		X	
¿Estamos solicitando más servicios de los que requiere el vehículo?		X	

DESPERDICIO POR DEFECTOS			
DESCRIPCIÓN	**SI**	**NO**	**CAUSA APARENTE**
¿Tiene usted personal capacitado en cada puesto?		X	Falta de capacitaciones continuas y estructuradas
¿Existen métricas estratégicas claras?		X	No existe métricas estandarizadas
¿Todas estas métricas se basan en el rendimiento?		X	No existen métricas de rendimiento
¿Disponemos de métricas de entrada de procesos clave (KPI)?		X	No hay un equipo que estandarice procesos, ni métricas
¿Tenemos métricas de resultados de procesos clave (KPO)?		X	No hay un equipo que estandarice procesos, ni métricas
¿Tenemos métricas durante el periodo del proceso?		X	No hay un equipo que estandarice procesos, ni métricas

DESPERDICIO POR PROCESOS ADICIONALES			
DESCRIPCIÓN	**SI**	**NO**	**CAUSA APARENTE**

¿Se está realizando algún trabajo de reelaboración visible?		X	
¿Medimos la cantidad de retrabajo?	X		Existe un sistema de datos donde se ingresa reprocesos
¿Recopilamos datos sobre mano de obra y materiales asociados con el reproceso?	X		Existe un sistema CRM que permite el ingreso de costos y trabajos realizados.
¿Estamos duplicando informes o información?		X	
¿Estamos realizando más trabajo del que se requiere para ese proceso?		X	
¿Estamos ingresando datos repetitivos?	X		Reingresos de vehículos – sistema de datos sin automatizar

DESPERDICIO POR ESPERA			
DESCRIPCIÓN	**SI**	**NO**	**CAUSA APARENTE**
¿Hay algún trabajo retrasado de un proceso anterior?	X		No concluye la compra de vehículos
¿Tiene usted una carga de trabajo desequilibrada?	X		Vehículos con dificultad alta, son designados a UN solo técnico.
¿Hay escasez de materiales?	X		Jefe de taller no mantiene un stock de insumos apto para el trabajo diario.
¿Tiene usted ausentismo laboral (muy pocos trabajadores)?		X	
¿Hay frecuentes tiempos de inactividad?	X		Falta de comunicación de la cantidad de trabajo
¿Tiene problemas con la calidad del producto en etapas iniciales?	X		Compras de vehículos en pésimo estado mecánico que necesitan reparaciones de dificultad ALTA

DESPERDICIO POR MOVIMIENTOS			
DESCRIPCIÓN	**SI**	**NO**	**CAUSA APARENTE**
¿Tiene usted las cantidades de material adecuadas?		x	No se realiza compras para mantener un stock de insumos constante

¿Están todas las herramientas disponibles?	x		
¿Tienes que buscar archivos en la computadora?		x	No hay información excesiva
¿Están todos los materiales donde se necesitan?		x	No se realiza compras para mantener un stock de insumos constante
¿Estás revisando constantemente los mismos manuales para obtener información?		x	No se posee manuales
Al momento de comprar repuestos ¿siempre es el correcto?		x	Encargado no compra los repuestos correctos

DESPERDICIO POR TRANPORTE			
DESCRIPCIÓN	**SI**	**NO**	**CAUSA APARENTE**
¿Se trasladan materiales entre edificios?	X		Insumos de taller son comprados al momento de necesitar
¿Haces lotes grandes?		x	
¿Compras materias primas a granel?		x	
¿Tiene muchos otros tipos de equipos de transporte?		x	
¿Los materiales se almacenan a largas distancias de donde se utilizan?		x	
¿Existen múltiples áreas de almacenamiento temporal?		x	No son necesarias

DESPERDICIO POR EMPLEADOS SUBUTILIZADOS			
DESCRIPCIÓN	**SI**	**NO**	**CAUSA APARENTE**
¿Conocemos las verdaderas experiencias y capacidades de nuestros empleados?		X	No hay iniciativa del jefe de taller para impulsar a conocer las capacidades del personal
¿Su proceso es tan frágil que los empleados no pueden ser asignados a proyectos especiales?	X		No existe organización y capacitaciones, se asignan a los

			mismos empleados para las mismas tareas
¿Se permite a los empleados experimentar con mejoras de procesos?	X		El personal es abierto al cambio
¿Pueden los empleados ayudar en otras áreas según sea necesario?	X		
¿Los gerentes imponen obstáculos o restricciones a los empleados?	X		El jefe de taller no promueve cultura organizacional de mejora continua
¿Existe una atmósfera de “no se puede hacer”?	x		Por parte del jefe de taller, eso se transmite a su equipo de trabajo

Anexo 3: Base Bibliográfica

ANEXO 1: BERNAL 2012 - MANEJO Y OPTIMIZACION DE LAS OPERACIONES DE MANTENIMIENTO

CAPÍTULO 1

1. ANTECEDENTES.

Actualmente la tecnología permea casi todas las partes de las vidas, desde la compra de alimentos, pasando las visitas al doctor y hasta el mismo salón de belleza o restaurantes. En todas partes se encuentra un software que agilite la administración de los negocios de manera más confiable y controlada.

En este medio los talleres mecánicos pequeños son muy pocos los que cuentan con una debida administración, además de estar equipados con las debidas herramientas técnicas informáticas y un plan o método. Por lo general se encuentra con la administración del maestro que ha sabido llevar adelante a su taller basándose en sus habilidades adquiridas con el pasar de los años, absteniéndose de moverse de su negocio y limitándole el crecimiento del mismo.

De los que utilizan algún medio de control en sus talleres mecánicos se tienen los siguientes como los más comunes: Hojas de Excel, libreta, expedientes, facturas, órdenes de servicio, registros, es decir, la mayoría lo hace todo de manera “manual” Además de que muy pocos talleres cuenten con un plan para la realización de los mantenimientos en el menor tiempo posible. Con esto se observa la poca importancia que se le ha prestado a los talleres automotrices menos aún en el medio, que es un país menos avanzado tecnológicamente a pesar de la importancia que tiene el sector de la industria automotriz y lo desatendido que éste se encuentra.

Problemas actuales de un taller mecánico

Existen algunos tipos de problemas comunes en los talleres mecánicos y entre los que se pudo notar en el taller automotriz Bernal están:

- Control de vehículos ingresados al taller,
- Control de repuestos,
- Inventarios,
- Control de reparaciones (órdenes de servicio),
- Pérdida de tiempo en la ejecución de los mantenimientos,
- Control de trabajos externos realizados,
- Otros.

ANEXO 2: GALO & PASCUA, 2023 - Productivity Improvement in an Automotive Workshop through Lean

Proceedings of the 9th World Congress on Mechanical, Chemical, and Material Engineering (MCM'23)Brunel University, London, United Kingdom - August 06-08, 2023
Paper No. ICMIE 145
DOI: 10.11159/icmie23.145

Productivity Improvement in an Automotive Workshop through Lean Manufacturing Methodology

Geovanna Noemy Galo Bruner[1], Paola Michelle Pascua Cantarero[2]
[1]Universidad Tecnológica Centroamericana/UNITEC Boulevard Kennedy,Tegucigalpa, Honduras geova12@unitec.edu; paola.pascua@unitec.edu.hn[2]Universidad Tecnológica Centroamericana /UNITEC Boulevard Kennedy,Tegucigalpa, Honduras

Abstract – This research study focused on improving the productivity of an automotive workshop in Tegucigalpa, Honduras, by analyzing and identifying the causes of delays and implementing process improvements by applying the tools and principles of the LeanManufacturing methodology using a quantitative approach with a descriptive scope. The problem for the automotive workshop lies in the long maintenance service times for the engine oil change, from the time the vehicle enters the workshop until it is delivered to the customer. The first step was to analyze the company's current situation and identify opportunities for improvement. Analyzing a probability sample of 3 employees in the maintenance department. The activities with the greatest negative impact on the process were analyzed, including unnecessary materials in the work area, the accumulation of waste, unnecessary transfers, and the fact that there is noperson in charge of maintaining order and cleanliness in the area. The operation of the workshop was then described using indicators to improve the overall performance of the company. Seven activities were reduced in the current analysis using process diagrams and flowcharts. The duration of the maintenance service was reduced by 36 minutes, identifying a 40.6% opportunity to improve the serviceand standardize the process. Finally, socialization was carried out with the company, where information was shared about the project andpossible implementation and execution.

Keywords: Productivity, VSM, Systems Improvement, Automotive Shop Floor

1. Introduction

The current environment of the automotive industry in Honduras, Central America, is characterized by competitiveness, speed of change, and instability of demand, which is due to an increase in the number of customers who demand products that meet their requirements and have the quality required to satisfy specific needs.

In Honduras, with the arrival of the Banana Transnationals, most of the elements of modernity were introduced to the country through companies such as the United Fruit Company, the first automobile that arrived in Honduras in 1905 in San Lorenzo [1]. The automobile was increasingly used in countless activities such as war, politics, entertainment, commerce, and of course, as a means of popular communication by families, the introduction of the automobile caused profound transformations in the Honduran society of the twentieth century, in 1910 the first transportation companies were installed in Honduras which were enacted a series of agreements and regulations.

The company where the research was conducted, Tecniwash and Autolote Fredys, located in Tegucigalpa, has observed this problem since its beginnings in the maintenance service. Previously, the company had less demand, but over the years there has been an increase in the number of vehicles in the country and therefore in the city, leading to an increase in demand and a growth in competition, therefore it becomes necessary to propose an improvement in the productivity of the workshop for greater competitiveness in the area. Considering the above-mentioned, the general objective of this research is to propose improvement methods for the productivity of the automotive workshop in the maintenance service using Operations Improvements Systems tools.

2. Methodology

The approach of the study was defined as quantitative because the results can be measured and, in this way, the data processing is carried out. The scope determined for this research is descriptive since the study of the processes of the automotive workshop and the measurement of productivity without manipulation of variables were carried out [2].

The study population was defined as the 3 employees who work in the workshop, specifically in the vehicle maintenance service area, as well as the sample. Their working days are Monday through Saturday from 9:00 am to 5:00 pm. The sampling used is probabilistic because the population that makes up the sample has an equal probability of being selected. The following research variables were considered: time, waiting time, total time, and productivity.

The tools used for the first objective were flowcharts, a process analysis diagram, and a Pareto chart. The process diagram is a visual representation of the operations that take place during the creation and delivery of a product or service and involves the various people responsible for fulfilling the process, it favors the knowledge of the real functioning ofthe process. [3]. The flowchart shows each step, order, and decision in a workflow or process [4]. The Pareto Diagram is a very powerful tool for showing the relative importance of problems [5].

Key performance indicators were used to meet the second objective. These were described to ensure measurement within the automotive shop as they previously did not have these metrics. We used indicators, a method or way that allows us to quantify or measure things, every day in our daily lives.

To propose improvements in the third objective, a VSM of the Current State and a VSM of the Future State were used, as well as a Muda Check to detect all the waste within the process. The VSM is a business method used to evaluate and improve the flow of information and resources needed to create a good or service. It is a planning tool to maximizeresults by eliminating waste [6].

The validation method of this study is carried out by comparing the time obtained in the current situation with the time obtained after the proposed improvement was made using a comparison table. The second verification was carried out through a meeting with the manager of the company under study, where the research was presented to him and possible improvement proposals for the workshop were evaluated. The main goal of validation is to demonstrate that the analytical method is fit for purpose [7].

3. Results and Analysis

A. Analysis of the company's current situation

In order to achieve the first objective of this research, we began with a diagnostic phase in which we observed andstudied the circumstances in which the primary processes of the workshop operate. With the help of some historical data and direct observation in the workshop area, several studies were carried out. With the help of the numerous tools suggested by an operational improvement system, proposals were made based on this knowledge to improve the internal processes of the workshop.

For the diagnostic phase, the direct observation technique was used for several weeks. The workshop performs different activities; among the main ones are vehicle maintenance service, painting, repair, as well as the acquisition, sale, or exchange of second-hand cars. The analysis was carried out specifically in the area of preventive maintenance of engine oil change, therefore this involves the mechanical personnel who perform the operation, the tools, and the materialresources that are used to carry out the service.

The first tool used for this research was the use of a flowchart to graphically represent the process, providing more information about the maintenance tasks of a car.

In order to understand the diagram, brief information about the processes is listed.

-Receiving: The consumer brings his car in the hope that, if the business has the necessary availability, he will be helped right away. If not, he will have to wait for the shop to make it available.

-Mechanical evaluation: Depending on the type of maintenance the vehicle requires; the technician evaluates the unit for the first time to confirm the need for replacement components.

-Unit maintenance: After receiving the customer's consent, the workshop manager assigns a mechanic to service the unit as soon as possible. The unit will be serviced as soon as possible.

-Delivery: Once the maintenance has been completed and the respective service inspections have been carried out,the vehicle is delivered to the customer. It happens that sometimes customers ask for their car to be washed before beingdelivered, however, it was not included in the diagrams since it is not part of the process since it counts as one more service.

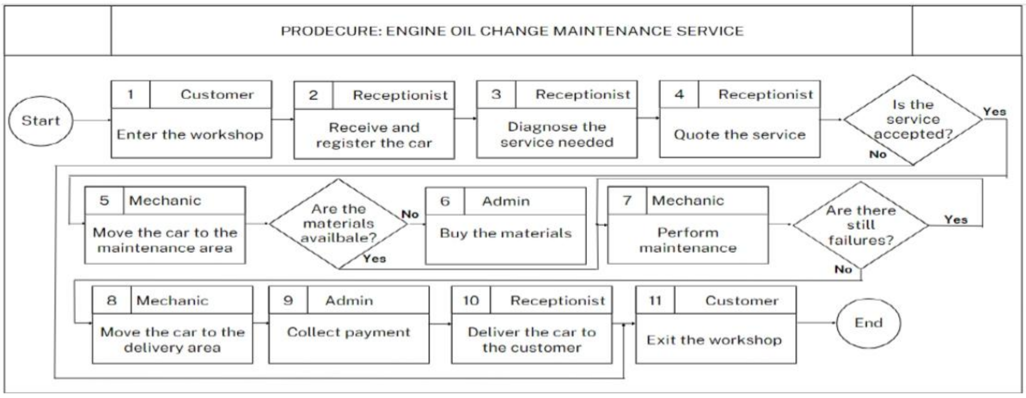


Fig. 1: Flowchart- Maintenance

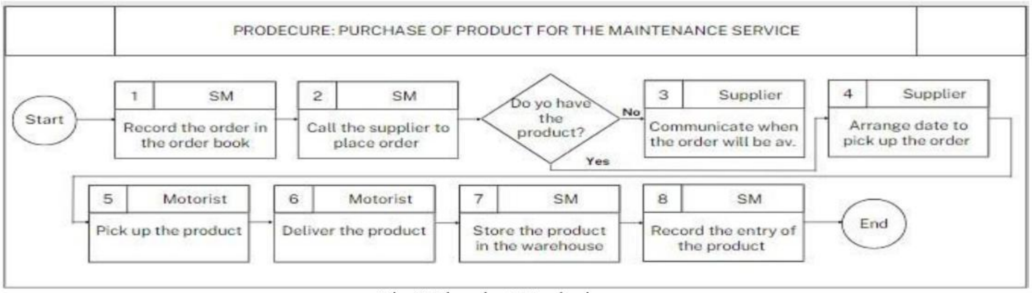


Fig. 2: Flowchart- Purchasing

The sequence of all activities, transports, inspections, delays, and taming that take place in each process are represented in the process diagram. The diagram was created with the intention of making it easy to distinguish between the actions that contribute to engine oil maintenance and those that do not. using a stopwatch and a visualization of the process.

Subprocess	Description	○	□	⊚	D	⇨	▽	Time (sec)
Reception	Receipt of vehicle	1						540
	Simple inspection		1					720
	Diagnosis and quotation	1						240
Maintenance	Vehicle entry to the maintenance area					1		36
	Placement of elevator clamps	1						105
	Lifting the vehicle	1						78
	Vehicle condition check	1						39
	Transfer of burnt oil collector under the vehicle					1		21
	Search for tools					1		135
	Remove oil filter cover	1						35
	Waiting for oil filter drainage			1				76
	Remove crankcase tap	1						12
	Waiting for crankcase tap emptying			1				218
	Place tap	1						7
	Transfer of burnt oil collector to its original place					1		19
	Tighten the tap	1						22
	Look for new filter and oil					1		87
	Open filter and install it	1						19
	Tighten the filter	1						15
	Lower the elevator	1						67
	Uninstall lift clamps on vehicle	1						53
	Open the bonnet and uncover the oil tank.	1						9
	Place the funnel	1						16
	Pour the oil	1						101
	Remove the funnel and cap the tank.	1						9
	Check oil level by starting the car	1						56
	Verify with dipstick	1						41
	Air filter cleaning	1						23
	Returns tools to their place					1		93
	Close bonnet	1						13
	Move vehicle to delivery area					1		139
Delivery	Verify that the service has been performed		1					200
	Invoicing and delivery of vehicle	1						125
	TOTAL							3369

Fig. 3: Oil Change Maintenance Service Analysis Diagram.

The diagram shows the average time taken on days of high demand for the process where it is shown that the engineoil change takes place. It was observed that for the development of the process, there are 22 operations, 7 transfers, 2 inspections, and 2 combined activities. The average service time is 56.15 minutes.

The analysis of the process revealed the following:

- Operating personnel: The company's management encourages teamwork. The personnel are familiar with the company's operations and the functions they perform, identifying their responsibilities and the positions for which they have been hired; however, the mechanics do not use any personal protective equipment. There is also no one responsible for cleaning the workshop.
- Tools: Although the tools they use are in excellent condition, there seems to be no order or designated place for them, and time is wasted when it is generally needed, slowing down the procedures being carried out.
- Methods: Since the status of the vehicle maintenance process is not visualized or regulated from incoming to outgoing, the organization lacks adequate techniques to improve the efficiency of the procedures performed.

In addition, the order of the process flow is not uniform, which occasionally causes bottlenecks for fast service.

- Material: Although the process requires a minimum amount of stock, occasionally work is stopped due to a shortage of supplies.

Overall, with the help of the expert, it was easy to determine that the approach is technically sound, and the observations also show which tasks are redundant and which can be eliminated to improve the method. It is essential to emphasize that the process analysis will make it possible to determine whether changing some activities will increase the productivity of the service by reducing the time required for its creation.

The automotive workshop handles different types of services, depending on the needs of the client and the needs of the unit. According to the Workshop Management Control Board and its evaluation, the most required types of services can be obtained, as shown in the Pareto diagram, it should be noted that for the following diagram, the non-routine services have been eliminated.

Board is from 06/02/2023 to 16/02/2023 were cut as research subjects. The objects that do not change at all the result ofthe most demanded service are:

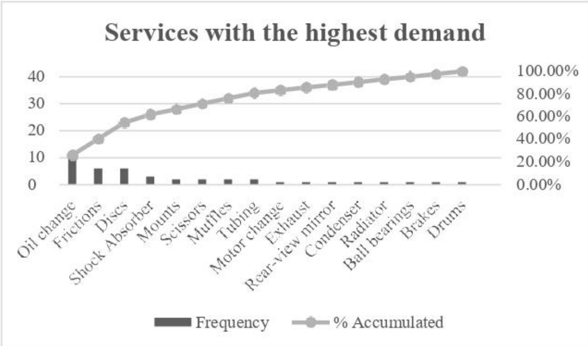


Fig. 4: Pareto diagram analysis of automotive workshop demand.

B. Description of KPIs for the workshop

The automotive workshop did not have established key performance indicators, therefore two key KPIs were proposed for the standardization of the process, such as customer waiting time at the reception, duration of the maintenance service, and the use of cycle times. It was observed that the employees did not pay attention to a sequence of activities, they focus on doing a quality job, but without keeping in mind the customer experience. By establishing indicators in different areas, the automotive shop can identify areas where improvement is needed.

Table 1: Key Performance Indicators

Indicator	Frequency	Formula	Goal
Waiting time at the reception	Every two weeks	Waiting time at reception= (Total waiting timeof customers/total number of customers)	8 min
Maintenance time	Every two weeks	Maintenance time= (total repair time/total number of repairs)	30 min

-Waiting time at the reception desk: the time that customers wait to be served at the business' reception desk was measured. This evaluated the quality of customer service and the competence of the reception staff. In order to establish a KPI for waiting time at the front desk, a specific goal in terms of time was defined. The waiting time at reception can be measured and monitored on a daily, weekly, or monthly basis to evaluate the performance of the business and take measures to improve customer service. This goal was established based on the creation of a prior appointment by the customer with the management for the attention of their vehicle, therefore, there should not be a long wait, an amount of 8 minutes was estimated in case the area is obstructed and proceed to clear the area for the entry of the next vehicle to be attended as scheduled.

Maintenance time: this metric measured the time required to complete a repair from start to finish. It is important to track the time taken for each service to get a clear picture of the company's productivity in this area. A specific goal in terms

of time was used; it is important to remember that the length of service time can vary depending on the type of service provided, the complexity of the service, the number of customers served, the personnel available, and other factors. The goal was set at 30 minutes based on the time of the sum of the service activities studied through the process analysis diagram, considering that some activities are eliminated, and others are improved.
With the proposed inclusion of these metrics, it was possible to reduce the operating times and service indicators, making the process more efficient and transmitting a culture of customer satisfaction.

C. Identification of process improvements.

The value stream map (VSM) was used as a tool to gain an in-depth understanding of processes, both within the company and in the supply chain. The value stream map was developed considering 6 macro-processes carried out in the automotive workshop: Reception, simple inspection, diagnosis, maintenance, payment, and delivery of the vehicle, maintenance, payment, and delivery of the vehicle.

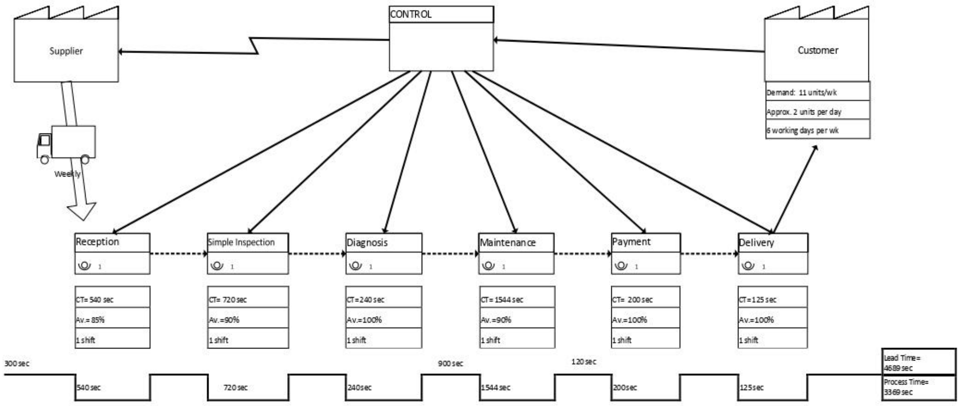


Fig. 5: Current VSM

In the current VSM, a slow, ambiguous process was observed, with many delays and long processes, which is whythe lead time obtained is quite high, 4689 seconds, which is a little more than 78 minutes, and a process time of 3369 seconds, it should be noted that this is an average, and often creates discomfort in the customer since it is a higher timefor the type of service.
Obtaining a Takt Time in minutes for the current flow: 405 min/2 units = 202.5 min.

This indicates that our available hours of work are above the "touch" or what the customer wants in time, which iswhy we must reduce the number of available hours, as it shows that there are many waiting times in the process. It is important to note that the total time available was based on the working day and the demand was taken only based on oil changes, considering that the workshop performs other types of service.
The following observations can be obtained from the VSM diagram:

The cycle time for the first two processes indicated is minimal. At the beginning of the activities, sometimes customers are already queuing outside the workshop to be attended, which varies the service time during the reception of vehicles. On the other hand, three processes (simple inspection, maintenance, and payment) have waiting time because the car must be moved to another area to be attended to. It is proposed to reduce time by redesigning processes and eliminatingactivities that do not add value.

- Dead times were identified within the processes. This is mainly due to time spent searching for tools to start work, the transfer of vehicles between different areas, and time wasted by the workers themselves, among others. For this reason,it is proposed to eliminate waste in the process.
- In the maintenance area, tools, machines, and materials are disorganized. When the vehicles are inspected, the personnel place the parts wherever they see fit, creating a tripping hazard for the workers and increasing maintenance time. In addition, it was detected that the objects are not returned to their respective places after being used, even though there is a shelf to organize the work tools. To avoid this, it is necessary to create a logical sequence of the activities of the process.

It is also proposed to establish the Gemba walk philosophy, in which an employee performs periodic walks in the workplace to observe and keep records of the process times, review the lean aspects, review the standards, and observe the key indicators of the process.

MUDA CHECK

Table 2: 7 wastes of Lean

Wastes	Description
Waiting	Loss of time while waiting for the arrival of material or for the availability of maintenance personnel. maintenance personnel. The supply of tools is slow, so it is possible that some of them are being used by technicians or are in disuse.
Transportation	Unnecessary transfers of the same cars, parts or maintenance personnel do not add value to theservice.
Duplication	More work than necessary, such as performing maintenance services at too short intervals or changing parts that do not yet need to be replaced.
Inventory	In the absence of a safety inventory in the company's warehouse, if any spare part is missing, it will have to be obtained immediately, completely delaying operations.
Motion	Operators often move around the facility sporadically in search of the equipment they need to complete tasks, such as tools or spare parts.
Talent	The company lacks a strategy to capture the inventiveness of its employees. The company's manager usually finds solutions to the problems that arise.
Defects	Performing improper repairs, using low-quality parts, or installing incorrect parts that cause the vehicle to fail and, therefore, require additional repairs.

The workshop has a collector for burned oil; however, for other waste there is no specific place to place the other waste; they use garbage bags. It is suggested to use a container specifically for these toxic materials. Use a suitable, corrosion- resistant container, such as a plastic or metal container, to store burned oil waste without spills, thus avoiding any accidentsor spills in order to avoid any accidents and damage to the environment.

When detecting the moults that influence the process, the most effective measure to improve the process is to eliminate them. This can be achieved by reorganizing the processes, implementing more advanced technologies, optimizing the workflow, reducing waiting times, among other measures.

Once best practices have been identified and moults have been eliminated, work standards can be established to ensure process consistency and quality. It is important that workers are trained in the new processes and work standards. It is also necessary to implement monitoring and control systems to ensure that the process continues to function properly, and that possible future shedding can be detected.

FUTURE VSM

The future state VSM shows how the process will look after implementing the proposed improvements. A waste reduction has been included, thanks to the MUDA Check performed, proposing measures to eliminate or reduce them in the future state. Reduced times in the areas that generate waiting times in the current process, and propose measures to reduce or eliminate them, as well as an improvement in productivity and quality by identifying the areas in which errors or problems are generated and the areas in which rejects, or rework are generated in the current process. A better sequence of the workflow has been proposed to achieve a clear optimization.

Previously there were 6 macro-processes, but by eliminating repetitive operations, unnecessary transfers, rejects, or rework in the current process, a better workflow sequence has been proposed to achieve clear optimization. By eliminating repetitive operations, unnecessary transfers, duplication of activities, inventory errors, idle time, and waste,among others, it was possible to reduce the customer's waiting time when requesting an engine oil change. The changes proposed to improve service productivity are detailed below:

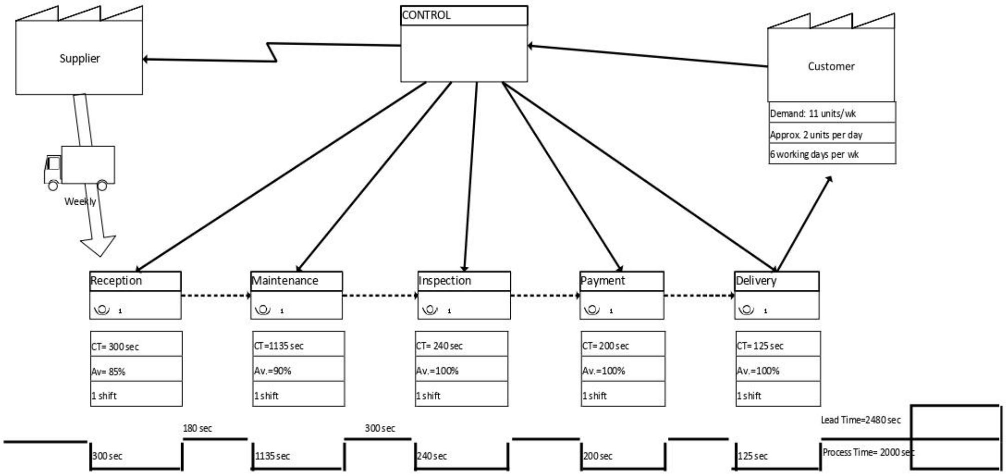


Fig. 6: Future VSM

Vehicle reception: This operation cannot be eliminated; however, it can be improved because the organization must verify the condition in which the vehicle enters the shop. In addition, the workshop manager should assign the work according to the volume of service. To reduce this, organization and planning are necessary; having a well-defined andorganized plan for the reception can help reduce the time it takes to serve customers. For example, having a well- organized reservation system and a checklist for preparing tools and materials before customers arrive. For customers who visit the workshop for the first time and do not have an appointment scheduled, it is important to ask a series of questions at the end of the service that can provide information regarding their satisfaction in the expected time, also mention that for a next visit, you can make an appointment so that the attention is much faster and effective reducing approximately 55% of the time they normally wait.

Simple inspection, diagnosis, and quote: Once the oil change is scheduled, the customer will be able to quote the price of the service through a message or phone call to the workshop, as well as provide details of the car model, type of filter, and the oil he wants for his vehicle in case he has a preference. It was decided to eliminate both operations to reduce these waiting times for the customer.

Maintenance: The operator was asked which activities he considered could be performed in less time, he commented that the installation and uninstallation of the elevator, but this operation cannot be eliminated because, due to health and safety issues at work, the mechanic must ensure that the car is well secured in the elevator and the engine cover. However, 3 activities are eliminated, including one operation and two unnecessary transfers, which are covered by preparing all the necessary tools for the service. The two waits of burn oil were merged with one operation, removing the crankcase tap, because when the mechanic is waiting for the oil to be drained, he can also be performing anotheractivity, either removing the tap or preparing the filter to be installed, and he can even clean his hands with the rag in any case that it has been spilled. There are several operations that are indispensable within the process since they were performed in the correct time, therefore, they were standardized in the same value. It is always recommended that the activities are not carried out with eagerness and that they are performed prudently.

A new activity is integrated at the end of the maintenance, and inspection, the mechanic performs it in two ways, by means of the measuring rod the dipstick, and when starting the vehicle, but this new inspection would be performed after the maintenance has been done so that the management can make sure it is delivering the customer's vehicle in proper conditions,thus avoiding future complaints or claims.

The payment of the service and delivery of the vehicle were standardized as they were in the current situation since normally the customer prefers to pay when the service has been completed and they can see that their vehicle is in good condition, and the keys are handed over so that the customer can leave the workshop.

In the Future VSM, all the proposed improvements were considered, leading to a notable decrease in the oil change, achieving a lead time of 2480 seconds when before it was 4689 seconds, thus increasing productivity by 47.1%; the process time went from 3369 seconds to 2000 seconds, improving by 40.6%. The number of operations was reduced from 22 to 19,from 7 transfers to 4, and one inspection was reduced. The improvement in the reception area was the key to the increase inproductivity, representing a significant time saving, as well as the standardization of the sequence of maintenance steps, alsoallowing the inclusion of activities such as inspections that add value to the service.

D. Validation by time comparison

The research study was facilitated to the management of the automotive workshop, through socialization with the objective of better understanding its operation, organizational culture, processes, and indicators, among other relevant aspects. This type of socialization was carried out in order to obtain valuable information for the research and to enrich the analysis of the results obtained, and an evaluation of the possible suggestions for improvement of the workshop was offered so that it can fulfill its purpose, which is to increase productivity in the workshop. Below is Table 3 as a summary of the before and after.

Table 3: Comparison of the current and improved process.

	Current	Future
Operations	22	19
Inspections	2	1
Transfers	7	4
Process Time	3369 s	2000 s
Lead Time	4689 s	2480 s

Table 3 was discussed in order to be able to state which were the proposals of improvements taken to obtain such data, eliminating 3 operations at the time of providing the service, eliminating the inspections that they made before the process and establishing a new one for the review of the process carried out, the reduction of transfers by avoiding the operator is moving unnecessarily since 2 of the 4 replacements that remained in the improved state include what is the entry and exit of the vehicle which means that the operator as such only moves 2 times for the collection of the burnt oil.

Turnaround times let us know how long it takes to manage the maintenance of a vehicle. By reducing this metric, both in process time and lead time with 1369 seconds and 2209 seconds, respectively, the company will gain greater flexibility and a better ability to cope with higher demand and thus obtain good customer references. The company is considering taking the suggestions into account, as they do not involve an initial cost, but rather a commitment between the company and its employees to comply with them and thus be able to implement the improvement suggestion and prevent future imperfections within the processes.

4. Conclusion

A diagnosis of the company's current situation was made through which the main activities with opportunities for improvement were identified and analyzed, and those with the greatest negative impact on the productivity of the automotive workshop were selected.

Specific measurements were established in front of the team, aligning the individual objectives with those of the workshop, and achieving the company's 2 KPIs, waiting time at reception which would take 8 minutes, and service time which would be 30 minutes. Understanding the automotive shop's objectives around these KPIs is the best way to achieve success in the service department.

A proposal was made to improve the activities that caused delays in the automotive workshop by using the value flow map and waste detection. Seven activities were reduced within the current analysis carried out by using process diagrams and flowcharts. The duration of the maintenance service has been reduced by 36 minutes difference, identifying a 40.6% improvement in the service and achieving standardization of the process.

The analysis in the use of Lean tools was crucial to help drive a culture of continuous improvement in the work teams and maximize available resources. The times of the productive process of the workshop were analyzed by comparing the state the company was in when the data was collected and how implementing the improvement proposal can significantly benefit the company's success.

References

[1] Amaya, J. A. (2010). "LET'S TAKE A LOOK...": A BRIEF HISTORY OF THE AUTOMOBILEOF THE AUTOMOBILE, MOTORING AND COMMUNICATIONS IN HONDURAS COMMUNICATIONS IN HONDURAS: 1905-2005.

[2] Álvarez, C. A. (2011). QUANTITATIVE AND QUALITATIVE RESEARCH METHODOLOGY.QUANTITATIVE AND QUALITATIVE. Retrieved From.

[3] Https://Www.Uv.Mx/Rmipe/Files/2017/02/Guia-Didactica-Metodologia-Dela-Investigacion.Pdfb. Klaus And P. Horn, Robot Vision. Cambridge, MA: MIT Press, 1986.

[4] Roger J. Kremer, D. T. (2007). "The Lean Office Pocket Manual (Lean Office).

[5] PEINADO, J., & GRAEML, A. R. (2007). Production Management: Industrial And Service Operations.Services. Obtained Of Https://Blogdelacalidad.Com/Diagrama-De-Flujo-Flujograma-Deproceso/

[6] Germanova-Kastreva, D. (2020). ANALYSIS OF DEFECTS AND THEIR IMPACT ON THE PRODUCTION LOSSES USING PARETO DIAGRAMS.

[7] Cabrera Valverde, H. S. (2016). Proposal For Quality Improvement Through The Implementation Of LeanService Techniques In The Mechanic Service Area Of An Automotive Company. Lima

[8] Gupta, P. C. (2015). Method Validation Of Analytical Procedures

ANEXO 3: CHEUNG & BAL, 1998 - PROCESS ANALYSIS TECHNIQUES AND TOOLS FOR BUSINESS IMPROVEMENTS.

Process analysis techniques and tools for business improvements

Yen Cheung
Monash University, Clayton, Victoria, Australia and
Jay Bal
University of Warwick, Coventry, UK

Introduction

A methodology for business improvement is really only as good as the tools and techniques that support it. Many companies have adopted a process-oriented view of their operations (Hammer and Champy, 1993), replacing the traditional functional viewpoint, with a view towards gaining benefit by better integration of operations. Many factors can affect the success of this approach, but a key one is the ability to represent and model the process. A range of techniques and tools are available to aid this activity.

Process analysis tools

This section gives a brief description of some of the frequently used tools. These tools are categorised under two headings, paper based software-supported tools and software enabled tools.

Paper based software supported tools

These tools are paper based; some are computer supported to the extent of providing drawing support and consistency checking. There are many tools available in this category, this paper focuses on the ones currently in favour.

The IDEF family

The IDEF (**I**ntegrated Computer Aided Manufacturing **Def**inition) method was defined by the US airforce to aid in its process of supplier development and evaluation (Defining IDEF, 1992). This method involves more than one technique, extending from $IDEF_0$ to $IDEF_6$ and an $IDEF_{1x}$ model. Six variants of the IDEF methodology were defined, for specific purposes. For process capture and analysis the key ones are :

$IDEF_0$

A detailed description of this tool is given in the $IDEF_0$ manual (US Air Force, 1981). The $IDEF_0$ is one of the most widely known tools for functional modelling. It is a top-down hierarchical method, which provides a description of functions and processes in manufacturing. An $IDEF_0$ model consists of an ordered collection of diagrams, text and glossary, all cross-referenced to each other. The graphical representation of $IDEF_0$ is derived, and hence similar to the activity diagram of SADT, as shown in Figure 1. It uses the ICOM (input,

Business Process Management Journal, Vol. 4 No. 4, 1998, pp. 274-290. © MCB University Press, 1463-7154

control, output and mechanism) codes for graphical representation. It can be decomposed indefinitely, depending on the level of detail intended. This top-down breakdown makes $IDEF_0$ a very suitable tool for the visualisation of complex manufacturing systems. Thus $IDEF_0$ provides a structured representation of the functions, information and objects which are interrelated in a manufacturing system (Kusiak *et al.*, 1994).

As there are few symbols in this technique, it should be easy for users to understand how it works. Any manual change in the system requires much effort, as these changes will have to be reflected in all its related nodes. Therefore computer support of this technique is essential. A disadvantage of $IDEF_0$ is that it is not suitable for modelling processes, which have a high frequency of change as the $IDEF_0$ method only provides a static representation of the process.

$IDEF_3$

$IDEF_3$ was created specifically to model a sequence of activities performed. It is a method by which an expert can express his knowledge of a particular process. It is different from the other models as it captures the description of what a system actually does. There are two basic components of the $IDEF_3$ description language:

(1) the process flow description; and

(2) object state transition network description (OSTN). The two components are cross-referenced to build the diagrams.

The process flow description is made up of units of behaviour (UOB), links and junctions boxes. A UOB represents a function or activity occurring in the process. It could be assemble parts, perform test etc.; all these activities can be represented as UOBs. The relationship between UOBs is modelled using three kinds of links: these are precedence links, relational links and object flow links. The precedence links simply indicate the sequence or precedence of the UOBs. The relational link highlights the existence of a relationship between two or more UOBs. The object link provides a mechanism to show the participation of an object between two UOB instances. The process of branching within a process is shown using junctions, the semantics used are *and* (&), or (O) and *exclusive or* (X). In addition the junctions can show if the processes are being

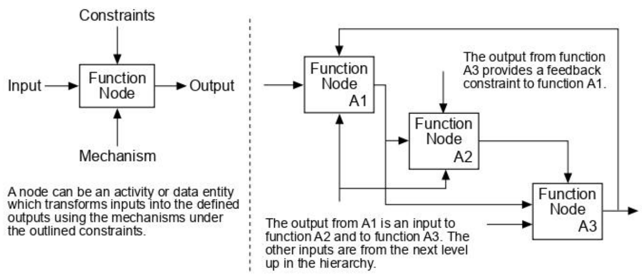


Figure 1.
The $IDEF_0$ diagram

carried out in synchronisation or asynchronisation. Figure 2 shows a process flow description for a simple painting process.

OSTN diagrams indicate the object centred view of a process. Each diagram is used for a single object. The basic elements are nodes, circles and arcs (arrows). The nodes in the diagram indicate the different states of the object. The arcs represent the transition that objects can make. Figure 3 shows an OSTN for the painting process, considering the paint to be an object.

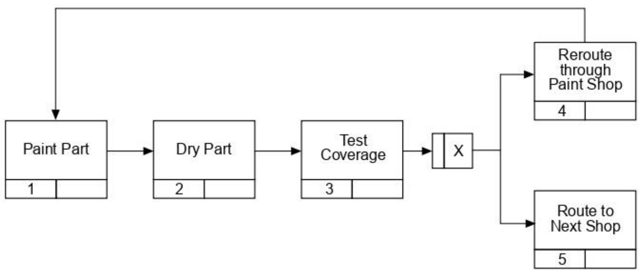
Both process flow diagrams and OSTNs can be decomposed as in $IDEF_0$. Some of the general uses of $IDEF_3$ are:

- Provides a method for recording and analysing data.
- Support system design and design trade-off analysis.
- Can help experts create a resource base and in the development of expert systems.

$IDEF_1$ & $IDEF_1x$

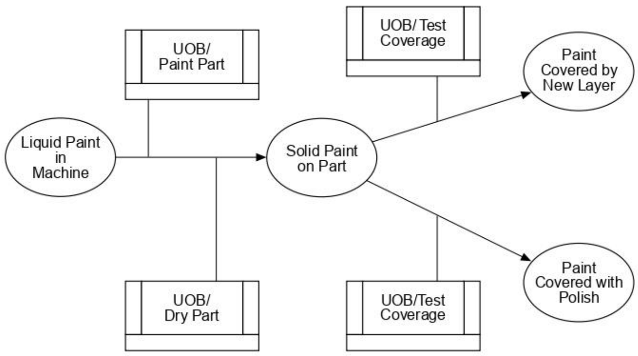
$IDEF_1$ is similar to Entity Relationship Modelling (Marker *et al.*, 1992) but merges this with relational modelling. It was designed as a method that helps organisations to analyse and determine the information resource management needs and requirements, rather than a database design. It captures information in both the automated and non-automated information systems. It can analyse the existing information systems in an organisation, determine the problems and chalk out the information needs of the "to-be" system.

The basic constructs of $IDEF_1$ contain entities, attributes and relationships. The entities are represented by boxes. Characteristics of an entity are represented by attributes within a box. The relationship between two boxes is represented by lines and diamonds, Figure 4. The diamonds at the end provide additional information about the relationship (i.e. if it is cardinal or dependent). The advantages of this information model is that it provides an invariant structure around which databases and applications systems can be designed to handle the constantly changing requirements of manufacturing information.



Source: Arison and Fitzgerald (1988)

Figure 2.
The process flow description



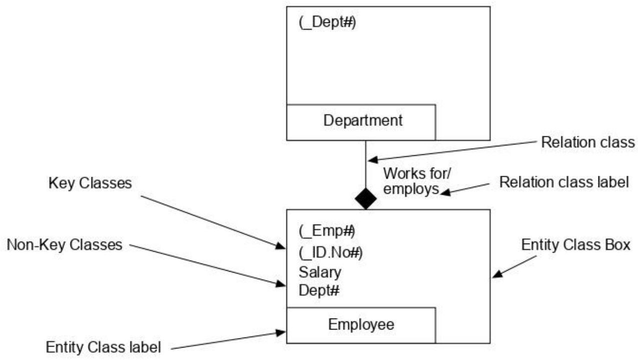
Source: Arison and Fitzgerald (1988)

Figure 3. The object state transition network

$IDEF_1x$ is an extension of $IDEF_1$. It has better graphical representation with enhanced semantic richness and simplified development procedures (The IDEF Home Page, 1997). It is not very effective for doing an "as-is" analysis. It is most useful for logical database design after the information requirements are determined and the decision to implement a relational database has been made.

$IDEF_2$

The $IDEF_2$ method describes the dynamic aspects of a system – the resources used to produce a product, the path an entity can take and the resources needed along the path, also can describe the status of the resources. Thus, it is a



Source: The IDEF Home Page (1997)

Figure 4. $IDEF_1$ diagram

simulation tool. IDEF2 does not provide any specific graphical tools. Instead, it allows the modeller to use language-specific diagrams.

Petri-nets
Petri-nets are a graphical and mathematical tool that can be used to represent procedures, processes, machines and organisations. They can be used to describe, analyse and study various business processes (Van der Aalst and van Hee, 1996). Petri-nets use tokens to reflect the dynamic nature of a process.

A classical Petri net is represented by a bipartite graph. This graph has two kinds of nodes, place and transition. They are represented by circles and rectangles. The nodes are connected by arcs as shown in Figure 5. The marked corner is used to distinguish a transition from a subnet. The figure shows a clerk processing a document. The tokens in *d_in* represent the documents that are to be processed by the clerk. The token in *c_free* indicates that the clerk is free and ready to process the document. Only when all the inputs of a transition have a token is it enabled. Thus the movement of the token through the places via the transition indicates the dynamic nature of the process.

Role activity diagrams (RAD)
RAD shows the dynamics of processes. They are oriented towards the people aspect of a process in relation to the organisation. A RAD shows the roles, their component activities and their interactions, together with external events and the logic which determines the sequence of the activities (Ould, 1995). A role involves a sequence of activities, which is carried out together as a particular responsibility. Figure 6 shows a typical RAD for a simple design process.

The shaded box indicates the role. The arrow shows an external event. An activity is indicated by the black box. The black box can be further decomposed to another RAD. A trigger for initiation of a new role is shown by the crossed box. An interaction between roles is indicated by the connector between two white boxes. The token shown by the diamond is useful in showing the dynamic aspects of the process, it indicates the change of state. RADs have been used

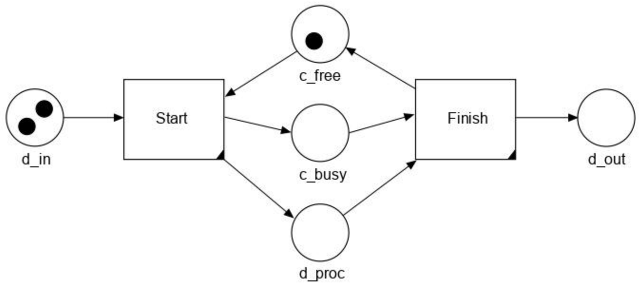
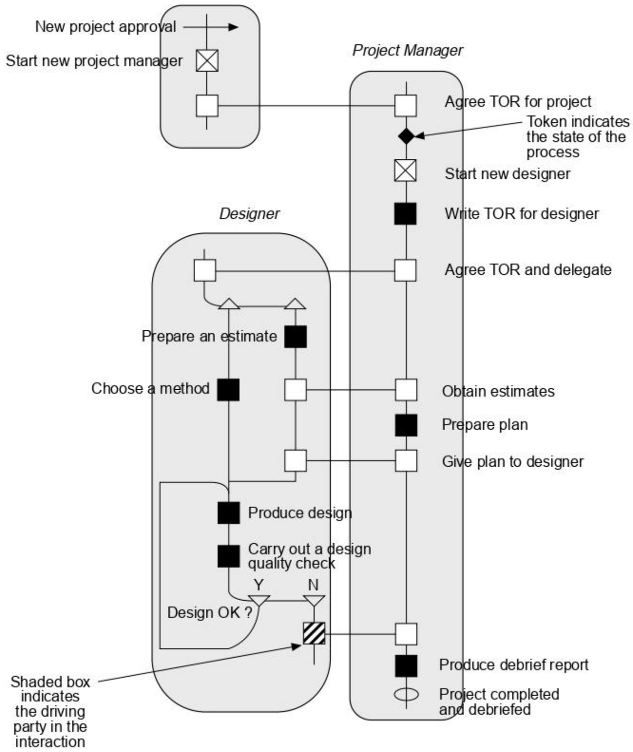


Figure 5.
A high level Petri-net

Source: Van der Aalst and van Hee (1996)



Source: Ould (1995)

Figure 6.
A role activity diagram

effectively in the restructuring of business processes and the analysis of various manufacturing and IT based processes

Conclusions regarding paper based tools
An evident characteristic of the paper-based tools is that they are limited by the medium of representation. Hence they are focused towards modelling only a few specific features of a process. To completely represent a process, many different characteristics such as resources, timing, decisions etc. may need to be captured along with the associated business logic, such as "we only process this on Wednesday mornings unless there is more than 400 units" and What-if analysis is also needed. Computer based tools are more focused towards these requirements.

They can look at the process with respect to key strategic factors such as cost, time and quality. The focus for the rest of the paper is towards software enabled tools.

Software enabled tools

There is a whole range of tools appearing in the marketplace which use computer facilities to aid in process mapping and analysis. These programs generate diagrams that resemble flow charts, but unlike typical flowcharting programs, they let you attach data such as number of units, processes, cost/resources consumed, required time and other information associated with the activity. Table I provides an overview of some of these tools and their functionality.

Every process is unique in its characteristic because of the way in which the people, equipment, tools and practices interact to generate the end result. Thus, the choice of a process analysis tool depends on the characteristics of the process investigated. The more complex the process the greater the demands on the tool.

We chose the product introduction process as a benchmark.

Product introduction process

The product introduction process of companies from the automotive industry was studied for determining the characteristics of this process. Figure 7 ideally indicates the structure that most companies are trying to achieve. Although the PIP of each of the companies varied in structure, some common process characteristics were clearly identified. These are:

- *Concurrency/overlap:* The activities within the process were carried out concurrently. Figure 7 ideally indicates the kind of concurrency seen.
- *Interdependency:* The different sub-processes working concurrently, constantly shared information.
- *Complexity:* The process indicated a high degree of complexity. The advancement in various engineering disciplines and their interaction indicated a complexity in technology and human interaction.
- *Innovativeness:* The degree of newness of the product development was another determining factor. A product that is new-to-the-world varied in the degree of uncertainty from a product, which is an improvement to an existing line of products.

Based on the identified characteristics a limited sample of process analysis tools was tested to determine their suitability for analysis and modelling of the product introduction process. Two tools, Octave and Process Charter were chosen and studied.

Comparison methodology

Based on the capabilities of currently available tools and the identified process characteristics, six categories were seen as essential in a process analysis tool (see Table II). This also includes factors, as identified in a previous research done by Bradley *et al.* (1995) and Hansen (1994).

These are:

Software tool	Description
ABC Flow Charter 6 The Roderick Manhattan Group London, UK	It provides 12 shape templates for different job classifications. Jobs such as quality control, network management and auditing. Contains a special feature known as Sensible Lines. These lines can be attached to certain shapes or can be totally independent. It also features Sensible Text, lets you type even if the shapes are too small. It hides the text. It includes a word processor. Supports colour printing
Business Process Generator Quondam, Dorringon, UK	The program lets you define business logic in an easy to understand English syntax. It can then, if required generate 3GL code for clients and servers in mixed client server environments. Complex business processes can be built and they can be linked to a user interface of your choice and also to Relational Database Management System tools
BPWin Logic Works Inc. Princeton, NJ, USA www.logicworks.com	It helps analyse and reengineer business processes. Describes business processes in terms of inputs (items required by the process) and outputs (items produced by the process). The effect of the information on the outcomes can be seen as well as the flow. You can determine whether processes would benefit from reengineering and by what degree. Utilises the IDEFO process modelling standard. It basically allows the analysis of complex business processes and uncovers potential areas for process improvement and redesign
FlowModel 2.0 Systemstar SoftTools	A charting tool that not only allows the creation of charts and diagrams but lets you integrate data into the diagrams. Offers ten graph templates with pick and place symbols. Data can be given to each of the symbols. The symbols and data can be exported to other programs if required, e.g. spreadsheets and databases. You can also define your own symbol if it is not already present
Octave Process Manager TOP-IX Business Systems Wigan, Lancashire, UK www.metabpr.com	Octave provides a centralised repository as a database of best practices. The specific users can model their respective processes and submit the changes or new models to the centre. The software provides a mechanism for modelling a process from a business level to an operational level (hand and eye motion). It provides analysis reports indicating the value-added content, also indicates the least cost path, longest path etc. of a process. It also has another module called TOPMAN which is useful for staff resource planning preparing activity based cost budgets and enables performance reporting
ProSLCSE Process Simulator International Software Systems Incorporated Austin, Texas, USA www.issi.com	The ProSIMULATOR tool in the ProSLCSE product line is a powerful graphical tool that can be used for understanding, debugging and refining the process. It can play a critical role in assessing a new process before implementation. Beneficial for training staff in new or revised processes

(*Continued*)

Table I.
The software enabled tools

Software tool	Description
Process Model ProMODEL Corporation Orem, Utah, USA www.processmodel.com	It is a flowcharting and simulation tool that provides businesses with a means to diagram, analyse and improve their processes. It has some key features like live animation, Onestep modelling and Visual Staffing. The key advantage of this tool is the visualisation it provides
Workflow Analyser Meta Software Corp. Cambridge, Massachusetts, USA www.metasoftware.com	This tool addresses the entire BPR life cycle, including data capture. process modelling, simulation, implementation and continuous improvement. Graphical language is used to express complex data sets pertaining to budgets, staffing and equipment requirements. The software enables users to test assumptions, analyse alternatives and measure results
SIMPROCESS CACI Arlington, VA, USA www.caci.com	This simulation tool is designed for business process modelling and analysis. When used in conjunction with the company's object-oriented simulation languages, SIMPROCESS can help reduce the time spent on mapping reengineering components
ReThink Gensym Corporation. Cambridge, Massachusetts, USA www.gensym.com	Gensym Corp. in Cambridge, MA, has combined object-oriented technology with interactive graphics to produce a BPR tool that provides user-friendly modelling and simulation. The software helps users monitor process performance and manage real-time operations
Extend+BPR Imagine That Inc. San Jose California, USA www.imaginethatinc.com	This software package includes 90 pre-built blocks to help users create reengineering models. The program, which supports five types of operating systems, features drag-and-drop modelling, animation, spreadsheet connectivity and customised reporting
Optima! AdvanEdge Technologies Inc. Tualatin, OR, USA www.advanedge.com	This process improvement software is an easy-to-use Windows application that features process modelling, simulation and reporting capabilities. The program, designed for the front end of reengineering projects, helps users quickly create and edit presentation-quality process map
SA/BPR Professional Popkin Software & Systems Inc. New York, NY, USA www.popkin.com	Popkin Software's comprehensive integrated tool supports the Federal Information Processing Standards for function and data modelling. The package analyses what controls the execution of a function, what performs the function, and what objects or data are produced by the function. It features a built-in reporting language with a graphical-user interface for creating customised reports
COOL– Gen/ Composer 4.0 Sterling Software. www.cool.sterling.com	This tool enables organisations to use model-driven development to rapidly design, build, test, install and maintain reengineering applications. Composer supports variety of computer platforms and databases
Framework 5.1 Ptech. Cambridge, Massachusetts, USA www.ptechinc.com	This offers an integrated set of object oriented tools that enable users to create interactive blueprints of business processes. Software code can be generated from the hierarchical layout, providing rapid and consistent application development

Table I.

(*Continued*)

Software tool	Description
FirstSTEP Interfacing Technologies Corp. Quebec, Canada www.interfacing.com	This business-process modelling and simulation too incorporates object-oriented technology. The software provides reporting and analysis on static and dynamic states of BPR models. Also included are export and import capabilities, in addition to workflow connectivity
SAP R/3[1] SAP Aktiengesellschaft, Systems, Applications and Products in Data Processing, Neurottstrasse 16, 69190 Walldorf, Germany www.sap.com.au	The Reference Model in this package is a methodology for assisting SAP implementation. It covers the conceptual design stage through to actual operation of the final system. It provides both the process view and the functional view of the business with a graphical drill down facility

Note: [1]Owing to the rising popularity of SAP R/3, it has been included in this paper. Users select from a generic pool of 800 processes to model their own processes, which are to be implemented in the SAP R/3 environment. In this case, users select the system before modelling the process(es)

Table I.

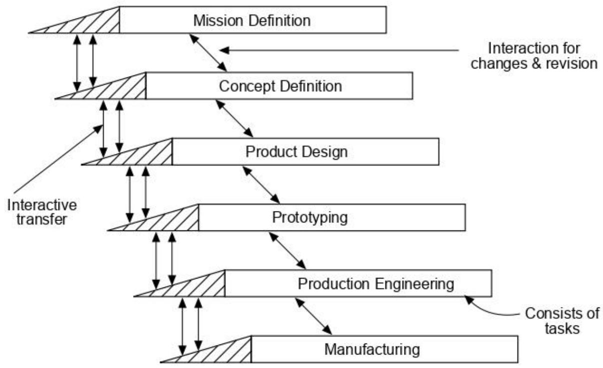


Figure 7.
Overlap between activities

(1) hardware and software features;
(2) user features;
(3) modelling capabilities;
(4) simulation capabilities;
(5) analysis capabilities;
(6) integration capabilities.

Each of these categories and their associated factors is described in more detail below:

Table II.
Feature comparison chart: comparison of tools

SL. No.	Categories	The factors	Octave process manager	Process charter
		Platform and operating system	DOS and Windows NT, 3.1, 95	DOS and Windows NT, 3.1, 95
1	Hardware and software	Database repository and client server	Yes, also has remote operation and systems mode	No
		External links	Yes, has OLE link	Yes, but no OLE link
		Tool integration	No	No
		User friendly	Average, has poor help features	Good, has good help features
		GUI	Yes	Yes
2.	User	Online documentation and help	Average	Good
	features	Special features	• Data dictionary • Shows critical paths like longest time, highest cost etc. • Considers the entity owners and other organisational issues	• Resource sharing • Provides different templates, e.g. ISO 9000, $IDEF_0$ etc. • Multiple triggers, so two independent processes can be run on the same sheet
		Goal centred	No	No
		Object-orientation	Yes	Yes
		Customisation of blocks	No	Yes
		Roles	No	No
		Data flow	No	Yes

(*Continued*)

SL. No.	Categories	The factors	Octave process manager	Process charter
3.	Modelling capability	Front end analysis	Yes	Yes
		Ease of modelling	Average	Good
		Concurrency	Average	Poor
		Interdependency	Poor	Not available
		Level of detail	High for tasks. But poor in terms of resource and time allocation	Average for activity levels, but good for resource and time allocation
		Time	No	Yes
		Cost	No	Yes
4.	Simulation capability	Discrete event modelling	No	Yes
		Statistical distributions	No	Yes
		Animation/graphical capability	No	Yes
		Reasoning	None	None
		Output analysis	Poor	Average
5.	Analytical capability	BPR expertise	None	None
		What-if? analysis	Poor	Average
		Top-down. bottom-up analysis	Yes	Yes
		Outside-in analysis	No	No
		Integration of views	No	No
6.	Integration	Business to physical link	No	No

Table II.

(1) *Hardware and software*

- *Platforms and operating systems:* the wider the number of platforms and operating systems the software can work on, the better it is. This is because different computer systems might be used by the people who work across the globe on the same development project.
- *Database repository and client-server:* if the software has a mechanism for storing the information in a common repository. This way all the information can be accessed from a single source and can be shared across the company. This is very useful for a development project, where the work is normally done in different places. By having a common repository and networking capability, the problem of redundancy is avoided.
- *External links:* does the software have export-import features? Can word processing tools, spreadsheet tools, reporting tools etc. be linked to it?
- *Tool integration:* does the tool have a capability to be integrated with other tools or vice versa? It could have one or more modes of integration, such as it can be integrated with other point-solution tools, have its own integration environment or have a open architecture so that it can be incorporated in an integration framework.

(2) *User features*

This is to indicate the various features that each tool has. This will also help identify the ones that are good for a PIP process:

- *User friendly:* if the tool is user friendly the input data can easily be entered and also the output is presented in a way that is easily understood by the user.
- *Graphical user interface(GUI):* does the software have a graphical user interface so that the required information or data can be entered through several windows?
- *Online documentation and help:* the user should be able to get help on a specific feature on the click of a button.
- *Special features:* the number of special features that improves the capability and user friendliness.

(3) *Modelling capabilities*

- *Goal centred:* this means the process or re-engineering goals are identified first and then the re-engineering or improvement is carried out in agreement with the goals.
- *Roles:* it determines if the software tool can identify the different roles in the process. Roles can be the activities as performed by a person. Hence this is very useful in re-engineering.
- *Data flow:* does the software use data flow diagrams, IDEF etc. to model the process?

- *Front-end analysis:* can the tool check the model for syntax and logic? This way the model can be corrected before it is run.
- *Ease of modelling:* how easy is it to model the process, in terms of the graphical drawing and the data entry?
- *Level of detail:* how much detail can the model be drawn to? This indicates the accuracy of the model.
- *Concurrency and interdependency:* this is very important for PIP, as a number of activities in PIP are operating concurrently and there is constant exchange of information between the activities.
- *Object-orientation:* can we assign attributes to the various objects defined in a process? Attribute, here, refers to the distinguishing characteristics of an object, such as height, size, cost, time scale etc.
- *Customisation of blocks:* a tool cannot have all the iconic blocks needed for analysis and graphical presentation; hence, does the tool have an open architecture so that user customised icons can be created?

(4) *Simulation capabilities*
This is to determine if the software has any simulation capability. Simulation is an effective means of analysing a process. The resource constraints, idle time, delays etc. can be identified by simulating a process iteratively. The factors considered are:

- *Time:* this feature identifies if the tool can track time. This is very useful in determining the duration of a "to-be" operation from start to finish.
- *Cost:* can the tool track the cost changes as it is simulated?
- *Distributions:* can statistical distributions be used to trigger the input or to simulate activities? This will help determine the variation in the behaviour of an activity under different situations.
- *Animation/graphical capability:* a visual representation of a tool is very effective in analysis.
- *Discrete event modelling:* most activities do not occur in a fixed predictable period of time. Thus an effective tool should be able to vary the timing of events, tasks and processes to generate a realistic picture.

(5) *Analysis capability*
The analytical capability in a tool is vital for finding means for improvement. Some of the factors considered are:

- *Reasoning:* can the software reason with the information supplied and take decisions accordingly? The reasoning would be done through a knowledge shell or expert system stored in the tool.
- *Output analysis:* can the tool analyse the output for a process run? It refers to the ability of the tool to present the output in an understandable and effective format and draw conclusions.

- *BPR expertise:* does the tool have any BPR expertise in relation to the factors identified by Hammer *et al.*, such as determining decision points, tracing the information source in a process, posing questions for dramatic improvement etc?
- *What-if? analysis:* does the tool provide a means for simulating and considering the effect of different scenarios?
- *Top-down:* bottom-up and outside-in analysis, does the tool have the flexibility to map the process from bottom-up and top-down? Can the effect of changes in external environment, such as technology etc. be considered?

(6) *Integration of views*

Does the tool have an enterprise perspective? An important aspect of enterprise integration is the ability to look at the process from different views: the information view, behavioural view, organisational view, decisional view etc. The information view can help in designing an appropriate information system tailored to the requirement. Similarly the decision view can help in determining if the decision-making power is in the right hands. Following are some of the factors:

- *Integration of views:* can the user look at the process from a data view and shift to an activity view? Does the tool consider the effect of interrelationship between the organisation and information etc?
- *Business level to physical level link*: can the tool link to the business goals, thus help in re-engineering the strategic issues of the process? Can the tool link to the physical systems in the firm, wherein the improvements suggested from the IT perspective are directly translated into the systems? Does the tool have the flexibility and intelligence to make improvements from the strategic to the tactical level of the business?

Discussions

The analysis indicated that most process analysis tools were more suitable for operational than strategic processes. They usually lacked the ability to model complex processes like the product introduction process. However, Octave Process Manager was found to be more suitable for organisational process management and resource planning. It ensures that multiple process definitions are not unnecessarily generated and much re-use of process primitives can be made. The semantics library and client-server based technology can prove effective for a geographically dispersed process like the product introduction process. Process charter is a PC based tool and its simulation capability is very effective in studying the process efficiency, resource constraints etc. Hence it can be used as an effective tool for workflow analysis. The simulation capability also aids in doing scenario and what-if analysis. Octave really emerges as a tool for actively managing processes for TQM activities, whereas Process Charter is

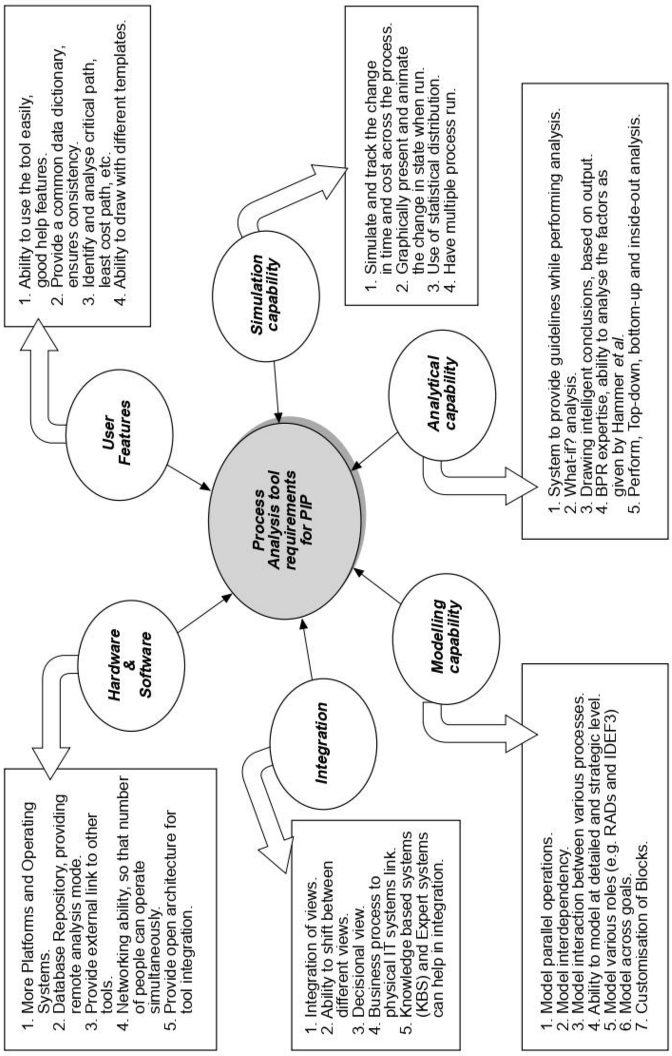


Figure 8.
The identified
requirements

much easier to use and more focused towards sub-process redesign and resource evaluation.

Both are limited in describing timing events and synchronising parallel linked activities.

Based on the limitations of the sample of tools analysed, Figure 8 identifies the requirements a process analysis tool should have to effectively study and re-engineer the product introduction process.

Conclusion

Process analysis tools are needed for business improvement projects such as business process re-engineering. A variety of such tools exists and software enabled ones are better than paper based ones, especially when modelling processes are in dynamic situations where changes often occur. General considerations for selecting a process analysis tool fall into six broad categories, i.e. hardware and software features, user features, modelling capabilities, simulation capabilities, analysis capabilities and integration capabilities. An emerging IT management practice where the system is determined before business analysis prevents integration problems but this may be restrictive where innovation is involved as in the product introduction process.

References

Avison, D.E. and Fitzgerald, G. (1988), *Information Systems Development,* Blackwell Scientific Publications, Oxford.

Bradley, P., Browne, J., Jackson, S. and Jagdev, H. (1995), "Business process re-engineering (BPR) – a study of the software tools currently available", *Computers in Industry*, Vol. 25, pp. 309-30.

"Defining IDEF" (1992), *Manufacturing Systems*, September, p. 36.

Hammer, M. and Champy, J. (1993), *Reengineering the Corporation: A Manifesto for Business Revolution,* HarperCollins Publishers, New York, NY.

Hansen, G.A. (1994), "Tools in business process re-engineering", *IEEE Software*, September, Vol. 11 No. 5, pp. 131-3.

Kusiak, A., Larson, T.N. and Wang, J.R. (1994), "Reengineering of design and manufacturing processes", *Computer and Industrial Engineering*, Vol. 26 No. 3, pp. 521-36.

Maker, R.J., Cullinane, T.P., deWitte, P.S., Knappenberger, W.B., Perakath, B.A. and Wells, M.S. (1992), "IDEF3 – process description capture method report", *Information Integration for Concurrent Engineering* (IICE), Armstrong Laboratory, Wright-Patterson AFB, OH.

Ould, M.A. (1995), *Business Processes – Modelling and Analysis for Reengineering and Improvement,* Wiley & Sons, New York, NY.

The IDEF home page (1997), Taken from, http://WWW.idef.com on 28 July.

US Air Force (1981), ICAM Architecture, Part II, Vol. IV, "Function modelling manual (IDEF0)", Wright – Patterson Air Force Base, OH.

Van der Aalst, W.M.P. and van Hee, K.M. (1996), "Business process redesign: a Petri-net based approach", *Computers in Industry*, Vol. 29, pp. 15-26.

Further reading

Carr, D.K. and Johansson, H.J. (1995), *Best Practices in Reengineering* McGraw-Hill, New York, NY.

Davenport, T. and Short, J. (1990), "The new industrial engineering information technology and business process redesign", *Sloan Management Review,* Summer, pp. 11-27.

Fitzgerald, B. and Murphy, C. (1994), *BPR: Myth and Reality*, Kogan Page, London.

ANEXO 4: GADATSCH, 2023 - ANALYSIS, MODELLING, OPTIMISATION AND CONTROLLING OF PROCESSES

1 Introduction to Business Process Management

Process management changes the world of work

Abstract

This introductory chapter first explains the concept and historical development of business process management. Subsequently, several basic concepts such as “function”, “business process”, “process”, “end-to-end process” and “workflow” are defined and distinguished from each other. The conclusion is formed by review questions and an exercise.

1.1 Concept Clarification

Why do we need business process management? This question is not only asked by students of business administration who go to a corresponding lecture with great expectations, but also by experienced practitioners. A look at history can help here a little bit. Since the beginning of the 19th century, the world of work has been characterized by a strong division of labor as a result of the previous industrial revolution. The *Taylorism* played an important role here, named after the US American Frederick W. Taylor (see the original work Taylor 1903).

The *business process management* (GPM) or simply *process management* was developed at the beginning of the 1990s in order to, among other things, eliminate the negative consequences of division of labor and poor coordination.

Business process management deals with the documentation, analysis and restructuring of workflows (processes). For a long time, the term “process organization” was common in German-language literature. The documentation of the processes is also referred

 1
A. Gadatsch, *Business Process Management*,
https://doi.org/10.1007/978-3-658-41584-6_1

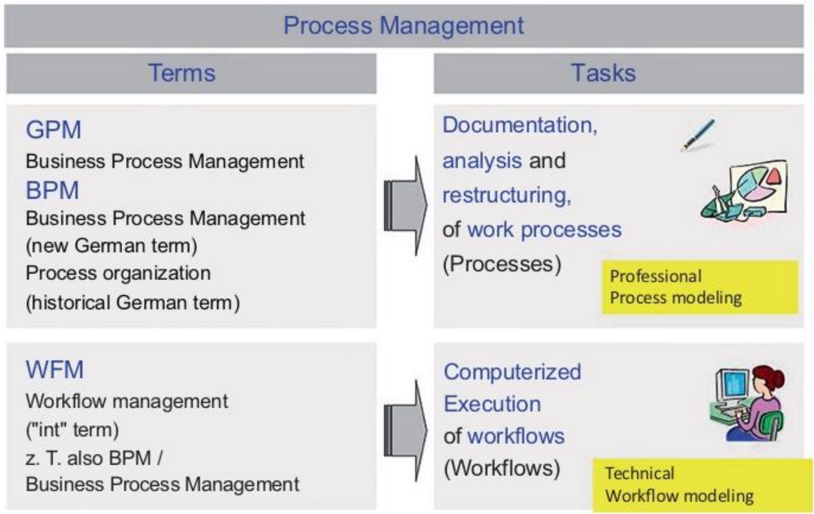


Fig. 1.1 Concept clarification process management

to as “technical process modeling”. In the international environment, the term “business process management (BPM)” is common.

To be distinguished from this is the term “workflow management” (WFM), which covers the computer-supported execution of workflows (so-called “workflows”). Here one also speaks of “technical workflow modeling”. In international usage, the terms “business process management and workflow management” are often not further differentiated, one usually speaks in both cases of “business process management (BPM)”. Figure 1.1 shows the terminology at a glance.

1.2 Historical Development

In the development of process management, four phases of development can be identified (see Table 1.1).

I. Phase: Breakdown of work into functions (Taylorism):

The early phase of process management begins with Taylorism, named after Frederic Winslow Taylor (1856–1915), who consistently separated planning and executive activities. This classical bureaucratic organizational structure prevailed in almost all companies of the 19th century and played a central role for departments (cf. Sua-Ngam-Iam and Kühl 2021, p. 46). According to the then prevailing business paradigm, the construction and operational organization were considered separately in this phase. This was first

DE GRUYTER OPEN

ECONOMIC THEMES (2018) 56(1): 57-78

DOI 10.2478/ethemes-2018-0004

IDENTIFICATION AND ANALYSIS OF KEY BUSINESS PROCESS MANAGEMENT FACTORS

Aleksandra Stoiljković Ranđelović

PhD student, University of Niš, Faculty of Economics, Republic of Serbia

✉ *aleksandras.randjelovic@gmail.com*

Radenko Milojević

University of Niš, Faculty of Economics, Republic of Serbia

✉ *radenko.milojevic@eknfak.ni.ac.rs*

Marija Radosavljević

University of Niš, Faculty of Economics, Republic of Serbia

✉ *marija.radosavljevic@eknfak.ni.ac.rs*

UDC 005.83

Original scientific paper

Abstract: Under the conditions of ever-increasing competition in the market, companies can reach a competitive advantage only with flexible offer of quality superior and cheaper products/services. In order to achieve this, companies must carry out innovative and effective business processes and manage them appropriately. In this regard, it is important to identify the factors that may be considered critical for improving business process management. The aim of this paper is to identify and analyse key business process management factors, on the example of the electronic industry. The research results show that the factors that are usually accepted as the most important for increasing the maturity of business process management are not recognised as dominant in the electronic industry. One of the findings of the survey reveals, as the main lagging factors, four of the six most important factors for the successful implementation of business processes. Based on results of the research, Managing employees has been detected as the most important maturity factor for companies from the electronic industry; hence, the suggestion is further improvement within this field.

Received: 02.02.2018
Accepted: 02.04.2018

Keywords: process orientation, business process management, maturity factors, and electronic industry

JEL classification: L25, M12

1. Introduction

The frequent changes that occur as the result of a high degree of adaptability of business systems, lead to the creation of new or modified organisational structures of high flexibility driven by users' requirements (Adamides & Karacapilidis, 2006). Usually, changes include process orientation. In process-oriented companies the focus has changed from business functions to business processes, with emphasised relation with customers and suppliers (Andersen, 1999). It is important to emphasise that process orientation is not synonymous with process organisational structure (Bosilj Vukšić et al., 2006). It represents an understanding the flow of business and is only the first step towards a process organisational structure.

Changes in the market indicate the need that companies constantly analyse and improve business processes, but also to find ways for efficient coordination of integrated business processes, providing at the same time value for customers and rational consumption of resources. Business Process Management (BPM) is a management discipline aimed at describing and managing the business processes in an organisation (Koster, 2009). The goal of BPM is for the organisations' objectives to be achieved by aligning the business processes with the strategy and to continually improve them.

The implementation of BPM is a very complex and time-consuming process that requires great effort, resources, discipline and commitment. Many companies have tried to change their business in compliance with a process orientation, but only few have managed to completely integrate their business functions into end-to-end processes (Bandara et al., 2009). Consequently, many BPM projects are unsuccessful in practice (Trkman, 2009). At the same time, successful implementation is a prerequisite for the benefits of this concept (Štemberger et al., 2009).

Since BPM is a multidisciplinary concept, its success depends on different factors (Bandara et al., 2009). Many research studies, aiming to find and to explore the success and failure factors of BPM implementation, have already been conducted, but the area remains an interesting research topic (Alibabaei et al., 2010).

In order to implement BPM properly, it is necessary to identify key processes and key success factors. This is the precondition for creating the matrix of processes and factors, which can help managers to identify the influence of factors to the processes. For this reason, the identification of BPM factors deserves special attention of managers, and especially process owners. Companies that have adopted process approach become aware of the power of their processes and the importance of providing the quality of their products, which, together with continuous improvement, ultimately leads to increased satisfaction of all stakeholder groups, but also to the success of the organisation (Chen, 2001).

The study presented in this paper aims to explore the state of business processes management maturity of the companies in the electronic industry. The purpose of this paper is to identify and analyse the achieved level of development of the maturity factors, on the example of electronic industry.

2. Business process management conditionality: Factors' identification

In line with changes in the environment, there are also changes in managing, and one of very important changes in the last few decades concerns business processes. Their importance has been changing during the last few decades, but it is certainly that doing business in the 21st century will not be possible without process orientation. As Harmon (2014) noted, all contemporary organisational structures, to a greater or lesser extent, emphasise the importance of business processes. Business process management is becoming an important part of organisations' operations (Lahajnar & Rožanec, 2016). Jeston and Nelis (2014) define BPM as achieving goals of an organisation through management, improvement, and control of essential business processes. According to Weske (2007) business process management includes concepts, methods, and techniques to support the design, administration, configuration, enactment, and analysis of business processes.

As part of these changes, it is possible to identify factors that influence the implementation of business process management. Dealing with this issue for some time, authors concluded that the most organisations identify around 7 key business processes and between 6 and 16 factors, critical for business process management. Many researchers are focusing on the six key factors of BPM, which include strategic alignment, culture and leadership, people, governance, methods and information technology (Fisher, 2004; Melenovsky & Sinur, 2006). Bandara et al. (2009) introduce nine success factors, namely: culture, leadership, communication, information technology, strategic alignment, people, project management, performance measurement and methodology. Ravesteyn and Batenburg (2010) conducted a survey among 39 Dutch consultants, developers and end-users of BPM-systems and found that communication, involving the right people in the project and making sure that there is support from top management and governance are the critical factors.

It is important to say that organisations should consider different factors as well as the links between them, rather than putting emphasis on only some of them. One study (Bai & Sarkis, 2013) found a number of direct and indirect relationships among the factors. According to the results of this study the four most important factors include: strategic alignment, top management support, project management and collaborative environment. Based on the studies of others authors in the field of process management (Jarrar et al., 2000; Ravesteyn & Versendaal, 2007; Trkman, 2010; Vom Brocke & Rosemann, 2010; Jeston & Nelis, 2014), but also on the basis of the authors' experience, 16 business process management factors have been identified, and they are:

- Strategic alignment,
- IT investment,
- Process measurement,
- Managing employees,
- Organisational changes,
- Appointing process owners,
- Managers' commitment,
- Continuous improvement,
- Business culture,
- Information technology,
- Methods,
- Employees' specialisation,
- Focus on customers and their requirements,
- Standardisation of processes,
- Implementation of proposed changes,
- Cooperation with suppliers.

Strategic alignment – According to Rosemann and de Bruin (2006), strategic alignment of BPM has to enhance the linkages that connect organisational priorities and company processes in order to achieve its business objectives. Although business process management assumes focus on realisation and improvement of everyday activities, it does not mean that it is separated from the company's strategy. On the contrary, business process management has to be balanced with the strategy, since otherwise the efficiency of business processes will not lead to the accomplishment of company's objectives and it will be useless. Therefore, some authors consider the relationship between strategy and operations function as crucial (Rhee & Mehra, 2006).

IT investment – Performing tasks in today, dynamic environment cannot be imagined without information technology. For this reason, investment in IT can be observed as a necessary condition for providing the competitive advantage, in the same way as other non-material resources are. The required level of IT investment depends on company's strategy, other organisational resources, which interact with IT and on the external environment (Duh et al., 2006; Melville et al., 2004). Certainly, IT technology is not enough by itself for gaining the competitive advantage, but it can contribute certainly to the improvement of the internal environment where all resources necessary for process realisation are connected and balanced and where managers must reengineer their core processes from a customer perspective (Terziovski et al., 2003).

Process measurement – According to process approach, it is very important to define critical control points during the process realisation, at least as important as to measure the characteristics of the process output, after the process has been finished. This claim comes from the fact that measurement during the process realisation can help managers to identify the mistakes and errors while process still lasts and to correct them before additional resources are used for their further processing. In the meantime, it is important that the obtained measurement results are used as feedback for improving process efficiency (Hassan, 2009).

Managing employees – Contemporary business is characterised by the perception of employees as the most valuable assets of the company. Starting from the fact that their knowledge, skills, communication largely determine the success or failure of the company, the management of employees is the most important and the heaviest part of the work of the managers. Empowering of employees and appreciating their ideas and suggestions are significant motivational tools, which are often crucial for the successful business process management. Empowering employees to make decisions independently can help make operations easier and faster (Trkman & McCormack, 2010).

Organisational changes – Usually authors who research the field of process management agree that it is not necessary to make changes in the organisational structure. But, potential problems with an inadequate organisation are reflected in the increase in costs, inconsistency in the execution of functional decisions between processes, and general decline in the efficiency (Silvestro & Westley, 2002). However, although the organisational units may still be separated formally, their employees have to make mutual communication more intensive in order to avoid the silo effects. This is in line with the attitude of Leavitt (2005), according to whom hierarchies must be changed so that employees perform their tasks more efficient.

Appointment of process owners – Identification of process owner is one of the most important issues of the implementation of the process approach. The existence of the process owner represents most noticeable difference between a process company and a traditional one (Hammer & Stanton, 1999). This is also the issue critical for providing the link between the organisational units that perform activities included in the same process. The process owner is in charge for taking care of the process from the beginning till its end and for accomplishing the defined objectives. According to Hammer and Stanton (1999), the owner of the process must have a constant role, real responsibility and authority in making decisions on process design, performance measurement, as well as training and improvement of employees who perform them.

Managers' commitment – Although the implementation of business process management often requires the engagement of appropriate professionals, the support and commitment of managers is very important. Responsibility of engaged experts cannot be delegated, so managers have to be completely involved in process management.

Continuous improvement – Striving to perfection has to be a light motive for all companies wishing to sustain achieved comparative advantage or trying to reach it.

This factor is emphasized as very important in the most of process management models. It assumes process observation, measurement and analysis, and based on the facts, improvement of process aspects that are main limitation for increasing process efficiency or effectiveness. Continuous improvements are based on encouraging the organisational culture and formal structure (Trkman, 2010). The top management must be the main support while process owners should be the main drivers for popularisation of continuous improvement (Savolainen, 1999).

Business culture – Business culture is a manifestation of attitudes, values and traditions in a company. Culture is often referred to as a key driver of BPM initiatives (Rosemann & vom Brocke, 2010), representing both a source of failure and success in BPM initiatives (Melenovsky & Sinur, 2006). In line with the fact that in the modern business environment the only constant is change, continuous adjustment to changes, or initiating changes to improve business processes must become an everyday task of all employees (vom Brocke & Schmiedel, 2011).

Information technology – Information technology includes the software, hardware and information management systems that support the process activities (Melenovsky & Sinur, 2006). Contemporary companies are aware of the importance of IT and do not ask the question of whether it is profitable to implement IT, but are in the dilemma about the technology they should choose. There are an increasing number of software packages related to business process management. However, it has to be emphasized that IT itself cannot bring any competitive advantages (Terziovski et al., 2003).

Methods – Rosemann & vom Brocke (2010) defined methods, in the context of BPM, as the set of tools and techniques that support and enable activities along the process lifecycle and within enterprise-wide BPM initiatives. Different kinds of methodologies have been developed to help organisations manage their business processes. Some can be used in any organisation. For example, many organisations have tried some process improvement methods (Six Sigma, Lean…) or new technology, such as Business Activity Monitoring, Service Oriented Architectures (Garimella et al., 2008).

Employees' specialisation - It is clear that employees' specialisation is very important for increasing process efficiency, because they can perform their tasks faster and more accurately. However, narrow specialisation may be an obstacle for flexibility, which is also important for providing sustainable competitiveness in modern conditions. For those reasons, it is necessary to provide optimal balance between employees that are specialist and those who are generalists (Mulyar & van der Aalst, 2005), primary bearing in mind processes and activities critical for providing customers' satisfaction (Reijers & Mansar, 2005).

Focus on customers and their requirements – Connection with customers is reflected in need to identify the ones that represent the target group for the company and then to define and specify their needs. This kind of specification is base for process standardisation and setting up the measurement system. Especially are important critical to quality customers' requirements, because the company has to provide them if it does not want to disappoint its customers. Customers' satisfaction is precondition for their loyalty, so the companies' task is to continuously analyse the level of their satisfaction and their suggestions for the improvement. In fact, customers play a crucial role both at the beginning and at the end of a business process, providing valuable input for the business process and assessing the quality of the output (vom Brocke & Schmiedel, 2011).

Standardisation of processes – Standardisation is a key condition for providing the reliability of the process; It assumes identification of the activities and tasks and resources for their performing, providing at the same time, their optimal combination and consistency in their usage. If it is based on customers' demands, then standardisation may be observed as a condition for providing quality of final outputs. Given that processes are carried out with the respect of specifications and rules, this means that BPM systems are positively connected to standardisation (Küng & Hagen, 2007). However, sometimes standardisation leads to impairment of innovation and performance. In such processes, companies should avoid setting rigorous rules (Benner & Tushman, 2003; Hall & Johnson, 2009).

Implementation of proposed changes – Very important factor of business process management success is company's capability to implement necessary changes. For the success of this process, it is necessary to provide support from the employees, the ones that may be considered as change agents and who will act as promoters of the change and advocates of the process manager. Also, middle management is significant to help employees in transition (Herzig & Jimmieson, 2006).

Cooperation with suppliers – Bearing in mind that quality of inputs determines the quality of processes realisation and, consequently, the quality of outputs, the relationship with suppliers is equally important as relationship with customers. For that reason, it is necessary that company demands from its suppliers to adopt the same business philosophy and to implement the same process approach. In this way, company extends the process management philosophy into supply chain, extending, in that way, the base for sources of competitive advantage. There are different forms of connection from a competitive relationship to a partnership relationship. Today, companies are increasingly cooperating with their suppliers and include them in the development of new products (Cantista & Tylecote, 2008; Valle & Vazquez Bustelo, 2009). Inclusion of suppliers into innovation process is especially important today when customers' demands are constantly changing and when the product life is shortened. Therefore, the role of suppliers is important in ability of company to innovate (Roberts, 2001).

One of the most explored and analysed model for business process maturity management - model formulated by Rosemann and de Bruin (2005), based on the Delphi study, suggests that the *critical factors* for the successful implementation of business processes, are: strategic alignment, process management, employee management, methods, information technology, and culture. The maturity of process management is determined exactly by the quality and the presence of these factors (Radosavljevic, 2015). Those six factors will be used for a comparison with the most significant factors of business process management maturity in the research presented in this paper.

3. Research framework

Previously introduced factors represent the framework for the researching conditionality of the business process management in the Republic of Serbia. The electronic industry has been selected as a basis for this research because it's increased growth due to the effects of the globalisation process and the inflow of foreign capital into our country, as well as its contribution to the economy. This study aims to explore the state of business processes management maturity of the companies in the electronic industry and the achieved level of development of the maturity factors. The research was conducted in November 2017.The research involved data collection including survey and interviews. The survey questionnaire was used as a research technique.

The objective of the research has been identification of the key factors of business process management maturity in the Republic of Serbia. Therefore, the research has been led by the following hypotheses, formulated based on the research of other authors and the pre-research conducted in September 2017:

- H1: Factors that are stated as *critical* for increasing business process management maturity are recognised as the most important in the electronic industry, based on the observed sample,
- H2: The main lagging factors are Managing employees and Information technology,
- H3: All of the observed factors influence business process management maturity.

The questionnaire that has used consists of two parts. Within the first part of the questionnaire, the questions aimed to collect the basic data about the companies and the information about the interviewed managers (the function of the interviewed manager, the gander, the years of age, and the years of work in business process management ...).

The second part of the questionnaire included two segments. The objective of the first segment was to determine, based on the respondents' answers, in which state - silos, tactical integration, process orientation, optimized company, intelligent network, the companies were. The second segment consists the questions that aim to identify the maturity level of each of the business process management factors (Strategic alignment, IT investment, Process measurement

(Process management), Managing employees, Organisational changes, Appointing process owners, Managers' commitment, Continuous improvement, Business culture, Information technology, Methods, Employees' specialisation, Focus on customers and their requirements, Standardisation of processes, Implementation of proposed changes, Cooperation with suppliers). Based on the processing of these data, one can notice the level of development of each factor individually, and determine whether some of the factors lag behind in comparison with others. By revealing factors that are at a lower level of maturity compared to others, appears the need to put emphasis on their correction and improvement.

The data were collected by surveying the top managers of the companies. For the purpose of analysis, a scale ranging from 1 to 5 has been used, where 1 means that the claim is completely incorrect, while 5 means that the claim is completely correct.

The data collected were processed with the software SPSS (Statistical Package for the Social Sciences). In order to test presented hypotheses, the authors used descriptive statistics, as well as regression analysis, correlation analysis and cluster analysis.

4. Research results and discussion

The research of the achieved maturity level, conducted in previous years (Radosavljevic, 2015), suggests the use of the following interval groups, which represent different levels of maturity:

- Up to 2.50 – the second level of maturity,
- From 2.51 to 3.50 - the third level of maturity,
- From 3.51 to 4.50 – the fourth level of maturity and
- Over 4.51 – the fifth level of maturity.

Starting from such interval groups, the factors' maturity level can be determined. For that cause, the collected data from the questionnaire are processed and the results of descriptive statistics are shown in Table 1.

According to the obtained results, it may be said that all factors are at the fourth level of development. However, this result has to be accepted with reserve, due to the (positive) subjectivity of the interviewed managers.

The obtained average marks and accepted interval groups, show that there is no factor which has reached the fifth level of maturity. The most developed maturity factors of the observed companies are Standardisation of processes (4.4815), Continuous improvement (4.3704) and Employees' specialisation (4.2963). Higher level of development also exists for factors such as Methods (4.2593), Cooperation with suppliers (4.2222) and Managing employees (4.1481). Comparing the factors that are most often mentioned as the dominant with most developed factors in the observed companies, it is discovered that the only ones that appear few times are factors Methods and Managing employees only. The factors that are usually accepted as the most important for increasing business process management maturity are not recognised as dominant in the electronic industry. In this sense, it can be said that *the first hypothesis should be rejected.*

Table 1. Descriptive statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Strategic alignment	27	1.00	5.00	3.5926	1.11835
IT investment	27	2.00	5.00	4.0370	.80773
Process measurement (Process management)	27	2.00	5.00	3.8519	1.09908
Managing employees	27	2.00	5.00	4.1481	.94883
Organisational changes	27	2.00	5.00	4.0741	.99715
Appointing process owners	27	1.00	5.00	3.5926	1.21716
Managers' commitment	27	2.00	5.00	3.6296	.96668
Continuous improvement	27	3.00	5.00	4.3704	.74152
Business culture	27	2.00	5.00	3.8148	.92141
Information technology	27	2.00	5.00	3.7778	.93370
Methods	27	2.00	5.00	4.2593	.94432
Employees' specialisation	27	3.00	5.00	4.2963	.82345
Focus on customers and their requirements	27	2.00	5.00	4.0370	.89792
Standardisation of processes	27	2.00	5.00	4.4815	.89315
Implementation of proposed changes	27	2.00	5.00	4.0461	.8423
Cooperation with suppliers	27	1.00	5.00	4.2222	1.01274

Source: Authors' calculation

There are few main lagging factors which should be improved and developed. Based on results the factors least developed are Strategic alignment (3.5926), Appointment of process owners (3.5926), Managers commitment (3.6296), Information technology (3.7778), Business culture (3.8148) and Process management (3.8519). Those results show that for further increasing of process management maturity very important things are missing and they are strategic approach to process orientation, intensive usage of IT, business culture change and improvement of process management. Bearing in mind that those four factors are stated as four of six the most important factors it may be said that they will represent the significant limitation for process orientation spreading through the observed companies. Also, it can be noticed that the main lagging factor in the electronic industry are not Managing employees, on the contrary, this is one of the six dominant factors. But, maturity factor Information technology is one of the factors weakest developed in the observed companies. *It means the second hypothesis is partially accepted.*

Table 2: Comparison of business process management maturity factors

Critical factors for the successful implementation of business processes (Rosemann & de Bruin, 2005)	The most developed factors of business process management maturity in researched companies	Main lagging factors of business process management maturity in researched companies	Factors with greatest significance individual contribution to business process management maturity
Strategic alignment	Standardisation of processes	Strategic alignment	Managing employees
Process management	Continuous improvement	Appointment of process owners	Appointing process owners
Employee management	Employees' specialisation	Managers' commitment	Strategic alignment
Methods	Methods	Information technology	Business culture
Information technology	Cooperation with suppliers	Business culture	Managers commitment
Culture	Managing employees	Process measurement (Process management)	Focus on customers and their requirements

Source: Authors' calculation

Through usage of regression analysis, the influence of all of the observed factors to business process management maturity may be observed. The values of 16 factors were included in the model as independent variables. The business process management maturity was taken as a dependent variable. The beta coefficients show the influence each dimension has on the model, while significance has to be below the 0.05. The null hypothesis assumes that all of the observed factors influence business process management maturity, while the alternative assumes that there is not influence of all of the observed factors to business process management maturity.

Starting from the value of the beta coefficient we compare the contribution of all maturity factors and found that the greatest individual contribution to business process management maturity has the factor Managing employees (beta = 2.613). It means that Managing employees has the greatest individually contribution to the explanation of the business process management maturity, when subtract variance explained by all other maturity factors. Also, two more factors, belonging to the group of six dominant factors according to the Rosemann and de Bruin model, have significant individual contribution to business process management maturity (Strategic alignment (beta =1.426) and Business culture (beta = 1.028)). Also, we see that the factors which have the greatest individual contribution to the business processes maturity, at the same time, they are the main lagging maturity factors in observed companies.

Based on the observation of the table, it can be concluded that several business process management factors (IT investment, Process measurement (Process management), Methods, Employees' specialisation, etc.) have low, insignificant predicting power for the business process management maturity. These factors have significance level higher than 0.05. Hence, it is important to emphasize that the obtained results can be the result of overlapping with other independent variables in the model. Based on the above, the null hypothesis is rejected, and the alternative is accepted. In this way *third hypothesis is rejected*, meaning that all of the observed factors do not influence business process management maturity.

Table 3. Regression analysis

Variables / Factors	**Unstandardised Coefficients**		**Standardised Coefficients**	**t**	**Sig.**
	B	Std. Error	Beta		
Strategic alignment	1.001	.374	1.426	2.676	.023
IT investment	.427	.339	.440	1.261	.236
Process measurement (Process management)	.276	.252	.386	1.094	.299
Managing employees	2.160	.751	2.613	2.878	.016
Organisational changes	-1.282	.435	-1.629	-2.950	.015
Appointing process owners	.974	.422	1.512	2.312	.043
Managers' commitment	.767	.399	.946	1.923	.013
Continuous improvement	.521	.271	.493	1.922	.014
Business culture	.875	.298	1.028	2.939	.015
IT	.435	.157	.518	2.764	.020
Methods	-.032	.325	-.038	-.098	.924
Employees' specialisation	.234	.276	.246	.849	.416
Focus on customers and their requirements	.685	.348	.784	1.966	.028
Standardisation of processes	.273	.298	.310	.916	.381
Implementation of proposed changes	.334	.359	.404	.931	.374
Cooperation with suppliers	.111	.214	.143	.519	.615
Dependent variable: business process management maturity					

Source: Authors' calculation

Based on the correlation analysis, strength and direction of the connection between maturity factors was examined. According to the results of the correlation analysis, between some factors there is a strong positive correlation.

Table 4. Correlation analysis of the process management maturity factors

		SA	ITI	PPM	ME	OC	APO	MC	CI	BC	IT	M	ES	FCR	SP	IPC	CWS
SA[*]	Correlation Coefficient	1.000															
	Sig. (2-tailed)	.															
ITI	Correlation Coefficient	.607**	1.000														
	Sig. (2-tailed)	.001	.														
PPM	Correlation Coefficient	.603**	.436*	1.000													
	Sig. (2-tailed)	.001	.023	.													
ME	Correlation Coefficient	.200	.328	.691**	1.000												
	Sig. (2-tailed)	.316	.095	.000	.												
OC	Correlation Coefficient	.108	.258	.573**	.888**	1.000											
	Sig. (2-tailed)	.591	.193	.002	.000	.											
APO	Correlation Coefficient	.400*	.217	.715**	.612**	.455*	1.000										
	Sig. (2-tailed)	.039	.277	.000	.001	.017	.										
MC	Correlation Coefficient	.038	.313	.351	.524**	.554**	.615**	1.000									
	Sig. (2-tailed)	.851	.112	.072	.005	.003	.001	.									
CI	Correlation Coefficient	.196	.314	.675**	.828**	.708**	.671**	.520**	1.000								
	Sig. (2-tailed)	.328	.111	.000	.000	.000	.000	.005	.								
BC	Correlation Coefficient	.231	.204	.469*	.744**	.604**	.463*	.448*	.570**	1.000							

[*] Explanation of the used symbols: SA-Strategic alignment, ITI-IT investment, PPM-Process measurement (Process management), ME-Managing employees, OC-Organisational changes, APO-Appointment of process owners, MC-Managers' commitment, CI-Continuous improvement, BC-Business culture, IT-Information technology, M-Methods, ES-Employees' specialisation, FCR-Focus on customers and their requirements, SP-Standardisation of processes, IPC-Implementation of proposed changes, CWS-Cooperation with suppliers.

		SA	ITI	PPM	ME	OC	APO	MC	CI	BC	IT	M	ES	FCR	SP	IPC	CWS
	Sig. (2-tailed)	.246	.308	.014	.000	.001	.015	.019	.002	.							
IT	Correlation Coefficient	.319	.329	.231	.377	.413*	.040	.098	.157	.213	1.000						
	Sig. (2-tailed)	.105	.094	.246	.053	.032	.842	.626	.433	.286	.						
M	Correlation Coefficient	.248	.021	.545**	.669**	.667**	.411*	.393*	.602**	.549**	.255	1.000					
	Sig. (2-tailed)	.212	.916	.003	.000	.000	.033	.043	.001	.003	.200	.					
ES	Correlation Coefficient	.333	.244	.582**	.490**	.522**	.420*	.419*	.371	.271	.392*	.659**	1.000				
	Sig. (2-tailed)	.090	.219	.001	.009	.005	.029	.030	.057	.171	.043	.000	.				
FCR	Correlation Coefficient	.398*	.321	.428*	.432*	.405*	.376	.537**	.344	.227	.477*	.613**	.716**	1.000			
	Sig. (2-tailed)	.040	.103	.026	.024	.036	.053	.004	.079	.255	.012	.001	.000	.			
SP	Correlation Coefficient	-.042	-.188	.365	.490**	.603**	.373	.567**	.358	.420*	.188	.729**	.558**	.557**	1.000		
	Sig. (2-tailed)	.837	.347	.061	.009	.001	.055	.002	.067	.029	.348	.000	.002	.003	.		
IPC	Correlation Coefficient	.079	-.065	.397*	.507**	.624**	.527**	.624**	.384*	.303	.385*	.574**	.648**	.651**	.791**	1.000	
	Sig. (2-tailed)	.696	.749	.040	.007	.001	.005	.000	.048	.124	.047	.002	.000	.000	.000	.	
CWS	Correlation Coefficient	.117	.125	.508**	.530**	.607**	.477*	.367	.412*	.112	.426*	.436*	.532**	.503**	.454*	.637**	1.000
	Sig. (2-tailed)	.563	.535	.007	.004	.001	.012	.060	.033	.577	.027	.023	.004	.007	.017	.000	.

Correlation matrix:
* Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

Source: Authors' calculation

More specifically, the results show that there is strongest statistically significant correlation between factors Organisational changes and Managing employees (correlation coefficient is 0.888). Based on calculated correlation coefficients observed existence of strong correlation between factors Continuous improvement and Managing employees (correlation coefficient is 0.828) and between Business culture and Managing employees (correlation coefficient is 0.744) It means that an increase in the maturity level of Managing employees is followed by an increase in the maturity levels of Organisational changes, Continuous improvement and Business culture. Also, concluded that Standardisation of processes strong positive correlated with factors Methods (correlation coefficient is 0.729) and Implementation of proposed changes (correlation coefficient is. 0.791).

Connection between the observed business process management factors and, therefore, possibility to use one of them as a driver for increasing the maturity level of the others, can also be seen from the dendrogram, presented in Figure 2.

Figure 2: Connection between the observed maturity factors[*]

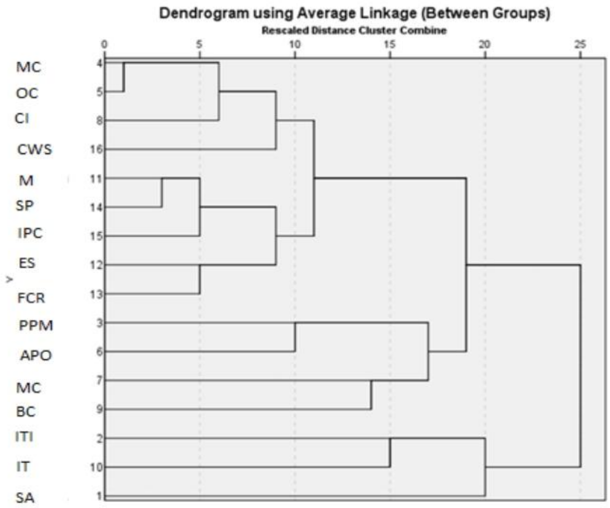


Figure 2 shows that the closest ones are variables (those variables are connected more than the others): Managing employees and Organisational changes, Methods and Standardisation of processes and Employees' specialisation and Focus on customers and their requirements.

In order to find drivers to achieve a higher level of business process management maturity, it is important to examine the correlation between maturity factors and business process management maturity. Based on results presented in Table 5, it can be said that there is the statistically significant strong correlation between Information technology and business process management maturity (correlation coefficient is 0.743).

[*] The explanation of the used symbols was given earlier

Table 5. Correlation analysis of the maturity factors and business process management maturity

Strategic alignment	Correlation Coefficient	.503**
	Sig. (2-tailed)	.007
IT investment	Correlation Coefficient	.429*
	Sig. (2-tailed)	.026
Process measurement (Process management)	Correlation Coefficient	.453*
	Sig. (2-tailed)	.018
Managing employees	Correlation Coefficient	.511**
	Sig. (2-tailed)	.006
Organisational changes	Correlation Coefficient	.430*
	Sig. (2-tailed)	.025
Appointing process owners	Correlation Coefficient	.272
	Sig. (2-tailed)	.170
Managers' commitment	Correlation Coefficient	.148
	Sig. (2-tailed)	.461
Continuous improvement	Correlation Coefficient	.471*
	Sig. (2-tailed)	.013
Business culture	Correlation Coefficient	.252
	Sig. (2-tailed)	.205
Information technology	Correlation Coefficient	.743**
	Sig. (2-tailed)	.000
Methods	Correlation Coefficient	.423*
	Sig. (2-tailed)	.028
Employees' specialisation	Correlation Coefficient	.423*
	Sig. (2-tailed)	.028
Focus on customers and their requirements	Correlation Coefficient	.478*
	Sig. (2-tailed)	.012
Standardisation of processes	Correlation Coefficient	.169
	Sig. (2-tailed)	.398
Implementation of proposed changes	Correlation Coefficient	.375
	Sig. (2-tailed)	.054
Cooperation with suppliers	Correlation Coefficient	.429*
	Sig. (2-tailed)	.026

Correlation matrix:
** Correlation is significant at the 0.05 level (2-tailed)*
*** Correlation is significant at the 0.01 level (2-tailed)*

Source: Authors' calculation

Compared to the others, higher correlation coefficients with business process management maturity can be found for Managing employees (correlation coefficient is 0.511) and Strategic alignment (correlation coefficient is 0.503). For the factors that have the significance level higher than 0.05, it can be concluded that there is no correlation between them and business process management maturity.

Conclusion

For many organisations, improving business process management capabilities is one of the main topics. The achieved level of process maturity is conditioned by the development of the dominant factors of business process management maturity. In that sense, research of the achieved factors' business process management maturity is important.

The factors that are emphasised as the *critical* for increasing business process management maturity (Rosemann & de Bruin, 2005) are not recognised as dominant in the presented research results. Based on those results, it is found that the most developed business process management factors are Standardisation of processes and Continuous improvement. In addition, these factors are not presented as critical in the mentioned model.

One of the findings of the survey reveals the limitations of further improvement of business process management maturity in the observed companies. Namely, the identified main lagging factors (Strategic alignment, Information technology, Business culture, Process measurement) in the model Rosemann and de Bruin (2005) are found as four of six the most important factors for the successful implementation of business processes.

The results of the correlation analysis show the existence of, stronger or weaker, correlation between a great number of maturity factors. The detection of a strong positive correlation between some maturity factors should serve as a guideline for electronics industry companies. It means that they should be aware that increase in the maturity level of Managing employees is followed by an increase in the maturity levels of Organisational changes, Continuous improvement and Business culture. For a more detailed analysis of established connections between maturity factors, dendrogram can be used and it shows that Managing employees provides synergy effects if change of factor Organisational changes is provided.

The results of correlation analysis show that achieving high maturity level of Information technology is followed by high state of business process management maturity. Also, Strategic alignment and Managing employees have strong positive correlation with business process management maturity.

The influence of the observed factors on business process management maturity may be evaluated based on regression analysis. Managing employees has greatest individual contribution to business process management maturity. This factor is one of most developed factors in researched companies and between it and maturity business process management is significant strong positive correlation. Also, Managing employees has strong correlation with other maturity factors (Organisational changes, Continuous improvement, Business culture, etc.). According to that, this factor can be considered as critical for electronic industry improvement.

This paper is an attempt to identify the key factors of the achieved level of business process management development in electronics industry. The authors tried to determine if they are matched with maturity factors that are marked as critical in the Roseman and de Bruin model (2005). One of the maturity factors (Managing employees) deserves special manager's attention, based on the degree of development, strength of correlation with other factors and in terms of contribution to increasing the maturity of business process management. The main recommendation for managers in the electronic industry refer to the necessity of improving the main lagging maturity factors, as they are marked as factors with a significant contribution to increasing the maturity of business process management.

For further research in the field of business process management maturity, a wide framework of business process management factors can be used. By applying this set of maturity factors framework in selected industrial areas in our country, it would be possible to compare the identified critical factors and the achieved level of business process management maturity in order to create a framework of factors characteristic for developing countries.

References

Adamides, E.D. & Karacapilidis, N. (2006). A knowledge centred framework for collaborative business process modelling. *Business Process Management Journal*, 5, 557-575.

Alibabaei, A., Aghdasi, M., Zarei, B., & Stewart, G. (2010). The role of culture in business process management initiatives. *Australian Journal of Basic and Applied Sciences*, 4, 2143–2154. Retrieved from: http://eprints.qut.edu.au/43250/, Assessed on 11. November 2017.

Andersen, B. (1999). *Business Process Improvement Toolbox*. Milwaukee: ASQ Quality Press.

Bai,Ch. & Sarkis, J. (2013). A grey-based DEMATEL model for evaluating business process management critical success factors. *International Journal of Production Economics*, 146 (1), 281-292

Bandara, W., Alibabaei, A. & Aghdasi, M. (2009). Means of achieving Business Process Management success factors. In: *Proceedings of the 4th Mediterranean Conference on Information Systems*, Athens University of Economics and Business, Athens.

Benner,M. J. & Tushman, M. L. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *The Academy of Management Review*, 28(2), 238–256.

Bosilj-Vukšić, V., Spremić, M., Omazić, M.A., Vidović, M. & Hernaus, T. (2006). Menadžment poslovnih procesa i znanja u hrvatskim poduzećima. *EFZG working paper series*, (05), 1-20.

Cantista, I. & Tylecote, A. (2008). Industrial Innovation, Corporate Governance and Supplier-Customer Relationships. *Journal of Manufacturing Technology Management*, 19(5), 516-590.

Chae, B., Yen, H. R., & Sheu, C. (2005). Information technology and supply chain collaboration: Moderating effects of existing relationships between partners. *IEEE Transactions on Engineering Management*, 52(4), 440–448.

Chen, I.J. (2001). Planning for ERP systems: analysis and future trend. *Business Process Management Journa*l, 7(5), 374-86.

Duh, R.-R., Chow, C. W. & Chen, H. (2006). Strategy, IT applications for planning and control, and firm performance: The impact of impediments to IT implementation. *Information & Management*, 43(8), 939–949.

Erl, T. (2005) *Service-oriented architecture concepts, technology and design*. Upper Saddle River. NJ: Pearson Education.

Fisher, D.M. (2004). The Business Process Maturity Model—A Practical Approach for Identifying Opportunities for Optimization. Bearing Point, Inc. Retrieved from: http://www.bptrends.com/publicationfiles/1004%20ART%20BP%20Maturity%20Model%20-%20Fisher.pdf, Assessed on 10. December 2017.

Garimella, K., Lees, M. & Williams, B. (2008). *BPM Basics for Dummies*. Software AG Special Edition. Wiley Publishing, Inc., Indiana: Indianapolis.

Hall, J. M. & Johnson, M. E. (2009). When should a process be art, not science?. *Harvard Business Review*, (March), 59–65.

Hammer, M., & Stanton, S. (1999) How process enterprises really work. *Harvard Business Review*, 77(6), 108–118.

Harmon, P. (2003). *Business process change: A manager's guide to improving, redesigning and automating processes*. San Francisco: Morgan Kaufmann.

Harmon, P. (2014). *Business Process Change.: A Business Process Management Guide for Managers and Process Professionals* (3nd ed). Burlington: Morgan Kaufmann.

Hassan, N. R. (2009). Using social network analysis to measure IT-enabled business process performance, *Information Systems Management*, 26(1), 61–76

Herzig, S. E. & Jimmieson, N. L. (2006). Middle managers' uncertainty management during organizational change. *Leadership & Organization Development Journal*, 27(8), 628–645.

Hiemstra,A. Ravesteyn,P & Versendaal J. (2009). An alignment model for business process management and service oriented architecture. In: 6th intenational conference on enterprice,systems,accounting and logistics.

Jarrar, Y. F., Al-Mudimigh, A. & Zairi, M. (2000). ERP implementation critical success factors-the role and impact of business process management. *In Management of Innovation and Technology*, ICMIT 2000.1, 122-127.

Jeston, J. & Nelis, J. (2014). *Business process management*. New York: Routledge.

Koster, S. (2009). An evaluation method for Business Process Management products. Graduation thesis. Business Information Technology. University of Twente. Netherlans

Küng, P. & Hagen, C. (2007). The fruits of business process management: An experience report from a Swiss bank. *Business Process Management Journal*, 13(4), 477–487.

Lahajnar, S. & Rožanec, A. (2016). The evaluation framework for business process management methodologies. *Management*, 21(1), 47-69.

Leavitt, H. (2005). *Top down: Why hierarchies are here to stay and how to manage them more effectively*. Boston: Harvard Business School Press.

Madison J. D. (2005). *Process Mapping, Process Improvement and Process Mangement – A Practical Guide to Enhancing Work and Information Flow*. Chico: Paton Press.

Melenovsky, M.J. & Sinur, J. (2006). BPM maturity model identifies six phases for successful BPM adoption, *Gartner Research*, ID No: G00142643 18 October.

Melville,N., Kraemer,K. & Gurbaxani,V. (2004). Information technology and organizational performance: An integrative model of IT business value. *MIS Quarterly,* 28(2), 283–322.

Mulyar, N. & van der Aalst, W. M. P. (2005). Towards a pattern language for colored Petri nets. Paper presented at the sixth workshop and tutorial on practical use of coloured Petri nets and the CPN tools. Retrieved from: http://www.pads.rwth-aachen.de/wvdaalst/publications/p288.pdf, Assessed on 20. November 2017.

Ngai, E. W. T., Law, C. C. H., & Wat, F. K. T. (2008). Examining the critical success actors in the adoption of enterprise resource planning. *Computers in Industry*, 59(6), 48–564.

Radosavljević, M. (2015). Process Orientation as a Basis for increasing supply chain Management Maturity. *Economic Themes*, *53*(3), 398-414.

Ravesteyn, P. & Batenburg, R. (2010). Surveying the critical success factors of BPM-systems implementation. *Business Process Management Journal*,16(3), 492-507.

Ravesteyn, P. & Versendaal, J. (2007). Success factors of business process management systems implementation. *ACIS Proceedings*, 60, 396-406.

Reijers, H. A. & Mansar, L. (2005). Best practices in business process redesign: An overview and qualitative evaluation of successful redesign heuristics. *Omeg*a, 33(4), 283–306

Rhee, M. & Mehra, S. (2006). Aligning operations, marketing, and competitive strategies to enhance performance: An empirical test in the retail banking industry. *Omega*, 34(5), 505-515.

Roberts, EB. (2001). Benchmarking Global Strategic Management of Technology. *Res. Technol. Manage*, 44(2), 25-36.

Rosemann, M. & De Bruin, T. (2005a) Application of a Holistic Model for Determining BPM Maturity.*BPTrends*, p. 1-20.

Rosemann, M. & De BruinT. (2005b).Toward s a Business Process Management Maturity Model. 13th European Conference on Information Systems (ECIS 2005), Regensburg, Germany.

Rosemann, M. (2006). Potential pitfalls of process modeling: part A. *Business Process Management Journal*, 12 (2), 249-254.

Rosemann, M. & vom Brocke, J. (2010). The six core elements of business process management. In: vom Brocke, J. & Rosemann, M. (Eds.), *Handbook on business process management, Introduction, methods and information systems* 1, 109–124, Springer, Berlin.

Savolainen, T. I. (1999). Cycles of continuous improvement: Realizing competitive advantages through quality. *International Journal of Operations & Production Management*, 19(11), 1203–1222.

Sharp, A. & McDermott, P. (2009). *Workflow modelling: tools for process improvement and applications development*. Norwood: Artech House, MA.

Sikavica, P., & Hernaus, T. (2011). *Dizajniranje organizacije*. Novi informator, Zagreb.

Silvestro, R. & Westley, C. (2002). Challenging the paradigm of the process enterprise: A case-study analysis of BPR implementation. *Omega*, 30(3), 215–225.

Štemberger, M. I., Bosilj-Vukšić, V., & Jaklić, M. I. (2009). Business process management software selection–two case studies. *Economic research - Ekonomska istraživanja*, *22*(4), 84-99.

Terziovski, M., Fitzpatrick, P. & O'Neill, P. (2003). Successful predictors of business process reengineering (BPR) in financial services. *International Journal of Production Economics*, 84(1), 35–50.

Trkman, P. & McCormack, K. (2010). Estimating the benefits and risks of implementing e-procurement. *IEEE Transactions on Engineering Management*, 57(2), 338–349.

Trkman, P. (2010). The critical success factors of business process management. *International journal of information management*, 30(2), 125-134.

Valle S, & Vázquez Bustelo, D. (2009). Involving suppliers in new product development: An Analysis of Spanish Firms. *Universia Bus. Rev.*, 24, 62-75.

vom Brocke, J. & Schmiedel, T. (2011). Towards a conceptualisation of BPM culture: Results from a literature review.15th Pacific Asia Conference on Information Systems, Brisbane, Australia: In PACIS Proceedings. Retrieved from: https://webdocs.uni.li/public/03805967.PDF, Assessed on 10. December 2017.

Weske, M. (2007). *Business process management. Concepts, languages and architectures.* Berlin Heidelberg, Springer.

IDENTIFIKACIJA I ANALIZA KLJUČNIH FAKTORA UPRAVLJANJA POSLOVNIM PROCESIMA

Rezime: U uslovima sve oštrije konkurencije na tržištu, kompanije mogu ostvariti konkurentsku prednost jedino bržom ponudom kvalitetnijih i jeftinijih proizvoda/usulga. U tom cilju, kompanije moraju izvršavati inovativne i delotvorne poslovne procese i upravljati njima adekvatno. U tom smislu, značajno je identifikovati faktore koji se mogu smatrati kritičnim za poboljšanje upravljanja poslovnim procesima. Cilj ovog rada je identifikacija i analiza ključnih faktora upravljanja poslovnim procesima u jednoj od zemalja u razvoju, na primeru elektronske industrije. Rezultati istraživanja pokazuju da faktori koji su obično prihvaćeni kao najvažniji za povećanje zrelosti upravljanja poslovnim procesima nisu prepoznati kao dominantni u okviru elektronske industrije. Jedan od zaključaka istraživanja otkriva kao glavne zaostajuće faktore četiri od šest najvažnijih faktora za uspešnu implementaciju poslovnih procesa. Na osnovu rezultata istraživanja, kao najznačajniji faktor zrelosti za kompanije iz oblasti elektronske industrije otkriveno je Upravljanje zaposlenima, stoga je predlog dalje unapređenje u ovoj oblasti.

Ključne reči: procesna orijentacija, upravljanje poslovnim procesima, faktori zrelosti, elektronska industrija

Authors' biographies

Aleksandra Stoiljković Randjelović is a PhD student at the Faculty of Economics, University of Nis, module Finance and Banking. In 2007, she graduated from the Faculty of Economic in Nis and in 2012 defended her Master's thesis at the Faculty of Economics in Nis, as well. Now she is working on a doctoral dissertation in the field of process management.Her key interest areas are the following: banking, risk management, process management, human resource management.

Radenko Milojević is a full-time professor at the Faculty of Economics, University of Niš, for narrow scientific field Business management. He is a lecturer for the following courses: Business resource management, Human Resources Management, Agency Business, International Human Resource Management, Business Process Management. He completed his undergraduate studies at the Faculty of Economics in Kragujevac, graduate at the Faculty of Economics in Belgrade, and he defended his doctoral thesis at the Faculty of Economics in Subotica. The key areas of his interest are: functioning of the enterprise, human resources management, business process management.

Marija Radosavljević is an Associate Professor at the Faculty of Economics, University of Nis, for narrow scientific field business management. In 2001, she graduated from the Faculty of Economic in Nis as a student of generation and defended her Master's and doctoral thesis at the Faculty of Economics in Belgrade in 2004 and 2009, respectively. She has participated in a few projects, financed by the Ministry of Science of Republic of Serbia. In 2005, 2009 and 2011 she was engaged as a researcher at the University for Economics and Business Administration, Department for Production Management, Wien, Austria and in 2013 as a visiting professor at Willamette University, Atkinson Graduate School of Management, Oregon, the US, as the only candidate from Serbia. Her key interest areas are: quality management, process management, cost management, human resource management. So far, she has published more than 130 papers in journals and proceedings from the conferences (national and international) and 5 monographs.

ANEXO 6: GAIARDELLI, SACCANI, & SONGINI, 2007 - PERFORMANCE MEASUREMENT SYSTEMS IN AFTER-SALES SERVICE: AN INTEGRATED FRAMEWORK.

Int. J. Business Performance Management, Vol. 9, No. 2, 2007 145

Performance measurement systems in after-sales service: an integrated framework

Paolo Gaiardelli*

Department of Industrial Engineering
Università di Bergamo
Viale Marconi 5 – 24044 Dalmine, Italy
E-mail: paolo.gaiardelli@unibg.it
*Corresponding author

Nicola Saccani

Department of Mechanical Engineering
Università di Brescia
Via Branze 38 – 25123 Brescia, Italy
E-mail: nicola.saccani@ing.unibs.it

Lucrezia Songini

Department of Accounting and Finance
Università Commerciale Luigi Bocconi
Via Sarfatti 25 – 20136, Milano, Italy
E-mail: lucrezia.songini@unibocconi.it

Abstract: In today's competitive markets, companies are shifting from a Productcentric to a Customercentric view. The After-Sales (AS) service can become a key differentiator and a major profit source. Thus, after-sales cannot be considered simply a set of operative activities; rather it plays a strategic role, affecting the definition of the product-service mix offered to the customer and the physical and organisational configuration of the overall logistics chain. Therefore, importance should be given to its strategic management and to the definition of a structured business performance measurement system. In this context, this paper a) provides a review of the existing body of knowledge about AS performance measurement systems, b) proposes an integrated framework for AS performance measurement consisting of four levels (business, process, activity, and development and innovation), and c) provides an empirical application of the framework to four case studies in durable consumer goods – automotive, home appliances and consumer electronics.

Keywords: After-Sales (AS) service; performance measurement systems; reference model; case studies; durable goods.

Reference to this paper should be made as follows: Gaiardelli, P., Saccani, N. and Songini, L. (2007) 'Performance measurement systems in after-sales service: an integrated framework', *Int. J. Business Performance Management*, Vol. 9, No. 2, pp.145–171.

Biographical notes: Paolo Gaiardelli is a Researcher in the Department of Industrial Engineering at the University of Bergamo. He is Assistant Professor in Operations and Industrial Plants Management. His main research topics are logistics and operations management, with a specific interest in the field of after-sales services.

Nicola Saccani graduated in Industrial Engineering at the Politecnico di Milano (Italy) and at the Ecole Centrale de Lille (France). He received his PhD in Operations Management at the University of Brescia (Italy) in 2005. His research topics and teaching assignments include production planning and control, buyer-supplier relationships and after-sales service configuration and management.

Lucrezia Songini is a Lecturer at Bocconi University, Milan and senior faculty member of the Accounting and Control Department of the SDA Bocconi School of Management. She is Professor of Management Accounting at the Università degli Studi del Piemonte Orientale 'Amedeo Avogadro', Novara, in the Business Administration Department, in Casale Monferrato. Her main research topics are strategic cost management and performance measurement and evaluation, with a specific interest in the field of after-sales services.

This work was developed and written jointly by the authors. Sections 1 and 5.2 were written by Paolo Gaiardelli, Sections 2, 4 and 6 Lucrezia Songini and Sections 3, 5.1 and 5.3 by Nicola Saccani.

1 Introduction

Over the past years, as sales growth in the durable-goods market has slowed and product margins have decreased, the After-Sales (AS) activities – those activities taking place after the purchase of the product and devoted to supporting customers in the usage and disposal of the goods – have become increasingly important as sources of differentiation and profit for manufacturers. Therefore, a shift in emphasis is taking place, from a traditional *productcentric* view to a more innovative *customercentric* view.

The relevance of AS service is demonstrated by the profit it generates, often higher than the one obtained with the products' sales: the service market can be four or five times larger than the market for products (Bundschuh and Dezvane, 2003) and may generate at least three times the turnover of the original purchase during a given product's life cycle (Alexander *et al.*, 2002; Wise and Baumgartner, 1999), contributing about 40%–50% of the total revenue, and a profitability of up to 20%–25% (McClusky, 2002; Alexander *et al.*, 2002; Craemer-Kühn *et al.*, 2002).

These figures may explain the change in the role attributed to the AS function, traditionally seen only as a cost generator and a 'necessary evil', in favour of a more proactive one, which considers the AS as a source of competitive advantage and business opportunity. Therefore, the AS service should not be viewed simply as a set of operational activities, but rather as an integrated process to be managed with a systemic approach. As a consequence, a significant effort should be devoted to the strategic management of AS and to the definition of a structured performance measurement system.

In this context, this paper proposes:

1 a review of the existing body of literature about AS performance measurement systems

2 a new integrated framework for AS performance measurement

3 to provide an empirical application to four case studies in durable consumer-goods industries.

Therefore, the paper is divided and organised as follows. The next section introduces the characteristics of the AS service. Section 3 presents a literature review on performance measurement systems and on their specific applications in the AS service. Section 4 describes the proposed reference framework for AS performance measurement. Section 5 provides an empirical application on four in-depth industrial cases belonging to the automotive, home appliance and consumer electronics industries. Finally, Section 6 draws some conclusions and suggests directions for future research.

2 Characteristics of the after-sales service

Several definitions of AS service can be found in managerial literature. They mainly differ with respect to both the extension assigned to the concept of AS and its role inside the value chain (Cohen and Lee, 1990; Ehinlanwo and Zairi, 1996; Asugman *et al.*, 1997; Urbaniak, 2001). Despite the different definitions, some peculiar features of the AS service can be pointed out (Patelli *et al.*, 2004a):

a AS represents a business

b AS is a service

c AS is a process, consisting of different activities, carried out by actors belonging to different functions and organisations

d AS constitutes an organisational unit, with different possible economic responsibilities (cost centre, profit centre and investment centre)

e the AS process of a single company (*e.g.*, the product manufacturer) is part of a wider network, involving different actors (dealers, suppliers, technical assistance centres, logistic service providers, the final customers and so forth).

First of all, in most organisations AS represents a *business*, which can generate significant profitability, often greater than the one generated by product sales. AS represents a division and its management has to reach adequate financial results (costs, revenues, operating profit, Return On Assets (ROA), cash flow) and competitive performance (market share, market penetration, customer satisfaction and loyalty, competitors' results). In order to develop the AS business, a proper balance between the orientation to profitability and the one to customer satisfaction and loyalty, as well as an adequate level of investments, have to be assured, both in the short term, as well as in the long term. The design of a performance measurement system has to consider the peculiar strategy of the AS business and to carefully monitor its implementation.

Secondly, being a *service*, some characteristics of AS that are typical of services, and which play a significant role in profit generation and customer satisfaction, have to be considered, such as the distinction between front-office and back-office activities, the relevance of some intangible assets, such as human resources, the proximity to the customer, the relevance of indirect costs, the focus on the service level (quality and timeliness) (Fitzgerald *et al.*, 1991). AS effectiveness depends mostly on front-office activities, while efficiency emanates from back-office ones. Moreover, in order to capture the contribution of intangibles, such as human resources or customer satisfaction, a set of operational and qualitative indicators which integrate the financial and quantitative ones has to be developed.

Thirdly, AS as a *process* can be viewed as the sum of activities needed to maintain product quality and reliability, sustained after the delivery took place, with the objective of increasing customer satisfaction (Ehinlanwo and Zairi, 1996). The efficiency and effectiveness of different activities impact on customer satisfaction, productivity and flexibility, in terms of responsiveness and service characteristics. Up until now, a recognised definition of activities which compose the AS process has not been developed. Only recently, a classification of AS activities has been proposed, which distinguishes between core and support macroactivities (Patelli *et al.*, 2004b). The activities require different emphasis to be put on the efficiency and effectiveness dimension, being either front-office activities or back-office ones.

Therefore, AS performance ought to be monitored at the activity level, in order to link explicitly resource consumption by each activity to the appraisal of the functionality delivered to the customers, pointing out efficiency and effectiveness drivers.

Fourthly, AS represents an *organisational unit*. The manager in charge of it needs a set of performance measures which can help him/her to analyse the variances between budgeted goals and actual results; to evaluate strengths and weaknesses of the organisational unit and to support decisions. An integrated system of measures is required to monitor the AS performance, which can relate specific Key Performance Indicators (KPIs) for the AS services to the company's overall strategic goals, in order to assess the consistency between strategic and operational objectives. Indicators, in order to be meaningful, have to consider both financial results (revenue, cost or profit) and operational drivers of performance, such as quality and responsiveness.

Finally, AS is characterised by a *network* perspective. In order to coordinate different actors along the value chain (Ehinlanwo and Zairi, 1996; Seuring, 2002), an integrated and multi-level set of measures needs to be properly designed.

The description of the main characteristics of AS provided in this section shows the need for a comprehensive evaluation of the AS service through a systemic perspective, which considers all the peculiar features of AS and identifies the characteristics that performance metrics should possess.

3 Performance measurement and after-sales service

3.1 Performance measurement systems

A plethora of literature concerning performance measurement systems exists. From the early 1980s up to the 1990s, most of the developed frameworks focused on the definition of performance attributes and on the classification of related measures. Different criteria for identifying critical aspects and attributes of performance have been pointed out

(Venkatraman and Ramanujam, 1986; Voyer, 1999). Most of the frameworks, in addition, addressed the corporate level and the strategic business areas (Kaplan and Norton, 1992; 1996; Olve *et al.*, 1997). Moreover, activities and processes were identified as relevant aspects of performance (Kaplan and Johnson, 1987; Johnson, 1992; Lorino, 1995; Lorino *et al.*, 1997; Wright and Keegan, 1997). The need to integrate traditional financial measures, such as Return On Equity (ROE) and Return On Investment (ROI), with both shareholder value indicators (economic value added, economic profit, *etc.*) and non-financial measures, was pointed out (Eccles, 1991; Stewart, 1991). The importance of integrating long-term measures, related to strategic planning, and short-term indicators, based on budget, in order to link those measures with the actions, organisation's mission and strategic objectives, was also pointed out (Tonchia, 2000). Moreover, the need to consider both tangible and intangible aspects, as well as efficiency and effectiveness, and innovation, was stressed. Different features of performance measures were pointed out: they have to be dynamic, relevant, timely, multi-dimensional, internal and external, in order to compare an organisation's results with competitors' performance, and also capable of predicting future outcomes rather than backward looking (Dixon *et al.*, 1990; Lynch and Cross, 1991; Bititci *et al.*, 2000).

The performance measurement frameworks developed in the 1990s most often aimed at linking strategy formulation to strategy implementation, long-term goals and decisions to short-term objectives and actions, by pointing out the relationship between performance drivers and competitive and financial results. Among the most popular and relevant models were the Performance Measurement Matrix (Keegan *et al.*, 1989), the Results and Determinants framework (Fitzgerald *et al.*, 1991), the SMART Pyramid (Lynch and Cross, 1991), the Balanced Scorecard (Kaplan and Norton, 1992; 1996) and the EFQM framework (EFQM, 1998; Olve *et al.*, 1997).

Since the late 1990s, on the other hand, frameworks in literature considered explicitly new dimensions of performance: stakeholder satisfaction (Atkinson *et al.*, 1997), corporate social responsibility and sustainability, intangibles and the supply chain.[1] Although all the listed frameworks consider several performance attributes and areas, they do not refer explicitly to the AS service. However, they can help in defining how to capture and measure the performance dimensions which distinguish the AS, since they point out all the relevant levels that can be applied to the evaluation of the AS service: strategic business area, process, activity and innovation.

3.2 Performance measurement in the after-sales service

Although in most manufacturing sectors AS is considered a key ingredient of competitive success (Cohen and Lee, 1990), few authors (if any) developed integrated performance measurement systems for the AS process as a whole. This section reviews the applications of the concepts of performance measurement to the AS service. From the analysis of existing literature, four theoretical perspectives and their approach to the measurement of AS service performance can be easily identified:

1 the product life cycle literature

2 the AS strategy literature

3 the spare-parts logistics literature

4 the supply chain and process-oriented approach.

3.2.1 The product life cycle literature

The life cycle perspective is founded on the integrated view of activities going from the initial product design to its dismissal/substitution by the end customer, so that a multi-activity and inter-organisational approach should be developed in order to measure all the relevant costs (Fabrychy and Blanchard, 1991; Artto, 1994). According to Shields and Young (1991), AS should be analysed from a customer perspective and a societal one, both being concerned with the support and disposal of the product. *Design for serviceability* is identified among the techniques for cost reduction in a life cycle perspective. Analysing the design and renewal of services, Meyer and DeTore (2001), moreover, apply a platform-based approach (*e.g.*, the modular architecture typical of product development) to the development of services. Target Costing and Life Cycle Costing (LCC), which allow a company to plan and manage jointly price, cost and profitability as well as time, quality and functionality in the product life cycle (Cooper, 1995; Cooper and Slagmulder, 1999; 2003), do not take into account the impact on AS (Lele, 1986; Goffin, 1990; 2000). In addition, Cohen *et al*. (1997) stress the importance of joint consideration of the AS service function revenue with the product sales revenue on a full ownership life cycle basis. Cohen and Whang (1997) developed a product life cycle model to study the strategic choices facing manufacturers of durable goods in the design of the product-service bundle. Moreover, the importance of assessing the costs suffered by the customer – either a business or a private one – with a life cycle horizon, arises; thus the diffusion of the Total Cost of Ownership (TCO) approach, which is aimed at understanding the true cost of buying a particular good or service from a particular supplier (Ellram, 1995).

A life cycle approach to the development of the product-service bundle is stressed in the works of Goffin (1998) and Goffin and New (2001): customer support is a focal element in the reach of customer satisfaction, and thus impacts on firms' market share and financial performance. Product-support requirements have therefore to be considered in the product design stage. Edvardsson (1997) highlights the role of new service development as the phases in which the foundations for total quality for customers should be built in. Along with Edvardsson (1997), other authors, such as Bullinger *et al*. (2003), Tax and Stuart (1997), Berry and Lampo (2000) and Froehle *et al*. (2000), analyse the new service-development process and propose conceptual frameworks for the design or redesign of services. As well as product innovation, service innovation is therefore widely recognised as a source of sustainable competitive advantage.

The foregoing brief review suggests that life cycle literature favours a financial accounting (costs and revenue are evaluated) and long-term (the product life-cycle) oriented perspective for the AS performance measurement. The AS is evaluated in terms of its contribution to value creation for the company, at a strategic business level. The innovation dimension, moreover, is taken into consideration. Only the TCO approach considers the activity performed, but for what performances are concerned, focuses on cost measures. On the contrary, specific short-term AS performance metrics stay conspicuous because of their absence.

3.2.2 The AS strategy literature

Only some years after the seminal work by Levitt (1983), business and operations management literature addressed the issue of the AS support strategy. Armistead and Clark (1991) related the AS strategy to the design of the delivery system, in particular in relation with the volume and the level of insourcing, the class of product (capital vs. consumer goods), and the phase of product life cycle. Moreover, the authors stressed the importance of the consistency of AS capabilities with the critical success factors of its context, as well as the strong dependency between AS and manufacturing delivery systems: both should be monitored in order to evaluate an overall customer confidence index. Frambach *et al.* (1997) set as the main success element of a product-service strategy the assessment of the relative importance of different AS services to market segments. Lawless and Fisher (1990) identified the related services (intangibles) as one of the elements affecting nonimitability, and thus durable competitive advantage of new products. Also, Mathieu (2001) defined goods-related services as a source of competitive advantage; moreover, she classified them into services which support the products and services which support the client actions in relation with the product. In the three previously quoted works, nothing is proposed at the level of performance measurement.

According to Lele (1997), the critical factor that influences the strategy formulation is the amount of costs, fixed (*i.e.*, not depending on the downtime) and variable (*i.e.*, depending on the downtime), incurred by the customer in the case of product failure. The proposed approach distinguishes three classes of strategies that allow reaching cost-effective configurations of the AS service on the basis of different customers needs and product specifications. The three classes pointed out by Lele are product-design–related strategies, strategies focused on service support systems and strategies aimed at reducing customer risks. More recently, Agnihothri *et al.* (2002) focused on field service, and on the role and influence of technology in creating an effective service organisation. They pointed out the management of three relationships – company-customer, company-(its own) employees, and (company's) employees-customer – as the critical factor for ensuring service effectiveness. In addition, they proposed measures to assess the three above-mentioned relationships.

However, the theoretical frameworks in literature concerning AS service strategy do not, in general, adopt an integrated view of the performance measurement system. Performance evaluation is considered only at the strategic business level, and no detail is given on the definition of metrics. Moreover, the innovation dimension is not considered by most of the quoted papers. Finally, only Agnihothri *et al.* (2002) suggested a set of performance metrics as a tool to test and verify the coherence between the strategic objectives and the effect of the actions undertaken.

3.2.3 The spare-parts logistics literature

Most existing literature related to the AS operations focuses on spare parts inventory and distribution management. From the point of view of performance measures, they are generally oriented to internal service-level metrics (*e.g.*, part fill rates), sometimes neglecting the assessment of the level of service received by the end customer. On these bases, a stream of publications studied the issues of spare parts inventory planning (Papadopoulus, 1996; Hopp *et al.*, 1999; Zhang *et al.*, 2001; Huiskonen, 2001; Kennedy *et al.*, 2002). Approaching the service performance issue from a logistics standpoint, Cohen and Lee (1990) defined some internal-oriented service measures, such as the part unit fill rate, the part dollar fill rate, the order fill rate, and external-oriented ones, such as the repair-order completion rate and the customer delay time. Moreover, they stressed the trade-off between inventory costs and service (response time). Among the other works emphasising the trade-off concept, we find Cohen *et al.* (1997), Agnihothri *et al.* (2002) and PRTM (2002). Overall, the literature in this section shows a narrow focus, often identifying AS with spare parts management and distribution activities. The performance measures proposed by literature in this area, even if often very detailed, are only operative and mostly focused on a specific activity or a set of activities. Moreover, they assess a single company's efficiency and effectiveness, rather than those of the entire supply chain, but their effects on the final customer (*e.g.*, response time, cost) are not considered.

3.2.4 The supply chain and process-oriented approach

A supply-chain oriented approach to the AS performance measurement is envisaged, for instance, by Cohen and Lee (1990), who highlight how:

a end-customer oriented service measures should be implemented

b service measures should be applied to all the parties involved in the supply chain (*e.g.*, suppliers, Original Equipment Manufacturers (OEMs), dealers, warehouses, technical assistance centres).

Nonetheless, the Supply Chain Operations Reference (SCOR) model (Supply Chain Council, 2003), a well-known model for the analysis and assessment of supply chain processes, does not encompass the AS as a process in itself. However, some of the activities constituting the AS process, *i.e.*, the logistic and materials management ones, are mapped and evaluated in the SCOR model. For instance, the 'return' process describes the reverse flows that may be related to AS. It encompasses three types of return flows: return of a defective product, return of a product for maintenance, repair or overhaul, and return of an excess product. On the other hand, other specific activities of AS, such as service delivery, customer care, training, and so forth, are not considered by the SCOR model.

Broadly speaking, works approaching the AS as a process and thus oriented to the definition of process measures are almost absent in literature. A notable exception is provided by Patton and Bleuel (2000), whose work touches the different areas related to AS, from budgeting and service forecasting to the definition of the organisational structure, the training, marketing and inventory management activities. Moreover, they deal with the issue of performance measurement; nonetheless, their perspective is only operational. In fact, they provide a list of service attributes to evaluate, along with possible indicators, without an integrated, multi-level approach. The aspects of customer satisfaction and service quality measurement are also approached, but no comprehensive framework is given. Together with Patton and Bleuel (2000), among the first attempts to fill the gap on integrated AS performance measurement is the work by Patelli *et al.* (2004b). They provide a definition of the AS service, seen as a business network process that involves different actors (Earl and Khan, 1994); moreover they identify AS's main constituting subprocesses and propose an activity-based costing (Innes and Mitchell,

1990) approach to its financial performance measurement. For the same sub-processes, Brun *et al*. (2004) identify performance attributes in the areas of service quality, timeliness, efficiency and costs.

This approach, as compared to others, proposes a rather integrated perspective, in terms of activities and actors involved. Nonetheless, it focuses mainly on operative aspects, often neglecting an assessment at the strategic level. This stream of literature is relatively new, and no established theory or conceptual model seems to exist at the moment.

In conclusion, literature dealing with the AS service presents a highly fragmented picture, where a systemic approach connecting the strategic aspects of AS service, a supply chain and process perspective, with a consistent set of performance metrics, is still lacking. None of the analysed streams of literature, in fact, addresses all the different characteristics of AS described in Section 2. Moreover, the attention devoted to performance metrics is weak. The only works that propose detailed performance metrics (Cohen *et al*., 1997; Supply Chain Council, 2003) deal with only some operative activities.

4 Reference framework

Despite the key role AS has assumed in the last few years, the previous section shows that there are few, if any, comprehensive and consistent applications of well-known performance measurement frameworks to the AS service in manufacturing firms, or little development of new ones. Therefore this paper aims at proposing an integrated reference framework for AS performance measurement.

A performance measurement system for the AS has to take into account its peculiar drivers of efficiency and effectiveness, by considering its typical features (see Section 2): at the OEM level, AS is at the same time a business, a process, a service and an organisational unit. Literature on performance measurement systems, described in Section 3.1, can suggest criteria to identify an organic system of metrics and indicators for AS:

1 An effective performance measurement system has to be articulated according to different *levels of analysis*, in order to drill down objectives and results consistently with relevant dimensions of analysis (strategic business area, process and activities, organisational units, supply chain).

2 At each level, it has to refer to different *aspects and dimensions of performance*, in order to capture all the critical aspects of AS.

3 It has to balance financial and non-financial indicators, in order to relate the operative drivers with the financial results (cause and effects).

4 It should encompass both long-term and short-term perspectives, tangible and intangible aspects, efficiency- and effectiveness-oriented measures.

The integrated framework proposed in Figure 1 is addressed to companies operating in the AS business. The adopted perspective is the one of a single enterprise. The framework aims at:

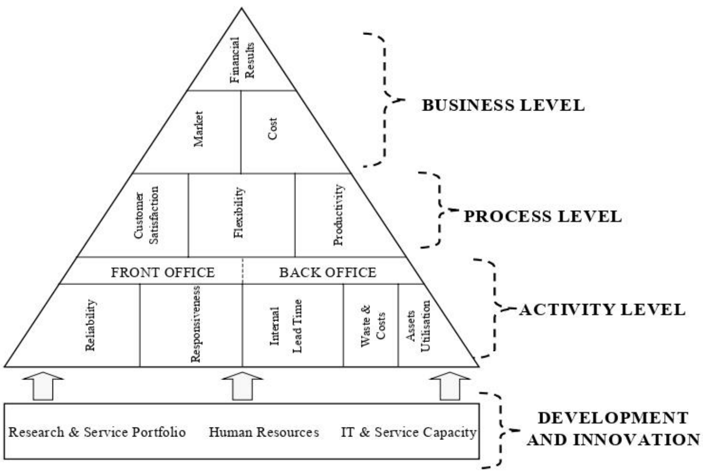
a supporting the design or the redesign of the performance measurement system

b assessing the consistency among the objectives and the performance metrics evaluated at the different levels of a single company framework

c describing and comparing the different performance systems adopted by different companies, or by one company across time.

The new framework is built by combining previous literature, in particular taking as a reference point Lynch and Cross (1991), Kaplan and Norton (1992) and the Supply Chain Council (2003). It links each peculiar feature of AS (see Section 2) with its related performance attribute, level and indicators. The framework distinguishes the short-term and the long-term perspective, as well as efficiency, effectiveness and innovation-related performance areas.

Figure 1 The After-Sales performance measurement framework



The framework is articulated in four levels:

1 business area

2 process level

3 activity level, whose performance measures are focused on the short term

4 development and innovation level, which considers a long-term perspective.

Only a stable and adequate investment in new services and products, human resources, IT and service capacity can ensure future profitability of AS. For this reason, the development and innovation level represents the base of the model, which 'feeds' in the long term all the other levels. The framework is systemic (encompassing different levels and areas), and hierarchical. The relationships exist among all levels, areas and indicators: the performance of lower levels and areas influence the results of the upper ones. The effectiveness-related performance areas (on the left side of the framework) are distinguished by the efficiency ones (on the right side of the framework). Different areas of performance at each level are defined and linked consistently with their impact on effectiveness or efficiency.

At the strategic business area level, the framework considers the overall AS performance, mainly a financial one. It can be measured by indicators such as operating profit, ROA and ROI. It is important to stress that the financial results are generated both by market results (market share, market penetration, *etc.*) that impact on revenue, and by the efficient consumption of resources (costs).

The second level in the framework is the process one. It is noteworthy to point out that the process is the linkage between the business's strategic objectives and specific activities carried out. A peculiar element of the model is the explicit recognition of the process as a specific measurement dimension. According to Lynch and Cross (1991), process performance can be measured with regard to customer satisfaction, flexibility and productivity. The customer satisfaction measures are devoted to identifying the existing gap between the expectations of the customer and the performance level of the firm, with regard to the characteristics of the output delivered. The flexibility measures the firm's ability to efficiently satisfy the customer expectations, both from an external perspective and an internal one (Lynch and Cross, 1991). The external perspective is related to the customisation of the output and its delivery time according to specific customer needs; the internal perspective is centred on the process lead time and the ability of the firm to manage activities fluently and without loss of time. Finally, the productivity indicators refer to the efficiency in resource consumption, and specifically to the link between resource consumption in the activities carried out and the output generated. It is noteworthy to point out that several companies evaluate those aspects, especially customer satisfaction, but focus on a single organisational unit or activity, without considering that they depend on the effectiveness and coordination at a process level.

The third level considers the performance of the AS organisational unit in dealing with its specific activities. It is useful to distinguish between front-office activities, impacting directly on customer satisfaction, and back-office ones, which are responsible for efficiency and lead times. Among the various works which dealt with the performance measurement at the activity level, we build on the Lynch and Cross and SCOR models. Lynch and Cross (1991) propose an explicit linkage between the performance metrics at the activity and at the process levels. For instance, reliability performance at the activity level impacts on the customer satisfaction at the process level. The performance attributes proposed at the activity level are quality, delivery, process time and cost. On the other hand, the SCOR model (Supply Chain Council, 2003) suggests other performance attributes to be evaluated at the activity level: reliability, responsiveness, flexibility, costs and assets. We agree with Lynch and Cross in considering flexibility to be a performance dimension characterising the process level. Therefore, integrating the two approaches, we point out five performance dimensions at the activity level: reliability, responsiveness, internal lead times, waste and costs, and asset utilisation. ***Reliability*** refers to the performance of the AS in delivering the correct product/service to the correct place, at the correct time, in the correct conditions and packaging, in the correct quantity, with the correct documentation, to the correct customer. ***Responsiveness*** is the speed at which AS provides products and services to the customer. *Internal lead times* represent the speed at which back-office activities are carried out, while *waste and costs* refer to internal efficiency in the consumption of resources. Finally, *asset utilisation* refers to the effectiveness of AS in managing assets (fixed and working capital) to support demand satisfaction. Different indicators can be associated with each performance attribute. Reliability and responsiveness performance can be evaluated in the case of front-office activities, while internal lead time, waste, costs, and asset utilisation are assessed with regard to back-office activities.

Finally, the need to integrate the short-term with the long-term perspective, as suggested by Kaplan and Norton (1992), Fitzgerald and Moon (1996) and Fitzgerald *et al*. (1991), leads to the definition of a fourth level of performance measurement, assessing the development and innovation dimension. It aims at capturing the drivers of stable and adequate future competitive and financial results, through:

a investments in new products and services, which assure the renewal of product and service portfolio according to customer needs and competitors' actions

b investments in intangibles, such as human resources, which are particularly relevant in a service activity

c investments in infrastructures (IT and service capacity), which allow a balanced growth of AS volumes and profits.

The possibility to integrate the operating unit level, focused on activities, with the business and process levels allows companies to relate corporate strategic performance with operative ones and to assess the cause-and-effect relationship between operational drivers and financial and competitive results. It allows the monitoring and evaluation of the AS performance not only from a strategic and managerial point of view, which is of interest to top and middle management, but also with regard to the efficiency and effectiveness of single activities at an operational level, which are relevant to employees in charge of specific activities (material procurement, product installation, maintenance, spare parts delivery, *etc*.). Therefore, the framework helps to align the strategic vision and goals with the tactical and operational objectives, and the results of different operative activities.

In Table 1 an exemplification of the possible metrics that can be used at different levels is proposed. It is not an exhaustive list of indicators, but rather a suggestion based on a literature analysis and on the empirical research performed by the authors. The following section details the application of the framework to four different case situations, and compares the performance systems adopted by different companies.

Table 1 An exemplification of AS metrics

Area	*Metric*	*References*
Business level		
Financial results	Share growth; Cash flow; Return On Equity (ROE); Return On Sales (ROS); Return On Investment (ROI); Return On Assets (ROA); gross margin; profit growth; net profit/loss (%); equity ratio on industrial operations; equity ratio on financial operations; net revenues; net profit/loss per year, cash to cash cycle time.	Fitzgerald *et al.* (1991); Kaplan and Norton (1992); Lynch and Cross (1991); Case Studies
Market	Market share; market penetration	Fitzgerald *et al.* (1991); Kaplan and Norton (1992); Lynch and Cross (1991); Case Studies
Cost	AS service costs, AS personnel costs (total); AS personnel costs on total number of employees; AS overhead costs	Fitzgerald *et al.* (1991); Kaplan and Norton (1992); Lynch and Cross (1991); Case Studies
Process level		
Customer satisfaction	Customer Satisfaction Index (based on reliability, responsiveness, comfort, courtesy, competence, access, availability, security, aesthetics); expected AS service; perceived AS service; loyalty customer index; Service Reliability index; Repurchase Intent Index	Fitzgerald *et al.* (1991); Kaplan and Norton (1992); Lynch and Cross (1991); Parasuraman *et al.* (1988); Case Studies
Flexibility	Stock Rotation Index, Time To Market (service – product); service redesign frequency; N° of products/N° of services; upside source flexibility; AS variable costs on total costs (variable + fixed)	Fitzgerald *et al.* (1991); Lynch and Cross (1991); SCOR (2003); Case Studies
Productivity	Global productivity on cost basis: output value/input value, Break Even Time, Activity Rotation Index, Profit margin (on total AS service value)	Fitzgerald *et al.* (1991); Kaplan and Norton (1992); Lynch and Cross (1991); SCOR (2003); Case Studies

Table 1 An exemplification of AS metrics (continued)

	Reliability	*Responsiveness*	*Internal lead time*	*Cost and waste*	*Assets utilisation*
Activity level					
AS planning	AS planning forecast accuracy; % resources unavailability under request; fill and error rate	Time horizon and time interval on forecast accuracy; start-up time for a new service; average spare parts delivery time; maximum spare parts delivery time		AS planning cost per employee; service planning cost as a % of AS Costs; AS planning cost per hour	Inventory days of supply; human resources productivity
Inventory and distribution management	Fill and error rate; transportation damages on delivery	Warehousing service level; warehousing picking time; delivery time	Warehousing download time; warehousing picking time	Cost of picking errors; stock value; stock management cost on AS cost; inventory obsolescence as a % of total inventory	Transportation damages on supply
Procurement and supply management	Supplier delivery performance; supplier quality performance; supplier price performance		Procurement response time	Cost to measure supply chain performance as a % of revenue	Assets as a % of non conformance cost
Customer care	Total number of claims received; first level calls; number of resolved complaints; % of resolved complaints; % of immediate responses; number of customer care calls not fulfilled	Mean time to response; customer calls abandon rate; percentage of responses below time limit		Cost per registered claim; customer care total cost	Number of customers per customer care employee
Service delivery (including maintenance, repair and reporting)	Mean Time Between Failure; First Right First Time	Mean Time To Repair (MTTR); product reuse time (diagnosis time + MTTR); average spare parts delivery time; maximum spare parts delivery time; number of parts delivered in delay; delivery perfomance; on time delivery	Mean Time To Repair (MTTR); diagnosis time	Repair cost	Number of technical reports per month; frequency of network reports; number of technical reports per employee
Warehousing	Number of stock-outs per month; fill and error rate	Delivery time	Time between repairing and spare parts receiving	Stock value; stock management cost on AS cost; Inventory obsolescence as a % of total inventory	Inventory days of supply
Reverse logistics	Correct authorisation on product return/total authorisation			Rework costs; reverse logistic on AS; defective products stock value	

Note: Metrics listed are issued from the SCOR model and the case studies analysed

Table 1 An exemplification of AS metrics (continued)

Area	*Metric*	*References*
Development and innovation		
Service portfolio	Number of services offered; number of requested services not offered previously; availability of internal tyres centre	Fitzgerald *et al.* (1991); Kaplan and Norton (1992); Case Studies
A human resources	Frequency of training courses, number of courses per employee per year	Fitzgerald *et al.* (1991); Kaplan and Norton (1992); Case Studies
IT and service capacity	Number of installed equipment, number of available ramps; number of repair shops; number of AS personnel; number of warehouses (echelons); number of workers per repair shop; number of technical assistance centres; number of tester for diagnosis on line; number of technical reports per month; frequency of network reports	Fitzgerald *et al.* (1991); Kaplan and Norton (1992); Case Studies

5 Empirical application

5.1 Research design and methodology

A multiple-case studies research design was used to evaluate the framework proposed in the previous section. The sample is composed of four companies, operating in Italy and belonging to the automotive (Companies 1 and 2), white goods (Companies 3 and 4) and consumer electronics industries (Company 3). The cases allowed testing of the ability of the proposed framework to describe the performance measurement system developed by each company, with a systemic and integrated perspective. Assessing the relative attention given by the companies to the dimensions constituting the framework, it is also possible to compare the different approaches to AS performance measurement adopted by the companies and to investigate their causes.

Case studies were performed through semi-structured interviews, a detailed questionnaire (available from the authors upon request), direct observation (warehouse and company tours) and the analysis of secondary sources (such as company documentation, corporate website, specialised press).

Four to six visits to each company's unit were made. Informants included the AS managing director, the manager(s) in charge for AS in Italy, the spare parts warehouse and material planning managers, and the customer care manager. Data collection focused on research variables describing the company, its AS unit and the AS supply chain configuration. In particular, the issues related to performance measurement were explored in the cases at each of the levels described in the framework in Section 4. The several informants and the different data sources used allowed for triangulation, to check the internal consistency of data (Voss *et al*., 2002).

Finally, cross-case comparisons were made to identify main differences and common behaviours among companies (Yin, 1994).

The cases are reported in Section 5.2, while the interpretation according to the framework and the discussion are treated in Section 5.3.

5.2 Case studies

5.2.1 Company 1

Company 1 is the Italian branch of a successful European group, one of the world-leading automobile and motorcycle manufacturers in the premium sector. All the Italian activities are carried out in accordance with the headquarters' strategic decisions, whose main aim is to gear to a long-term profitable growth. The adopted sales and AS strategies are focused on protecting the 'brand experience': the company supports its profitable growth thanks to the strength and image of its brands, kept through price positioning and customer satisfaction, provided by high quality, technological innovation, safe and reliable products and a complete set of services.

The AS is organised as a division, directly reporting to the CEO and responsible for profits and losses. The AS organisation is divided into five main areas:

1 marketing spares and service

2 sales promotion and budgeting

3 product and warranty

4 customer relationship management

5 field operation management.

The latter has been specifically created to manage and control the 180 exclusivist assistance centres located in Italy. The strategic role that AS service plays for the company has affected the whole performance measurement system. The main indicators measuring the financial performance are revenues, equity, long-term debt, net profit/loss per year, equity ratio on industrial operations, cash flow, cash flow on capital expenditures and equity ratio on financial operations. The main competitive results evaluated are volume growth and market share.

From an operative point of view, the AS emphasis shifts towards the measurement of customer satisfaction. At an operative function level the focus is mainly on reliability (*e.g.*, fill and error rate, number of stock-outs per month), responsiveness (*e.g.*, warehousing service level, average and maximum spare parts delivery time, mean time of service delivery, mean time to response) and internal lead time (*e.g.*, mean time to repair). Company 1 measures several cost and asset utilisation indicators, of which the single-area operative cost, the order-line cost, and the stock rotation index are the main ones. Although the business unit invests a large amount of money and resources in research and service development, the performance measurement in this area is not highly developed. Only the network service capacity is evaluated: the main indicators used are the number of the installed equipment and the number of available ramps and repair shops.

5.2.2 Company 2

Company 2 is the Italian branch of one of the world-leading automakers, offering a full range of models from minivehicles to large trucks. The automotive business (including sales financing and services) accounted for a total sales of around €110 billion worldwide in 2003. The adopted strategy focuses on customer satisfaction, retention and loyalty in the long term. This strategy is pursued through four key principles:

1 product reliability, to satisfy customer needs through functionality, quality and time to market

2 supply chain configuration and coordination efforts, in accordance with the Just In Time (JIT) principles

3 hierarchical dependence of the logistic function on the marketing function, in order to emphasise the goals of brand image, customer satisfaction and retention

4 a performance measurement system consistent with the strategic and organisational variables.

The Italian branch (established in 1990) has known an exponential sales growth (from 15 000 vehicles sold in 1996 to 120 000 sold in 2002). Today it employs 170 people (80 in the AS unit). The AS business unit is responsible for the activities concerning spare parts management and distribution, assistance services, customer care, network support management and control. It can be divided into four main areas: customer and network technical support, logistics, customer relationship management and AS business development. In order to assure an efficient and effective JIT operation system, the headquarters project and apply in the subsidiaries and the network (192 exclusivist official centres in Italy) the same organisational model, in which every function has to pursue the goal of customer retention in the whole product life cycle and where the management of processes is standardised according to the headquarters' guidelines.

At the business level, the emphasis is mostly put on evaluating and controlling competitive results, such as the market share. The AS performance measurement and control system, both at the company and the assistance network level, mainly focuses on operative aspects. Several indicators of efficiency and effectiveness are monitored. Depending on the process they refer to (logistics, AS business development, customer relationship management, technical assistance), they are related to customer satisfaction, flexibility and service productivity.

At the AS function, a wide set of metrics can be found, to evaluate reliability (*e.g.*, transportation damages on delivery, number of order-lines fulfilled, percent of requested parts available, total number of claims received, number of resolved complaints, percent of immediate responses, number of customer care calls not fulfilled); responsiveness (warehousing service level, customer calls abandon rate, response and repair time, percentage of responses below time limit, number of parts delivered in delay); internal lead time (*e.g.*, warehousing picking and download time, procurement response time, diagnosis time, time between repairing and spare parts receiving, transport time, invoice delivery time); costs (spare parts and technical assistance costs, cost of picking errors, transportation damages on supply); and asset utilisation (stock available per month, stock rotation index, number of technical reports per employee, percentage of equipment used).

Finally, the vision of AS service as a long-term competitive weapon, is reflected in many performance indicators used by Company 2 to evaluate and measure its service portfolio (*e.g.*, number of services offered, frequency of introduction of new services, number of requested services that were not offered previously); the AS human resources (*e.g.*, frequency of training courses, number of courses per employee per year); IT & Service Capacity (number of AS personnel, number of workers per repair shop, number of installed ramps, number of tester for diagnosis online, availability of internal tyre centres, number of technical assistance centres); and the communication effectiveness in the Service Supply Chain (*e.g.*, frequency of network reports, number of technical reports per month).

5.2.3 Company 3

Company 3 is a subsidiary of a successful multinational group operating in the household appliances and consumer electronics sectors, established in 1945. The company consists of 103 employees in Italy, where the company structure includes a sales-agent network, technical assistance centres and call centres (both outsourced). The total turnover of the Italian branch was around €630 million in 2003. Range of products (the company markets products from telephony, domestic refrigerators, screen monitors to professional printers), innovation and technology, positioning in the middle-high price sector, organisational flexibility and customer satisfaction are the factors on which the business strategy is based. The AS strategy stresses customer satisfaction as its main objective; it is carried out by supplying the customer with additional services and transferring technical culture to the assistance network, in order to improve its effectiveness.

The Italian branch is structured as follows: under the Managing Director there are four trade divisions subdivided by product lines, the general Marketing and the Service and Quality function. The AS service is managed by the Service and Quality function. The Italian assistance network is composed of 300 specialised centres. A required target service level is defined by contract, as well as a list of indicators and metrics used by Company 3 to evaluate and control the required performance of each assistance centre.

The focus on customer satisfaction and on operative activities of the AS function affects the whole performance measurement system. Although responsible for losses and profits, the AS division does not emphasise the performance measurement of economic and financial aspects. On the contrary, several indicators and metrics are used to evaluate, control and measure the AS processes and functions. At the process level AS is measured through customer satisfaction indicators. At the function level several indicators are managed and controlled to evaluate the service activities, considering reliability (*e.g.*, number of nonconformity, quantity of requested parts available in stock); responsiveness (*e.g.*, customer calls abandon rate, mean time to response, mean time of service delivery and spare parts delivery time); internal lead time (*e.g.*, immediate shipment ratio, time for providing spares at the network, procurement dispatching time, mean time to repair); and costs and asset utilisation (*e.g.*, percentage of warranty repairs on total, number of substitutions, spares consumption, warranty costs, rework costs, cost of quality control). Finally, for research and service development, Company 3 does not evaluate its service portfolio, AS human resources and IT & Service Capacity through specific indicators.

5.2.4 Company 4

Company 4 is a white-goods manufacturer based in northern Italy, with three production plants in Italy. Born as a capacity supplier, Company 4 developed into a major Italian manufacturer of cooking appliances, and then to a relevant player in the top-range cooking and other white-goods markets. It experienced a significant growth in the last decade, and now counts around 1000 employees worldwide and a turnover near €300 million. With a focus on the medium-high market segment, Company 4's overall strategy is aimed at preserving and improving its brand image.

The AS service function, organised on a geographical basis, employs about 50 people and reports directly to the CEO. Along with a few staff personnel (accounting, statistics, technical documentation), the AS function has three main branches, namely the spare parts logistics, the Italian technical assistance, and the foreign technical supervision and support. The objectives of AS service focus on customer retention and on the development of the company's brand through customer satisfaction. The technical assistance in Italy is carried out by around 200 centres throughout the country, all autonomous businesses authorised by Company 4 to repair products and to sell accessories. The AS function recently shifted from a cost centre to a profit one with the aim, first reached in 2003, to self-finance its activities. A structured performance and cost measurement system has been in place in Italy for five years. In 2004 it was replicated in the foreign branches. The emphasis is, on one side, on the operational performance of the service while, on the other side, a tight control on AS operating costs is maintained, in order to preserve the self-subsistence of the AS function.

At a business unit level, the Profit & Loss is realised for each branch of AS (spare-parts logistics, the Italian technical assistance, and the foreign technical supervision and support). At the process level, AS is measured considering customer satisfaction indicators. At the activity level the main evaluated indicators are responsiveness (*e.g.*, mean and max time of service delivery, average spares delivery time, percent of services below time limit) and reliability (number of service parts available, percent of resolved complaints, percent of immediate responses). Moreover, assistance centres' fidelity in buying spare parts from Company 4 is also assessed. Internal service (*e.g.*, spare parts shipments) is compared to the delivery time promised to direct customers. Finally, indicators are measured by Company 4 to evaluate its service portfolio (*e.g.*, number of offered services, number of service contracts), the AS human resources (*e.g.*, number of training courses per employee) and IT & Service Capacity (*e.g.*, number of AS personnel, number of warehouses).

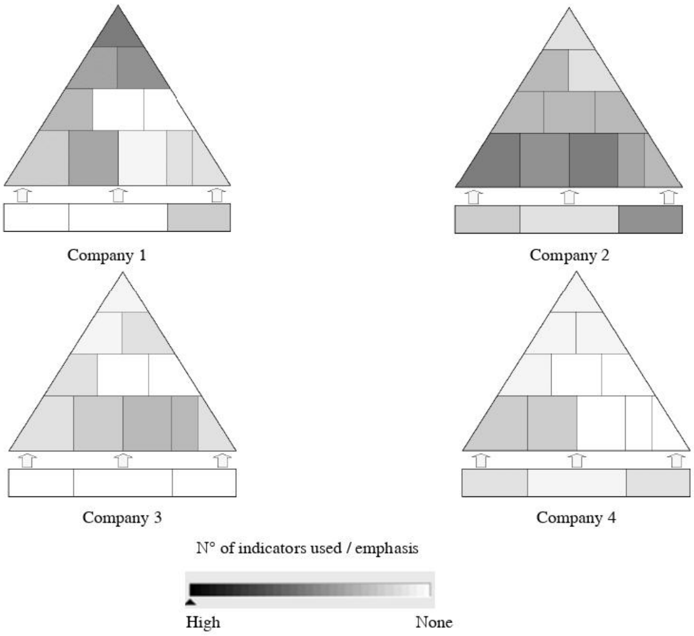
5.3 Discussion

In Figure 2, the main dimensions of performance measured by each company are mapped through the framework proposed in Section 4. For each company, each section of the diagram is filled with a different shade of grey according to the emphasis given to that specific performance area. Dark grey indicates that the company puts a high emphasis on the performance measurement of that area, and that multiple indicators are measured. The darker the grey, the higher the emphasis: the colour white indicates that a performance area is not measured by the company.

From Figure 2, it is possible to notice that case study companies implement AS performance measurement systems focusing on different aspects. Firms monitor different dimensions, going from financial results to competitive performance, from customer satisfaction and loyalty to process and activities' efficiency, and from lead time to quality. The dimensions differ according to the organisational level considered: strategic business unit, process, activities. The monitored aspects are both AS results (such as costs, profit, market share) and performance drivers (such as lead time, quality, customer satisfaction), but rarely is the relationship between drivers and results pointed out by companies. Some similarities among the case studies can be also pointed out. For instance, processes and activities are rarely considered dimensions of AS performance. Only Company 2, which adopts JIT principles, developed a common definition of processes also at the dealer level and implemented a consistent set of metrics. Company 2 is also the only firm which seems to cover adequately all the performance dimensions identified by the framework. The other companies rely on a (more or less) wide range of operational indicators not integrated in a systemic framework. All case studies show a link (explicit or, more often, implicit) between corporate strategic objectives, AS strategies and goals, and AS performance measures and indicators. However, apart from Company 2 and partially Company 4, firms adopt a short-term perspective on performance measurement, considering a budget horizon rather than a strategic planning one.

Figure 2 shows how the developed framework can serve as a descriptive tool, in order to draw a picture of a company's AS performance measurement system, and to develop a critical comparison of a set of companies, or a company over time.

Figure 2 Dimensions of performance mapped adopting the proposed framework



Interpretative power may be added to the framework by pointing out drivers that can explain the differences among companies and/or to evaluate the internal consistency of each company's behaviour. Table 2 shows a list of drivers and the positioning of each company, hereafter commented.

Table 2 Relevant drivers for the four case-study companies

Driver	*Company 1*	*Company 2*	*Company 3*	*Company 4*
Industry	Automotive	Automotive	Consumer Electronics – White goods	White goods
Size	Multinational	Multinational	Multinational	National
Product segment	Premium sector	Mass market	Mass market	Medium-high end
Time perspective (AS)	Short	Long	Short	Short
Economic responsibility (AS)	Profit and losses	Profit and losses	Profit and losses	Profit and losses

Company 1 and 2 belong to the automotive industry, which is rather advanced in terms of management techniques and which traditionally puts a high emphasis on performance measurement. This fact may explain why these companies assess, with high to medium emphasis, most performance areas. The generally lower emphasis on performance measurement of Company 4, moreover, may be due to its smaller size: it is the only national (Italian) company, addressing a European market. The product segment can act as a possible driver for the relative emphasis of efficiency and effectiveness measures: Company 1, acting in the premium segment, is more concerned with response times and customer satisfaction than with costs. The attention on the long term experienced by Company 2 shifts the relative emphasis towards the lower levels of the framework. Activity measures (low level of the pyramid) help in explaining the causes at the root of overall performance, while monitoring the capability of development and innovation supports the long-term strategic planning. Finally, the economic responsibility given to the AS organisation may explain the relative emphasis given to costs and competitive (revenue, market, *etc.*) measures.

6 Conclusions

The paper points out a gap between the increased relevance of AS services and the lack of integrated and systemic frameworks for AS performance measurement in literature. In order to fill (at least partially) this gap, we proposed a framework for AS performance measurement that combines the features of some existing models (such as the SMART Pyramid, the SCOR model and the Balanced Scorecard), integrating and adding to those models, in order to deal with the AS peculiarities described in Section 2. Consequently, the framework:

a addresses the AS at different levels from the business (and overall-results–oriented) one, to the process, activity and development/innovation ones

b addresses several performance areas at each level, giving emphasis to both efficiency and effectiveness performance

c at the same time, addresses internal and customer-oriented measures.

We believe that the framework may serve different situational needs: describing and comparing existing situations (*e.g.*, in an industry), interpreting and evaluating the differences, and designing or redesigning the AS performance measurement systems of a company, assessing the consistency among the objectives and the performance metrics.

The developed framework was applied to case studies from the automotive, consumer electronics and white-goods industries, showing the ability of the model to act as a descriptive tool for analysing and comparing different industrial realities (or a company across time). Interpretive power can be added to the framework by analysing specific drivers, such as the ones proposed in Table 1. This way, the different emphasis placed by firms on different levels and/or on different performance areas can be explained. The list of drivers proposed in this paper is not meant to be exhaustive, but rather should be viewed as a preliminary set, to be further developed. The aim of an enlarged and extended driver list is the analysis of the internal consistency of a company's AS performance measurement system (*e.g.*, with the company strategy, the product segment, the product life cycle phase, the type of customer addressed, the time perspective emphasised and so forth), and a benchmark analysis (with direct competitors or cross-industry). Suggestions on how to align or improve the performance measurement system may stem from this analysis.

In addition, in the broader empirical research undertaken by the authors, a general need for managerial reference models in the AS service organisations emerged. The proposed framework, therefore, can be applied as a tool on which to base the definition of a company's AS performance measurement system. The example proposed in Table 1 should be further developed in order to define a practical measurement system for practitioners. The utilisation of the framework does not imply that each company should put emphasis or even monitor each performance area included in the framework, but rather should understand its priorities and define its own profile of measures and indicators consistently with the company strategy and the other drivers quoted above. However, the validation of the proposed framework requires a more extensive and exhaustive research, based mostly on in-depth case studies. Moreover, as already stressed, the list of interpretation drivers has to be completed, and their action mechanism explained in detail.

Finally, in the cases studied, relatively little emphasis has been placed on the companies' supply chains. The framework, like the SCOR model, is meant to be interlinked with similar frameworks for the other supply chain actors. Thus, all supply-chain tiers have to be studied in depth (through case studies and extensive research) and their performance measurement systems assessed, in order to check the internal and overall consistency of the AS performance measurement systems. This last issue constitutes perhaps the most important indication for future direction of research, in accordance with the development of the *supply chain, process-oriented approach* pointed out in Section 3.2.

Acknowledgements

This work is part of a two-year Italian research project, named After-Sales Advanced Planning (ASAP), undertaken by five Italian universities (Università di Bergamo, Politecnico di Milano, Università di Brescia, Università L. Bocconi, Università di Firenze). The authors acknowledge the financial support received from the Italian Ministry of the University, Scientific Research and Technology (MIUR). For more information about the ASAP Project, please refer to the official website: www.progettoasap.org.

References

Agnihothri, S., Sivasubramaniam, N. and Simmons, D. (2002) 'Leveraging technology to improve field service', *International Journal of Service Industry Management,* Vol. 13, No. 1, pp.47–68.

Alexander, W.L., Dayal, S., Dempsey, J.J. and Vander Ark, J.D. (2002) 'The secret life of factory service centers', *The McKinsey Quarterly*, No. 3.

Armistead, C. and Clark, G. (1991) 'A framework for formulating after-sales support strategy', *International Journal of Operations and Production Management,* Vol. 11, No. 3, pp.111–124.

Artto, K.A. (1994) 'Life cycle cost concepts and methodologies', *Cost Management*, Fall.

Asugman, G., Johnson, J. and McCullough, J. (1997) 'The role of after-sales service in international marketing', *Journal of International Marketing*, Vol. 5, No. 4, pp.11–28.

Atkinson, A.A., Waterhouse, J.H. and Wells, R.B. (1997) 'A stakeholder approach to strategic performance measurement', *Sloan Management Review*, Spring.

Berry, L.L. and Lampo, S.K. (2000) 'Teaching an old service new tricks – the promise of service redesign', *Journal of Service Research*, Vol. 2, No. 3, pp.265–275.

Bititci, U.S., Turner, T. and Begemann, C. (2000) 'Dynamics of performance measurement systems', *International Journal of Operations and Production Management*, Vol. 20, No. 6, pp.692–704.

Brun, A., Novakova, G., Gaiardelli, P., Corti, D. and Pinto, R. (2004) 'Design of a performance measurement architecture for integrated after-sales processes', *Performance Measurement Association (PMA'04) Conference Proceedings, EICC*, Edinburgh (UK), 28–30 July, pp.163–170.

Bullinger, H.J., Fanrich, K.P. and Meiren, T. (2003) 'Service engineering – methodical development of new service products', *International Journal of Production Economics*, Vol. 85, pp.275–287.

Bundschuh, R.G. and Dezvane, T.M. (2003) 'How to make after sale services pay off', *The McKinsey Quarterly*, No. 4.

Cohen, M.A. and Agrawal, V. (2000) 'After-sales service supply chains: a benchmark update of the north American computer industry', *Internal Report, Fishman-Davidson Center for Service and Operations Management*, The Wharton School, University of Pennsylvania.

Cohen, M.A., Cull, C., Lee, H.L. and Willen, D. (2000) 'Saturn's supply-chain innovation: high value in after-sales service', *Sloan Management Review*, Summer, pp.93–101.

Cohen, M.A. and Lee, H.L. (1990) 'Out of touch with customer needs? Spare parts and after sales service', *Sloan Management Review*, Winter, pp.55–66.

Cohen, M.A., Zeng, Y. and Agrawal, V. (1997) 'Service parts logistics: a benchmark analysis', *IEEE Transactions – Special issue on Supply Chain Management*, Vol. 29, pp.627–639.

Cohen, M.A. and Whang, S. (1997) 'Competing in product and service: a product life-cycle model', *Management Science*, Vol. 43, No. 4, pp.535–545.

Cooper, R. and Slagmulder, R. (2003) 'Strategic cost management: expanding scope and boundaries', *Journal of Cost Management*, January–February.

Cooper, R. and Slagmulder, R. (1999) *Supply Chain Development for the Lean Enterprise: Interorganizational Cost Management*, Portland: Productivity Press.

Cooper, R. (1995) *When Lean Enterprises Collide*, Boston: Harvard Business School Press.

Craemer-Kühn, D., Junghans, M. and Krönig, J. (2002) *BER Revised – The Aftermarket Battle*, a McKinsey presentation.

Dixon, J.R., Nanni, Jr., A.J. and Vollmann, T.E. (1990) 'The new performance challenge: measuring operations for world-class manufacturing', *Business One Irwing*, Homewood, Ill.

Earl, M. and Khan, B. (1994) 'How new business process redesign?', *European Management Journal*, March.

Eccles, R.G. (1991) 'The performance measurement manifesto', *Harvard Business Review*, January–February.

Edvinsson, L. and Malone, M.S. (1997) 'Intellectual capital: realizing your company's true value by finding its hidden roots', *Harper Collins*, New York.

Edvardsson, B. (1997) 'Quality in new service development: key concepts and a frame of reference', *International Journal of Production Economics*, Vol. 52, pp.31–46.

European Foundation for Quality Management (EFQM) (1998) *EFQM Self-Assessment Guidelines*.

Ehinlanwo, O.O. and Zairi, M. (1996) 'Best practice in the car after-sales service: an empirical study of Ford, Toyota, Nissan and Fiat in Germany – Part 1', *Business Process Re-engineering and Management Journal*, Vol. 2, No. 2.

Ellram, L.M. (1995) 'Total cost of ownership. An analysis approach for purchasing', *International Journal of Physical Distribution and Logistics Management*, Vol. 25, No. 8, pp.4–23.

Epstein, M. and Wisner, P.S. (2001) 'Using a balanced scorecard to implement sustainability', *Environmental Quality Management*, Winter.

Fabrychy, W.J. and Blanchard, B.S. (1991) 'Life-cycle cost and economic analysis', *Englewood Cliffs*, New Jersey, Prentice Hall.

Figge, F., Hahn, T., Schaltegger, S. and Wagner, M. (2002) 'The sustainability balanced scorecard – linking sustainability management to business strategy', *Business Strategy and the Environment*, John Wiley & Sons, No. 11.

Fitzgerald, L., Johnston, R., Brignall, S., Silvestro, R. and Voss, C. (1991) 'Performance measurement in service businesses', *CIMA*.

Fitzgerald, L. and Moon, P. (1996) 'Performance measurement in service industries: making it work', *CIMA*, London.

Frambach, R.T., Wels-Lips, I. and Gundlach, A. (1997) 'Proactive product service strategies', *Industrial Marketing Management*, Vol. 26, pp.341–352.

Froehle, C.M., Roth, A.V., Chase, R.B. and Voss, C.A. (2000) 'Antecedents of new service development effectiveness', *Journal of Service Research*, Vol. 3, No. 1, pp.3–17.

Goffin, K. (1990) 'Design for support: results of a UK survey', *AFSM International, The Professional Journal*, Vol. 14, No. 10, pp.24–29.

Goffin, K. (1998) 'Evaluating customer support during new product development – an exploratory study', *The Journal of Product Innovation Management*, Vol. 15, No. 1, pp.42–56.

Goffin, K. (2000) 'Design for supportability: essential component of new product development', *Research Technology Management*, Vol. 43, No. 2, March–April.

Goffin, K. and New, C. (2001) 'Customer support during new product development – an exploratory study', *International Journal of Operations and Production Management*, Vol. 21, No. 3, pp.275–301.

Johnson, H.T. (1992) *Relevance Regained*, New York: The Free Press.

Hopp, W.J., Zhang, R.Q. and Spearman, M.L. (1999) 'An easily implementable hierarchical heuristic for a two-echelon spare parts distribution system', *IEEE Transactions*, Vol. 31, pp.977–988.

Huiskonen, J. (2001) 'Maintenance spare parts logistics: special characteristics and strategic choices', *International Journal of Production Economics*, Vol. 71, pp.125–133.

Innes, J. and Mitchell, F. (1990) 'Activity based costing: a review with case studies', *CIMA*, London.

Kaplan, R.S. and Johnson, H.T. (1987) *Relevance Lost – The Rise and Fall of Management Accounting*, Boston: Harvard Business School Press.

Kaplan, R.S. and Norton, D.P. (1992) 'The balanced scorecard. Measures that drive performance', *Harvard Business Review*, January–February.

Kaplan, R.S. and Norton, D.P. (1996) *The Balanced Scorecard. Translating Strategy into Action*, Boston: Harvard Business School Press.

Kaplan, R.S. and Norton, D.P. (2004) *Strategy Maps*, Boston: Harvard Business School Press.

Keegan, D.P., Eiler, R.G. and Jones, C.R. (1989) 'Are your performance measures obsolete?', *Management Accounting*, June, pp.45–50.

Kennedy, W.J., Pattersonb, J.W. and Fredendall, L.D. (2002) 'An overview of recent literature on spare parts inventories', *International Journal of Production Economics*, Vol. 76, pp.201–215.

Lawless, M.W. and Fisher, R.J. (1990) 'Sources of durable competitive advantage in new products', *The Journal of Product Innovation Management*, Vol. 7, pp.35–44.

Lele, M. (1986) 'How service needs influence product strategy', *Sloan Management Review*, Fall, pp.63–70.

Lele, M. (1997) 'After-sales service – necessary evil or strategic opportunity?', *Managing Service Quality*, Vol. 7, No. 3, pp.141–145.

Levitt, T. (1983) 'After the sale is over…', *Harvard Business Review*, September–October, pp.87–93.

Lorino, P. (1995) 'Comptes et récits de la performance', *Les Editions d'Organisation*, Paris (in French).

Lorino, P., Demeestere, R. and Mottis, N. (1997) *Contrôle de gestion et pilotage*, Nathan, Paris, in French.

Lynch, R.L. and Cross, K.F. (1991) *Measure Up – The Essential Guide to Measuring Business Performance*, Mandarin, London.

Mathieu, V. (2001) 'Product services: from a service supporting the product to a service supporting the client', *The Journal of Business and Industrial Marketing*, Vol. 16, pp.39–58.

McClusky, M. (2002) 'Service lifecycle management (Part 1): the approaches and technologies to build sustainable advantages for services', *AMR Research*, August.

Meyer, M.C. and DeTore, A. (2001) 'Creating a platform-based approach for developing new services', *The Journal of Product Innovation Management*, Vol. 18, pp.188–204.

Neely, A.D., Adams, C. and Kennerley, M. (2002) *The Performance Prism: the Scorecard for Measuring and Managing Business Success*, London: FT Prentice Hall.

Olve, N.G., Roy, J. and Wetter, M. (1997) *Performance Drivers. A Practical Guide to Using the Balanced Scorecard*, Chichester: John Wiley & Sons.

Papadopoulos, H.T. (1996) 'A field service support system using a queueing network model and the priority MVA algorithm', *Omega*, Vol. 24, pp.195–200.

Parasuraman, A., Zeithmal, V.A. and Berry, L.L. (1988) 'SERVQUAL: a multiple item scale for measuring consumer perception of service quality', *Journal of Retailing*, pp.12–40.

Patelli, L., Pistoni, A. and Songini, L. (2004a) 'The appraisal of after-sales service contribution to value creation. A conceptual framework', *27th Annual Congress of the European Accounting Association*, Prague, April.

Patelli, L., Pelizzari, M., Pistoni, A. and Saccani N. (2004b) 'The after-sales service for durable consumer goods. Methods for process analysis and empirical application to industrial cases', *13th International Working Seminar on Production Economics*, Igls (Innsbruck), Austria, 16–20 February, Pre-prints, Vol. 3, pp.289–299.

Patton, J.D. and Bleuel, W.H. (2000) *After the Sale*, New York: The Solomon Press.

PRTM – The performance Measurement Group (2002) *PMG's Scorecard Users Guide – Service Supply Chain*, www.prtm.com.

Seuring, S. (2002) 'Supply chain costing. A conceptual framework', in S. Seuring and M. Goldbach (Eds.) *Cost Management in Supply Chain*, Heidelberg, Physica-Verlag.

Shields, M.D. and Young, S.M. (1991) 'Managing product life cycle costs: an organisational model', *Cost Management*, Fall, pp.39–52.

Stewart, G.B. (1991) *The Quest for Value: The EVA™ Management Guide*, HarperCollins Publishers, Inc.

Stewart, T.A. (1999) *Intellectual Capital. The New Wealth of Organizations*, New York: Doubleday.

Supply Chain Council (2003) *Supply-Chain Operations Reference Model (SCOR)*, Ver. 6.0, Pittsburgh PA, USA, www.supply-chain.org.

Sveiby, K.E. (1997) *The New Organizational Wealth. Managing and Measuring Knowledge-based Assets*, San Francisco, CA: Berret Koehler.

Tax, S.S. and Stuart, I. (1997) 'Designing and implementing new services: the challenge of integrating service systems', *Journal of Retailing*, Vol. 73, No. 1, pp.105–134.

Tonchia, S. (2000) 'Linking performance measurement system to strategic and organisational choices', *International Journal of Business Performance Management*, Vol. 2, Nos. 1–3, pp.15–29.

Urbaniak, A.J. (2001) 'What really happen to customer service', *The American Salesman*, Burlington, Vol. 46, No. 2, pp.14–17.

Venkatraman, N. and Ramanujam, V. (1986) 'Measurement of business performance in strategy research: a comparison of approaches', *Academy of Management Review*, No. 4.

Voss, C., Tsikriktsis, N. and Frolich M. (2002) 'Case research in operations management', *International Journal of Operations and Production Management*, Vol. 22, No. 2, pp.195–219.

Voyer, P. (1999) 'Tableaux de Bord de gestion et indicateurs de performance', *Presses de L'Uiversité du Quebec*, in French.

Wise, R. and Baumgartner, P. (1999) 'Go downstream – the new profit imperative in manufacturing', *Harvard Business Review*, September–October, pp.133–141.

Wright, P.D. and Keegan, D.P. (1997) 'Pursuing value: the emerging art of the reporting on the future', *PW Papers,* Price Waterhouse LLP.

Yin, R.K. (1994) 'Case study research – design and methods', *Applied Social Research Methods Series*, 2nd ed., Thousand Oaks, CA: Sage Publications, Vol. 5.

Zhang, R.Q., Hopp, W.J. and Supatgiat, C. (2001) 'Spreadsheet implementable inventory control for a distribution center', *Journal of Heuristics*, Vol. 7, pp.185–203.

Note

1 The Value Reporting (Wright and Keegan, 1997), the Performance Prism (Neely *et al.*, 2002), the Sustainable Balanced Scorecard (Epstein and Wisner, 2001; Figge *et al.*, 2002), the SIGMA Sustainable Scorecard (www.sigma-project.org), the Intellectual Capital Model (Edvinsson and Malone, 1997; Sveiby, 1997; Stewart, 1999); the new balanced scorecard (Kaplan and Norton, 2004).

INTRODUCCIÓN A LA GESTIÓN DE INDICADORES

1. UN ENFOQUE SISTÉMICO DE GESTIÓN

Un elemento clave para empezar a gestionar de manera exitosa una organización es comprender cómo se generan los resultados del negocio.[1] En términos sencillos, los resultados se producen debido a la interacción de diversos componentes, tanto al interior como fuera de la organización.

– Un modelo para describir el enfoque sistémico de gestión

Esa comprensión, descrita líneas arriba, puede darse a través de lo que llamamos un “enfoque sistémico de gestión”. “Sistémico” porque actúa a través de componentes que se interrelacionan y que en su conjunto forman un sistema que produce resultados organizacionales.

Ciertamente, la definición que se ha hecho puede llevarnos a considerar el “enfoque sistémico” como un “modelo”, una abstracción, o una simplificación teórica. Los modelos son ampliamente utilizados en disciplinas diferentes a la administración y no hay razón por la que no podamos usar uno para intentar explicar cómo funciona una organización.

En el caso de una organización cualquiera, un modelo ayuda a comprender cómo se interrelaciona los resultados con la gestión.[2] Estos componentes del modelo son, a nuestro entender:

+ Los resultados financieros y económicos, en el caso de empresas. Para el caso de las organizaciones de gobierno, o de aquellas que puedan definirse como sin fines de lucro, los resultados están relacionados con el cumplimiento de la misión institucional.

1 La palabra “negocio” debe comprenderse en esta publicación como sinónimo de “la actividad de la organización”. En ese sentido, la expresión es usada de forma indistinta, tanto cuando se hace referencia a una empresa como a una organización de gobierno o sin fines de lucro.

2 La elección de estos componentes ha sido tomada del Modelo Malcolm Baldrige (2013). Sin embargo, la gráfica que representa el modelo de enfoque sistémico, o de la relación entre los componentes, es desarrollo propio del autor.

- Los resultados con los clientes y en el mercado, esto para el caso de las empresas. Si se tratase de organizaciones de gobierno o sin fines de lucro, el concepto, más amplio, de grupos de interés suele reemplazar a los clientes.
- La gestión de los procesos, que se encarga de asegurar que las operaciones sean las mejores para generar los resultados deseados con los clientes y en el mercado y con los grupos de interés. El objetivo de la gestión de los procesos es asegurar que no exista variación en la calidad, es decir, que las cosas se hagan siempre bien para asegurar los resultados esperados.
- La gestión de los recursos humanos, que tiene como papel asegurar que el personal sea el idóneo en lo que respecta a sus competencias (que incluyen conocimientos, habilidades, valores, actitudes, el autoconcepto, motivaciones y rasgos). Además de facilitar la puesta en marcha de sistemas de organización del trabajo acordes con el modelo de negocio.
- La gestión de la estrategia, la que, a partir de un análisis tanto interno como externo, permite precisar cuáles son las capacidades organizacionales a desarrollar y establecer los objetivos y las metas desafiantes para el futuro. Además, la gestión de la estrategia debe encargarse de asegurar el despliegue o implementación de la estrategia definida.
- El manejo de indicadores de gestión, componente que tiene como objeto la identificación de mediciones para representar a los principales objetivos o propósitos organizacionales. Asimismo, comprende la realización de un efectivo seguimiento al desempeño,[3] a través de los indicadores, para orientar a la organización hacia la mejora continua.
- El liderazgo organizacional, que se ocupa de las acciones sistemáticas, para asegurar que los líderes modelen comportamientos asociados a los principales propósitos organizacionales. Con ello se busca que influyan en los colaboradores para que se enfoquen de manera decidida a alcanzar los resultados esperados para el negocio.

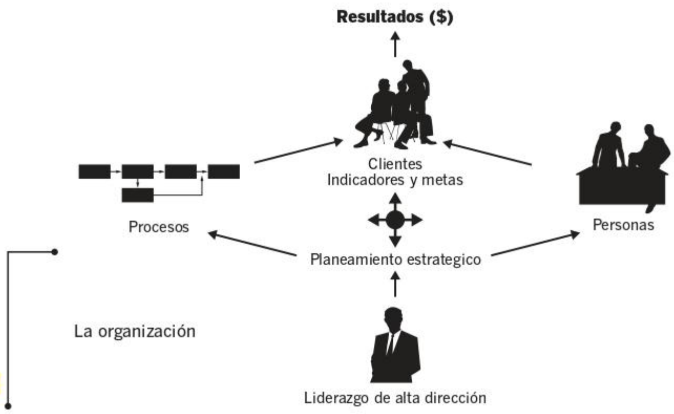


Figura 1. El enfoque sistémico de gestión.

3 "Desempeño" es un concepto asociado a los resultados alcanzados en el cumplimiento de determinados objetivos o propósitos organizacionales, ya sea, a nivel de la organización como un todo, de las áreas o gerencias, o de las personas. El nivel de desempeño alcanzado siempre se establece, a diferencia de los resultados, por comparación. Estas comparaciones pueden realizarse con base en: (1) metas propias, (2) los desempeños de organizaciones, gerencias por personas similares o, (3) con respecto a los mejores niveles de desempeño conocidos, a los que les suele llamar también como desempeños de "clase mundial". En algunas regiones de habla española el término usado para referirse a este concepto es "rendimiento", así como también, el término en inglés *performance*.

Es importante señalar que esta delimitación de componentes no tiene por qué ser única o la mejor, simplemente es la que usamos porque mejor nos ha funcionado en la práctica; sin embargo, puede haber otras maneras de estructurar un modelo de gestión en función a otros criterios.

Las características principales de este enfoque son:

+ La existencia de relaciones causa-efecto entre los componentes.
+ La interdependencia simultánea entre los componentes del sistema.
+ La aplicabilidad o pertinencia para todo tipo de organizaciones.

En las siguientes líneas entramos en detalle respecto al significado de estos tres aspectos característicos del enfoque sistémico.

– Las relaciones causa-efecto

Las relaciones de tipo causa-efecto comprendidas en el enfoque sistémico de gestión se pueden apreciar mejor partiendo de los efectos, para luego, regresar a las causas.

Así, en el caso de una organización empresarial o con fines de lucro,[4] los resultados financieros se producen cuando se logra que los clientes y clientes potenciales se comporten de la forma más conveniente para el negocio (compran, recomiendan o permanecen, por ejemplo).

A su vez, los clientes y clientes potenciales se comportan así porque la organización cuenta con procesos estructurados para inducir estos comportamientos (una buena gestión de procesos implica cuidar los detalles de producto y servicio que afectan los comportamientos deseados de los clientes y clientes potenciales). Al mismo tiempo es necesaria una gestión de recursos humanos coherente con lo que necesitan los procesos (personal bien seleccionado, capacitado o motivado, según sea el caso). La conjunción de una buena gestión de procesos y una adecuada gestión de recursos humanos, de manera simultánea, permite influir de la mejor manera en el comportamiento del cliente y del cliente potencial.[5]

Continuando con esta lógica, para alcanzar una adecuada gestión, tanto para los procesos, como para los recursos humanos,[6] es necesario planear. Es decir, decidir, en función a lo que es más conveniente para el negocio y, sobre la base de un análisis previo, cómo vamos a gestionar el negocio, qué capacidades son clave para ser desarrolladas, en qué debemos ser buenos. La pregunta clave en este punto es, ¿qué características deben tener, tanto la gestión de los procesos como la gestión de los recursos humanos, para influir de manera óptima en los clientes y clientes potenciales? Este es el rol del planeamiento estratégico: identificar cuáles son los aspectos que deben ser incorporados, reforzados o replanteados, básicamente en la gestión de los procesos y en la gestión de los recursos humanos, para asegurar un buen desempeño con los clientes y en el mercado. De esta manera debe ser posible alcanzar los objetivos desafiantes para el futuro, ya sean de tipo financiero o de cumplimiento de la misión institucional.

Los indicadores de gestión, otro de los componentes clave del modelo sistémico, entran en juego sólo cuando los propósitos u objetivos organizacionales están claros. Los indicadores sirven en la medida que representan prioridades o acuerdos importantes para la organización. Una buena gestión de indicadores ayuda a comprender la efectividad de las acciones planeadas y a evaluar la certeza de los

---

4 El mismo ejemplo aplica a organizaciones sin fines de lucro y de gobierno. Sólo es necesario reemplazar lo financiero por el cumplimiento de la misión, y a los clientes por los grupos de interés.

5 Es importante señalar que la organización espera distintos comportamientos de los clientes y de los clientes potenciales. En el caso de los clientes podría buscarse, por ejemplo, la recompra, la recomendación o la permanencia; en el caso de los clientes potenciales el comportamiento podría ser la preferencia.

6 En algunas situaciones, sin embargo, la importancia relativa entre la gestión de procesos y la gestión de recursos humanos puede variar. Este es el caso de algunas empresas industriales o de tipo productivo, altamente estructuradas, donde la gestión de los procesos suele ser algo relativamente más importante que la gestión de recursos humanos. Inversamente, en el caso de una empresa de servicios personalizados, la gestión de los recursos humanos puede resultar relativamente más importante que la gestión de los procesos. En otras organizaciones existe un equilibrio, por lo que es necesario destinar recursos similares tanto a la gestión de procesos como a la gestión de recursos humanos.

9. Conclusiones y recomendaciones

El uso de diagramas de flujo permitió entender de manera óptima las operaciones, tareas y procesos en cada una de las 3 áreas mencionadas: Mecánica Rápida, Mecánica Especializada, Latonería y Pintura. La implementación y ejecución de estudio de tiempos permitió observar y establecer las brechas entre los temparios establecidos por fábrica y los tiempos que requieren los técnicos para realizar dichas operaciones. Se llegó a la conclusión de que los temparios establecidos por la marca actualmente y los métodos de planificación permiten obtener un indicador de operatividad del taller con respecto al mantenimiento y reparación de vehículos, sin embargo, no es posible determinar la efectividad real de los técnicos sobre los tiempos de operación, para lo cual es necesario seguir trabajando en el estudio de tiempos en las 3 áreas establecidas que permitan estandarizar las operaciones.

Por otra parte, una medición real de la eficacia se lograría si la forma en la que el taller está planificando sus operaciones actualmente potencializa la utilidad de este.

Dentro la de eficiencia actual, se logró determinar que, para los mantenimientos preventivos, el taller tiene una ventaja ya que está muy por debajo del tiempo de operación establecido por fábrica lo que permite potencializar y mejorar este servicio para perfeccionar los ingresos en esta área. Sin embargo, los cambios de aceite no tienen este mismo comportamiento ya que el tiempo estándar calculado está muy cerca del tiempo establecido. Gracias a la investigación realizada se logró identificar oportunidades de mejora que permitirían impactar la eficiencia de estos procesos que se constituyen cerca del 25% del volumen total de ingreso de vehículos mensuales.

El diagnóstico de la situación del taller haciendo uso del ciclo PHVA permitió determinar el progreso de gestión de productividad en las 3 áreas del taller, así mismo mostrar la importancia

6 CONCLUSIONES Y RECOMENDACIONES

6.1 Conclusiones

Con la implementación de mejoras en la distribución de actividades del taller y la eliminación de procesos que no aportan al sistema como: el estacionamiento del vehículo, las charlas con el cliente y revisiones innecesarias; se logró una reducción de 239 a 127 minutos/tarea en la operación del cambio de banda; así mismo, la disminución de 306 a 206 minutos/tarea en la actividad de cambio de embrague. Cabe señalar que estas actividades son las de mayor complejidad en el taller y que más tiempo demandan en la reparación y mantenimiento.

A través del rediseño del modelo de servicio, el mejoramiento de la tecnología y la actualización de conocimientos, en los procesos de mantenimiento y reparación de vehículos, se logró un aumento en la productividad del taller de 0,7 a 0,9 vehículos/h-h, en el periodo de enero 2019 a diciembre del 2021.

Los retrasos en la entrega y reprocesos fueron las principales causas de quejas de los clientes; sin embargo, a través de las capacitaciones al personal en temas de: atención al cliente y mecánica básica, se mejoró el indicador de satisfacción al cliente de 3,7 a 4,2 desde enero 2021 a diciembre 2021.

La empresa no contaba con un control financiero, sin embargo, con la información disponible se estableció una línea base en el año 2019 en el cual se identificó un valor de 20,8% para el indicador de utilidad/ventas; tras las mejoras implementadas en el año 2021 incrementó su valor a un 35.6%.

ANEXO 10: ASTUDILLO, 2024 - ¿CÓMO ESTÁ LA DEMANDA DE VEHÍCULOS USADOS EN ECUADOR?

MENU EL COMERCIO

Giovanni Astudillo

Por cada **vehículo** nuevo que se comercializa en **Ecuador**, se venden de tres a cuatro **seminuevos**. Esa estadística es manejada por las empresas del sector y la proyectan en función de las transferencias de dominio comprobadas de forma legal.

La caída de las ventas de automotores nuevos en **Ecuador** incidió en el negocio de los usados. Durante el primer cuatrimestre de 2024 se comercializaron 39 974 unidades nuevas. Fueron 6 849 menos que en el mismo período de 2023.

En el caso de los seminuevos tuvo un importante crecimiento, sobre todo, en marzo de 2024, señala **Pierre Janineh**, gerente de Toma de Carros y Operaciones de **1001Carros.com**. Esta empresa tiene operaciones en todo el país.

Ana Gabriela Arcos, representante de **Patio Tuerca**, indica que en los últimos meses aumentaron las visitas a su portal web debido al mayor interés por los vehículos usados en **Ecuador**. En diciembre de 2023 registraron 646 178 visitas y en marzo de 2024 fueron 782 882.

ANEXO 11: GADATSCH, 2023 - BUSINESS PROCESS MANAGEMENT: ANALYSIS, MODELLING, OPTIMISATION AND CONTROLLING OF PROCESSES

Introduction to Business Process Management 1

Process management changes the world of work

Abstract

This introductory chapter first explains the concept and historical development of business process management. Subsequently, several basic concepts such as "function", "business process", "process", "end-to-end process" and "workflow" are defined and distinguished from each other. The conclusion is formed by review questions and an exercise.

1.1 Concept Clarification

Why do we need business process management? This question is not only asked by students of business administration who go to a corresponding lecture with great expectations, but also by experienced practitioners. A look at history can help here a little bit. Since the beginning of the 19th century, the world of work has been characterized by a strong division of labor as a result of the previous industrial revolution. The *Taylorism* played an important role here, named after the US American Frederick W. Taylor (see the original work Taylor 1903).

The *business process management* (GPM) or simply *process management* was developed at the beginning of the 1990s in order to, among other things, eliminate the negative consequences of division of labor and poor coordination.

Business process management deals with the documentation, analysis and restructuring of workflows (processes). For a long time, the term "process organization" was common in German-language literature. The documentation of the processes is also referred

A. Gadatsch, *Business Process Management*,
https://doi.org/10.1007/978-3-658-41584-6_1

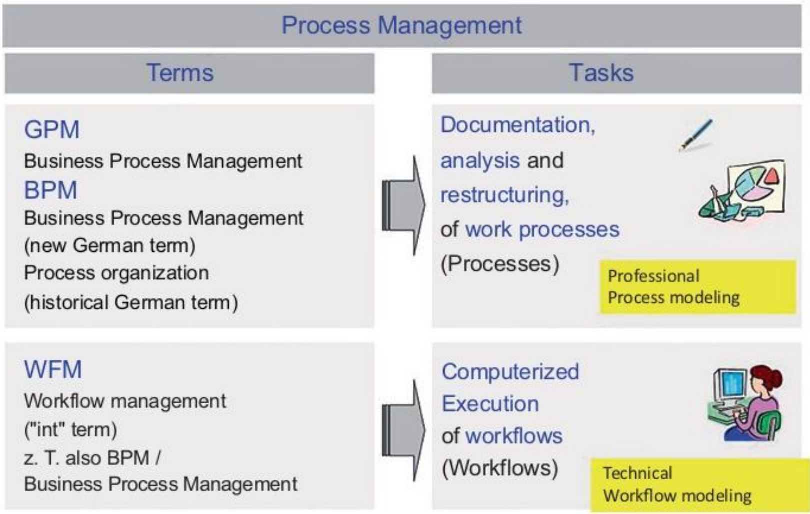


Fig. 1.1 Concept clarification process management

to as "technical process modeling". In the international environment, the term "business process management (BPM)" is common.

To be distinguished from this is the term "workflow management" (WFM), which covers the computer-supported execution of workflows (so-called "workflows"). Here one also speaks of "technical workflow modeling". In international usage, the terms "business process management and workflow management" are often not further differentiated, one usually speaks in both cases of "business process management (BPM)". Figure 1.1 shows the terminology at a glance.

1.2 Historical Development

In the development of process management, four phases of development can be identified (see Table 1.1).

I. Phase: Breakdown of work into functions (Taylorism):

The early phase of process management begins with Taylorism, named after Frederic Winslow Taylor (1856–1915), who consistently separated planning and executive activities. This classical bureaucratic organizational structure prevailed in almost all companies of the 19th century and played a central role for departments (cf. Sua-Ngam-Iam and Kühl 2021, p. 46). According to the then prevailing business paradigm, the construction and operational organization were considered separately in this phase. This was first

ANEXO 12: HUNG, 2011 - BUSINESS PROCESS MANAGEMENT AS COMPETITIVE ADVANTAGE: A REVIEW AND EMPIRICAL STUDY

Total Quality Management
Vol. 17, No. 1, 21–40, January 2006

Routledge
Taylor & Francis Group

Business Process Management as Competitive Advantage: a Review and Empirical Study

RICHARD YU-YUAN HUNG

Department of International Business, Toko University, Chia-Yi, Taiwan

ABSTRACT *From a process perspective, Business Process Management is regarded as a best practice management principle to help companies sustain competitive advantage. This paper reviews and examines the key concepts of Business Process Management and its effects on organizational performance. Process Alignment and People Involvement are two key concepts for successful implementation of Business Process Management. The purpose of this paper is to discuss how these two constructs affect organizational performance. A questionnaire was used to collect data; resulting in useful insights into the field of Business Process Management. This research contributes to a better understanding of the field of Business Process Management. The findings provide in-depth interpretation of previous studies regarding the relationship between the effects of alignment on organizational effectiveness. The results present evidence that Process Alignment and People Involvement are positively associated with organizational performance. Through regression analysis, we also confirmed that Process Alignment is a mediator that contributes to People Involvement and on Organizational Performance.*

KEY WORDS: Business Process Management, competitive advantage, process alignment, people involvement

Introduction

In order to sustain a competitive advantage and so face the rapidly increasing global competition, companies must continuously implement best practice management principles, strategies and technologies (Carpinetti *et al.*, 2003). Business Process Management (BPM) is a best practice management principle that helps companies sustain competitive advantage (Kilmann, 1995). From a process improvement perspective, BPM has become an important topic in the language of many organizations (Pritchard & Armistead, 1999). Increasingly, research has focused on developing concepts within BPM and in identifying the operational and strategic importance of business processes (Elzinga *et al.*, 1995;

Correspondence Address: Richard Y. Hung, Toko University, Department of International Business, No. 51 Sec. 2, University Road, Pu-Tzu City, Chia-Yi County 613, Taiwan. E-mail: ryyhung@mail.toko.edu.tw

1478-3363 Print/1478-3371 Online/06/010021–20 © 2006 Taylor & Francis
DOI: 10.1080/14783360500249836

Hammer, 1996; Zairi, 1997). As concepts within BPM, the alignment of business operations with strategic priorities is seen as core to competitiveness. Moreover, the importance of people's involvement, as an underlying concept, also plays an important role in the quest for long-term competitiveness.

However, the roots of Business Process Management can be traced back to the 1980s and that of TQM philosophy, and in the 1990s of Business Process Reengineering (BPR) (Hammer, 1990; Zairi & Sinclair, 1995). Previous studies (Davenport, 1993; Zairi & Sinclair, 1995) indicate that TQM is incremental, evolutionary and continuous in nature. Conversely, Process Re-engineering is radical, revolutionary and a one-time undertaking. Based on a holistic perspective, BPM integrates TQM and a Process Re-engineering approach, and can be regarded as suitable for performance improvement in most circumstances. Restated: BPM integrates TQM whilst BPR attempts to provide a model for an organization to practice.

The primary aim of BPM is to improve business processes and so ensure that the critical activities affecting customer satisfaction are executed in the most efficient and effective manner. It may involve small steps of improvement; and ongoing learning from best practices, resulting in a radical redesign of business processes to achieve superior performance (Hammer, 1996; Zairi & Sinclair, 1995).

Purpose of this Research

This research relates how well organizations manage their core processes to how well they perform. The purpose of this research is: (1) to review concepts of Business Process Management; (2) to develop constructs that contribute to our understanding of Process Alignment and People Involvement while implementing BPM; (3) to propose and examine research hypotheses based on the relationship between Process Alignment and Organizational Performance; (4) to propose and examine research hypotheses based on the relationship between People Involvement and Organizational Performance; and (5) to propose and examine the mediating impact relationship of Process Alignment between People Involvement and Organizational Performance. Specifically, this study addresses the following research questions:

(1) What is the relationship between Process Alignment and Organizational Performance, especially when organizations practice BPM? To what extent, do the components of Process Alignment affect Organizational Performance?
(2) What is the relationship between People Involvement and Organizational Performance, especially when organizations practice BPM? To what extent, do the components of People Involvement affect Organizational Performance?
(3) To what extent is Process Alignment a mediator that contributes to People Involvement and on Organizational Performance, especially when organizations practice BPM?

Theoretical Development

This section first reviews the literatures concerning Business Process Management, then Process Alignment, and finally People Involvement is examined.

Business Process Management

The term Business Process Management as a field of study is still in its infancy, yet the interest in BPM has grown steadily over recent years. However, the term 'process' is found in many disciplines. Systems thinking (Checkland, 1981), cybernetics (Beer, 1966) and systems dynamics (Senge, 1990) give a richer meaning to the term 'process'. Organizational theorists have also talked in terms of social and organizational processes (Monge, 1990). Porter's (1985) value chain emphasizes the competitive advantage of unique processes. According to Hammer (1996), thinking in terms of business processes provides a new analytic framework that helps break the mould of thinking that is based only on functional units. In turn, new team concepts, tools and methodologies are emerging to support the analysis, the improvement and management of processes. To maintain corporate competitive advantage, a process view of management is extremely important.

Business Process Management is considered to be a more general approach to organizational improvement. Zairi (1997: 78) defines BPM as:

> an approach which is dependent on strategic elements, operational elements, use of modern tools and techniques, people involvement and, more importantly, on a horizontal focus which will best suit and deliver customer requirements in an optimum and satisfactory way.

He also argues that Business Process Management has to be governed by seven rules: (1) Major activities have to be properly mapped and documented; (2) BPM creates a focus on customers through horizontal linkages between key activities; (3) BPM relies on systems and documented procedures; (4) BPM relies on measurement activity to assess the performance; (5) BPM has to be based on a continuous approach; (6) BPM has to be inspired by best practice; (7) BPM is an approach for culture change.

Hung (2001) in defining principles of Business Process Management included a holistic view – Strategic imperative – enabled by information technology, corporate-wide impact, and emphasizes cross-functional process management. More details are described as follows.

Holistic view

BPM takes a holistic view and attempts to overcome the piecemeal improvements in isolated parts of a business process that often result in sub-optimal solutions. It addresses the interdependence of strategy, people, processes and technology in achieving business objectives. According to DeToro &McCabe (1997), Business Process Management is an approach that presents a more comprehensive array of improvement options.

Strategic imperative

Davenport (1993) states that acknowledging the impact of strategic issues on process management early in strategic planning would make operational decisions easier and more consistent. BPM should focus on a coherent process to strategy, in order to add value on the customer.

Enabled by information technology

Information technology is a powerful enabler for managing business processes and in transforming business (Gates & Hemingway, 1999). Through implementing IT strategies, companies can integrate performance management systems to manage business processes.

Corporate-wide impact

BPM affects every aspect of an organization, from its structure (organized around processes) to its management (process leaders versus functional heads). The involvement of all staff in cooperative efforts must be promoted if organizations are to excel in process management (Oakland & Sohal, 1996; Waldman, 1994). When an organization adopts BPM it should be prepared to acknowledge that its impact would be corporate-wide.

Emphasizes cross-functional process management

Business process management attempts to overcome problems raised by the Taylorist view of structural specialization. Therefore, BPM includes changing the design of its organizational structure. According to Kirkham (1996), the key element of process orientation is the flattening of the hierarchy. Keen (1997) argues that most organizational processes are cross-functional and cut across traditional functional boundaries. The process-based approach attempts to improve customer focus by avoiding the limitations of managing by vertical functions (McAdam, 1996). DeToro & McCabe (1997) state that core processes are those processes that are strategically important to the organization's success, and have a high impact on customer satisfaction. Core processes describe the end-to-end work that starts from the customer and ends with the customer, and always using cross-functional activities (Hung, 2001).

BPM and TQM have many similarities. Deming's (1986) system thinking of TQM is the pursuit of internal and external cooperation. Moreover, to fit within the external environment, the dynamic equilibrium between internal processes and their environment should be maintained (Bailey, 1994). BPM provides a systematic and holistic thinking to align business processes, as does TQM (Hung, 2001).

It may be argued that much of the aforementioned share the underlying assumptions with system theory (Wang, 2004), and TQM (Deming, 1986). BPM supports the notion that every employee, not only the managers, have responsibility for supervising and controlling their work (Hung, 2001; Wang, 2004). Juran (1993) and Deming (1993) also encourage people involvement; and moreover should be considered as an important concept of BPM.

In this research, however, BPM is defined as an integrated management philosophy and set of practices that includes incremental change and radical change in business process, and emphasizes continuous improvement, customer satisfaction, and employee involvement (Ross, 1995). Organizations may take full advantage of both incremental change and radical change. Both incremental change and radical change can be viewed as parts of the continuous improvement efforts in organizations. BPM initiatives include any that seek to improve performance of a company and that comply with the principles of BPM. Continuous Improvement, Process Reengineering and Benchmarking can be classified as BPM initiatives as these approaches to improvement share the characteristics of BPM. In addition, McKay & Radnor (1998) point out that many organizations develop their own approach to managing business process, and that the important concepts of Business Process Management should include: (a) Process Alignment and (b) People Involvement.

Process Alignment

How well an organization manages the fit between its processes and its institutional elements is captured by the concept of Process Alignment. Process Alignment can be interpreted as the organizational effort needed to make processes the platform for organizational structure, for strategic planning, and for information technology (Hammer, 1996). The aim of Process Alignment is to arrange the various parts of the company to work in harmony in pursuit of common organization goals, in order to improve performance and sustain competitive advantage (Weiser, 2000). Organizational theory sees organizations as requiring designing their structures and systems to align the contingencies of environment, strategy, technology, and so on for survival and success (Daft, 1998; Lewin, 1999). Previous studies have empirically demonstrated the positive effect of alignment on organizational effectiveness (Roth *et al*., 1991).

Alignment theory (Semler, 1997) also suggested that employee behaviour is correspondent with organizational goals through structural change, strategy usage and culture transformation. Weiser (2000) posits that in order to link all areas of the organization and serve as an informational lifeline throughout the change and alignment process, the organizational structure needs to be redesigned to accommodate cross-functional requirements. Hall (2002) argues that alignment requires continual focus on customers and their constantly changing requirements and should also focus on strategic direction.

Horizontal Structure Alignment

Daft (1998) acknowledges that organizational structure must accomplish two things for the organization. It must provide a framework of responsibilities, reporting relationships, and groupings, and it must provide a template for linking and coordinating organizational elements into a coherent whole. From an institutional perspective (Powell, 1991), sees institution orientation as the extent to which a firm faces social and cultural processes demanded emanating from outside the marketplace. According to Budros (2000), organizational changes are guided by market (efficiency) forces, or non-market (institutional) forces. Empirical work on institutionalism has focused on the social and culture processes impact on organizations (Mezias, 1990; Powell, 1991). Facing the contemporary world of today, the structure of organizations is dynamic.

Early organization theorists stressed the vertical design of organizational charts and relied on vertical structures. In recent years, some scholars have argued that organizations should depart from any form of a vertical structure to a pure horizontal structure (Spector, 1999). Many companies have started to move from a relatively functional and hierarchical structure to one with a focus on cross-functional teams and flattened organizational structures (Ostroff, 1999). A horizontal management style should structure around core processes and increase the interaction of employees from different departments and foster close working relationships and better communication (Ostroff, 1999; Spector, 1999). However, some of these structural changes have been successful while others have not. Alignment theory provides an alternative theory perspective to explain the importance of structure alignment in BPM. Mintzberg's (1991) work divided the organization model context into: (a) simple style; (b) machine style; (c) professional style; (d) divisional; (e) advocacy style; and (f) missionary style. No matter what kind of organizational models, he suggested that organizational alignment is the major underpinning for the organizational models; and this can facilitate analysis of the organizational context through his model of organizational configuration whilst also supporting an analysis of approaches to strategic planning. Snow & Miles (1983) addressed an external-internal alignment model as an organizational theory to identify organization types: each with its own strategy and a particular configuration of technology, structure and process consistent with its strategy.

Strategic Alignment

In strategic theory, the external-internal alignment model is widely employed (Andrews, 1980; Snow & Miles, 1983), that focuses on organizational strategies to fit with the various parts of work in an organization. Strategic management is the process of formulating and implementing strategies to achieve strategic goals, given environmental and internal influences (Thompson & Strickland, 1999). It is oriented toward reaching long-term goals, weighs-up environmental elements, considers the organization's internal characteristics and involves developing specific strategies. In companies that practice BPM, strategies are first developed for core processes such as Customer Acquisition or Supply Chain Management. From these, functional strategies (Marketing, Sales, Distribution and Operations) are developed so that each business unit knows its contribution to core process objectives. The aim is to achieve greater alignment between functional actions and organizational goals (Hinterhuber, 1995; Lee & Dale, 1998; Zairi, 1997). Strategies for end-to-end processes that sit above and cascade into functional strategies are a defining feature of the guiding principles of Business Process Management.

Sustainable world-class performance will not occur if there is a misalignment between a firm's strategic objectives and actual market requirements. In order for a firm to compete successfully through its strategic objectives, alignment must exist between the firm's strategies, actions and performance measures (Keen, 1997; Pritchard & Armistead, 1999).

An effective strategy makes improvements and then so to does the capability of core and support processes. If changes to processes are undertaken via a BPM programme but the fit with strategy is not strong, the changes may drain value. In extreme cases, the process paradox may occur; benefits accrue as a result of process improvement but these do not translate into increases in customer and economic value (Keen, 1997). Understanding the strategic context of a BPM program is essential to maximizing the value from process improvement. (Gagnon & Dragon, 1998; Hinterhuber, 1995; Powell & Dent-Micallef, 1997, Zairi, 1997).

Information Technology Alignment

IT is a primary driver of strategic change and reshaping processes. Both of hardware and software technologies are evolving in ways that make it possible to maintain extensive amounts of information online and to be able to access this information in conjunction with the communication networks from almost any location (Wladawsky-Berger, 1999). IT also offers the capability to redefine the boundaries of markets and structural characteristics, alter the fundamental rules and basis of competition, redefine business scope, and provide a new set of competitive weapons.

Equally, IT facilitates the integration of business functions at all levels in an organization by making corporate-wide information more readily accessible (Scott-Morton, 1991). Watkins (1998) stressed that IT systems support leadership, management control, and employee participation. Ostroff (1999) advocates that changes in IT systems accompany the transformation to a horizontal management style. IT alignment to support changes in core processes is therefore critical to the implementation of BPM.

Numerous studies have shown that new IT systems alone do not produce sustainable performance, but must be carefully integrated into the way an organization operates its business processes (Gagnon & Dragon, 1998; Powell & Dent-Micallef, 1997). The alignment of IT systems with end-to-end processes is needed to obtain the full benefits from IT investment (Sohal & Ng, 1998). When a process view is taken in designing and installing an IT system, it becomes possible to reduce integration problems and enhance the performance of cross-functional, cross-divisional and even cross-company processes (Short & Venkatraman, 1992; Venkatraman *et al*., 1993). Thus, alignment theory concurs that IT implementation is a component of organizational performance (Davenport, 1993; Hammer & Champy, 1993; Ostroff, 1999; Spector, 1999; Thompson & Strickland, 1999). Grover *et al*. (1995) argue that IT as a transformational subsystem is imperative in culture transformation. Therefore, when an organization is appropriately aligned, organizational structure, strategic planning and IT correspond to organizational core processes and objectives, ensuring competitive advantage. Several studies (Burns & Wholey, 1993; Pyburn, 1983) suggest that effective alignment of strategy and IT require appropriate choices to position the firm with respect to dynamic and uncertain organizational structures. From the previous discussion the following hypothesis is derived:

Hypothesis 1. Process alignment, principally horizontal structure alignment, strategic alignment and information technology alignment, have a positive impact on organizational performance

People Involvement

In this research, People Involvement as one BPM concept has two components: executive commitment and employee empowerment. How well an organization involves people at all levels in the management of its processes has been discussed as a critical factor in organizational success (Ulrich, 1997). How well an organization involves people at all levels in the variable *People Involvement* captures the management of its processes. People Involvement is measured as the extent of active participation of all members of an organization in decision-making and problem solving. It consists of sponsorship and support from top-level management and the realignment of power, knowledge, and information to lower levels in the organization (Bounds *et al*., 1994; Keen, 1997).

Executive Commitment

Leadership is a key element for organizational success and results in greater productivity and competitive performance. The necessary commitment of senior managers in supporting change programmes has been extensively treated in the literature (Tichy & Cohen, 1997; Ulrich, 1997). Executive commitment has always played an important role in organization change activities. Business Process Management as one form of organization change needs top management support. Specifically, top management support can bring the whole organization to perform well in Business Process Management activities.

Employee Empowerment

Employee empowerment can generate greater value in organizations (Yeung & Berman, 1997). Empowerment connects the individual with others and creates a sense of personal worth that brings a sense of self-fulfilment. Pickering & Matson (1992) noted that empowerment means building a shared vision of where the organization needs to go and crafting an organizational culture and climate with operating values that enable all employees to participate actively and creatively in pursuing the firm's vision. This involves realignment of power, knowledge and information to front-line staff.

According to Lawler *et al.* (1998), employee empowerment can improve employee motivation. Motivation is the willingness to exert high levels of effort toward organizational goals in order to satisfy some individual need (Robbin *et al.*, 1994). Motivation is translated into improved performance when people have the necessary skills and knowledge to perform well and when the technology and work situation allow people to impact performance significantly. Previous studies have indicated that employee empowerment can result in positive outcomes for an organization (Arthur, 1994; Kallenberg & Moody, 1994).

The link between employee empowerment and organizational performance is also supported (Vogt & Murrell, 1990), and they suggest employee empowerment will increase employee participation and communications, to produce their best when working individually and/or as part of a team. However, as the above literature review identifies, alignment is the main feature of BPM (Hall, 2002). As a result, the following hypothesis is constructed:

Hypothesis 2. People involvement, principally executive commitment and employee empowerment, has a positive impact on organizational performance.

Hypothesis 3. Process alignment has a mediating impact on the relationship between People Involvement and organizational performance.

Methodology

Questionnaire Development and Data Collection

The questionnaire consisted of four parts, which related to the evaluation of general information about the company; how well organizational structure, information technology, and strategy aligned with core processes; commitment of top executive management and employee empowerment; and organizational performance. The items that comprise each of the six composite variables in the questionnaire are provided in Table 1.

The scales used to measure the above variables were developed based on previously discussed literature. As shown in Table 1, the Structure Alignment variable measurement was adopted from Spector's (1999) study. The scale to measure IT Alignment was adapted from Guimaraes & Bond (1996). The measures for Strategic Alignment were mainly defined from Byrne (1993) and Venkatraman *et al.* (1993). The scale to measure Executive Commitment was adapted from Ahire *et al.* (1996) and Shabana (1995). The scale to measure Employee Empowerment was adapted from Ahire *et al.* (1996) and Powell (1995). Organizational Performance was an aggregate measure that was adapted from Lawler *et al.* (1998) and Powell (1995). For each of the scale items, respondents were asked to indicate their level of agreement with the statement on a five-point Likert scale ranging from 1 being 'strongly disagree' to 5 being 'strongly agree'. Firm size

Table 1. Items of the six composite variables

Variable	Items in the variables
Horizontal Structure Alignment	High barriers between departments Frequent use of process teams Cross-functional teams have more authority in making day-to-day decisions than departmental managers Customer satisfied with response time Well practice horizontal communication A flat organizational structure Managerial tasks to front-line staff delegated
IT Alignment	Technology enabled business processes to perform well State-of-the-art technology Amount of data shared by employees increasing IT important to improvement of business processes Well integrated IT systems across functional units
Strategic Alignment	Developed strategies based on customer needs Management team identified core processes Core processes important input into strategic plan Operational improvements had direct impact on ability to compete Sufficient measures permit clear tracking of performance Current strategic plan identified actually undertaken Strategic planning process actually encourages information sharing and cross-functional cooperation
Executive Commitment	Executive has received adequate training in managing core processes Executive has sufficient knowledge on how to manage core processes Executive expressly recognizes the need to identify core processes Executive allocates adequate resources to improve core processes Executive actively communicates to employees on how best to manage core processes
Employee Empowerment	Employees increasing involvement in the way their work is planned Employees increasing autonomy in making decisions that affect work Employees are given the necessary resources to fix problems they encounter Employees encouraged to fix problems they encounter Employees interacting more with external customers
Organizational Performance	Organization's competitive position improved over last two years Productivity of employees increased over last two years Organization's profitability increased over last two years Quality of products and services improved over last two years Average cost per unit of product or service decreased over last two years

and type of industry served as control variables in this study. Firm size was the natural log of sales volume. Type of industry was classified into service and manufacturing industry. It was measured with dummy variables.

Two steps were followed in this study to establish the validity and reliability of the questionnaire. First, data were collected and analysed on the content and face validity by a panel of experts. The questionnaire was also changed per their suggestions. Second, the revised questionnaire was piloted with 90 post-graduate final year Executive MBA students from the Australian Graduate School of Management. The data collected from the pilot test, with a response rate of 52%, was coded and analysed using statistical software SPSS to find any unanticipated difficulties, however, no significant problems were found. Cronbach's alpha was used from the pilot test to determine the internal consistency of the instrument. All scales were highly reliable and consistent ($0.6298 < \alpha < 0.8389$).

The target population of this survey is the Top 1000 companies in Australia based on market capitalization, as reported by MOS Data Assembly Specialists. A preliminary opinion survey suggested that a top executive would be the most appropriate key informant. The actual number of companies surveyed (950) was fewer than 1000 because of acquisitions and mergers. Several efforts such as a telephone follow-up campaign and a follow-up fax were made to encourage the respondents to complete and return the questionnaire. A total of 333 questionnaires were received from the 950 questionnaires distributed. Within the 333 questionnaires, 77 questionnaires were discarded. This comprised 62 questionnaires that were returned unfilled, 14 that were incomplete, and one that was undeliverable. A total of 256 questionnaires were usable, which represents a response rate of 27%. This response rate compared favourably with other surveys in the Business Process Management related fields of TQM, Process Reengineering and Benchmarking, which yielded response rates of between 9% and 28% and averaged 18% (Lawler *et al.*, 1998; Zairi & Sinclair, 1995).

Of the 256 analysable questionnaires, 79.3% had practised Continuous Improvement (203 companies), 69.1% had practised Process Re-engineering (177 companies), and 76.2% of companies had practised Benchmarking (195 companies). Almost 54% of companies had performed Continuous Improvement, Process Reengineering and Benchmarking (138 companies), 92.2% of companies had performed at least one process improvement initiative (236 companies), and 7.8% did not conduct any process improvement initiative (20 companies). The Venn diagram in Figure 1 represents the number of companies performing each BPM related process improvement initiative. A total of 236 questionnaires met the criteria to be included in this research. Table 2 is a statistical description of the sample collected.

It was next necessary to know whether the effect of non-response bias is significant between those who responded early with those who responded late. We compared the total sales volume, size of organization, type of industry, and Business Process

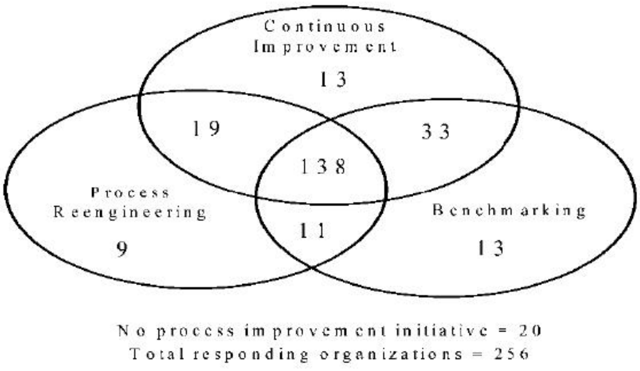


Figure 1. Number of companies performing each process improvement initiative

Table 2. Descriptive statistics

Sample firms	Frequency	%
(a) Sample distribution by industrial classification		
Manufacturing	155	65.7
Service	81	34.3
Total	236	100.0
(b) Sample distribution by age of firms		
≤20 years old	75	32.1
>20 years old	159	67.9
Missing value	2	
Total	236	100.0
(c) Sample distribution by size of firms		
≤322 Million	108	50.2
>322 Million	107	49.8
Missing value	21	
Total	236	100.0

Management related programme between those who responded early with those who responded late. Chi square tests and t tests were performed. The null hypothesis of this analysis is that an early respondent has the same characteristics as a late respondent. The observed significant level p for all variables is much higher than 0.05. This result indicates that in this research the extent of non-response bias is insignificant, and the results are generalizable to the sampling frame. The following section reports results from factor analyses and regression analyses.

Analyses and Results

Using exploratory factor analysis for Organizational Performance (OPER), Process Alignment (PALI) and People Involvement (PINV), the principal component method with varimax rotation was employed. Items in the same factor with a loading factor greater than 0.40 were grouped together.

Tables 3 and 5 show the results of factor analysis. For organizational performance (OPER) in Table 3, one factor is chosen to explain 49.86% of the variances.

For the 20 Process Alignment items, there were three factors: Horizontal Structure Alignment, IT Alignment and Strategic Alignment; their variances were 49.06%, 49.70% and 45.23%, respectively. One item in Horizontal Structure Alignment was removed due to small loading factor. The loading of each item on all common factors is shown in Table 4.

Table 5 shows the ten People Involvement items grouped into two factors: Executive Commitment and Employee Empowerment. One factor with five items was chosen to explain 62.74% of the variance in Executive Commitment. In addition, one factor with five items was chosen to explain 65.23% of the variance in Employee Empowerment.

As shown in Tables 3 to 5, the results of the factor analysis suggest that the factor structure of the questionnaire is stable and provides strong evidence for the discriminant validity of the measurement instrument.

Table 3. Factor matrix for Organizational Performance variables

Variable	Eigenvalue	% of Variance
Organizational Performance	2.48	49.57
Items		Item Loading
Organization's competitive position improved over last two years		0.804
Productivity of employees increased over last two years		0.809
Organization's profitability increased over last two years		0.766
Quality of products and services improved over last two years		0.583
Average cost per unit of product or service decreased over last two years		0.502

Table 4. Factor matrix for Process Alignment variables

Variable	Eigenvalue	% of Variance
Horizontal Structure Alignment	3.33	49.06
Items		Item Loading
High barriers between departments		0.661
Frequent use of process teams		0.581
Customer satisfied with response time		0.419
Well practice horizontal communication		0.712
A flat organizational structure		0.686
Managerial tasks to front-line staff delegated		0.533
IT Alignment	2.49	49.70
Items		Item Loading
Technology enabled business processes to perform well		0.828
State-of-the-art technology		0.812
Amount of data shared by employees increasing		0.557
IT important to improvement of business processes		0.468
Well integrated IT systems across functional units		0.781
Strategic Alignment	3.17	45.23
Items		Item Loading
Developed strategies based on customer needs		0.598
Management team identified core processes		0.789
Core processes important input into strategic plan		0.754
Operational improvements had direct impact on ability to compete		0.495
Sufficient measures permit clear tracking of performance		0.602
Current strategic plan identified actually undertaken		0.768
Strategic planning process actually encourages information sharing and cross-functional cooperation		0.648

Table 5. Factor matrix for People Involvement variables

Variable	Eigenvalue	% of Variance
Executive Commitment	3.14	62.74
Items		Item Loading
Executive has received adequate training in managing core processes		0.821
Executive has sufficient knowledge on how to manage core processes		0.841
Executive expressly recognizes the need to identify core processes		0.729
Executive allocates adequate resources to improve core processes		0.771
Executive actively communicates to employees on how best to manage core processes		0.795
Employee Empowerment	3.26	65.23
Items		Item Loading
Employees increasing involvement in the way their work is planned		0.743
Employees increasing autonomy in making decisions that affect work		0.812
Employees are given the necessary resources to fix problems they encounter		0.618
Employees encouraged to fix problems they encounter		0.674
Employees interacting more with external customers		0.569

Assessment of Reliability

The internal consistency method is by far the most common test for assessment of reliability (Cooper & Emory, 1995), and was used in this research. For assessing reliability, Cronbach's alpha was used to measure the internal consistency of the measurement scale. Although an alpha higher than 0.50 may be considered adequate, an alpha higher than 0.60 is recommended for acceptable scale reliability (Nunnally, 1978). As shown in Table 6, Crobach's alpha values are detailed for each factor and the reliability coefficients ranged from 0.661 to 0.851. Hence, the measurement instrument can be considered as reliable.

Table 6. Cronbach's Alpha Reliability Measures

Dimension/variable	Cronbach's Alpha
Process Alignment (PALI)	
Horizontal Structure Alignment (HORI)	0.661
IT Alignment (TALI)	0.734
Strategic Alignment (SALI)	0.792
People Involvement (PINV)	
Executive Commitment (ECOM)	0.851
Employee Empowerment (EEMP)	0.688
Organizational Performance (OPER)	0.735

Correlation Analyses

Table 7 contains the means and standard deviations of PALI, PINV, the three components of PALI and two components of PINV along with their correlations relative to organizational performance (OPER). We also measured and controlled for firm size and type of industry in this research.

Table 7 shows that all three variables of PALI – Horizontal Structure Alignment, IT Alignment and Strategic Alignment – were significantly positively correlated with performance, as predicted. For PINV, two variables – Executive Commitment and Employee Empowerment – were significantly correlated with performance. For control variables, firm size was significantly correlated with performance, whereas it was not significantly correlated with PALI and PINV. For type of industry variable, comparing service industry to manufacturing industry, both of them were not significantly correlated with OPER, PALI and PINV.

Linear Regression Analysis

Tables 8 and 9 show the results of hierarchical regression analyses. From Models 1 to 3, hypotheses were tested by comparing the increase in variance. Model 1 represents the regression of the control variables on organizational performance. Model 3 is the set of regression that included Process Alignment, People Involvement and two control variables (type of industry and firm size) on organizational performance.

In Table 8, Model 2a indicates adding two control variables and Process Alignment to the set of regression; in Table 9, Model 2b indicated adding two control variables and People Involvement to the set of regression. In both instances, the F test statistic for the change in the multiple squared correlation statistic (R^2) is significant.

For independent variables, the results in Table 8 indicate that Process Alignment is positively and significantly related to organizational performance ($t = 7.330$, $p < 0.01$). In addition, the results in Table 9 indicate that People Involvement is positive and significant related to organizational performance ($t = 6.124$, $p < 0.001$). Therefore, Hypothesis 1

Table 7. Means, standard deviations and correlation analysis results

Variable	Mean	SD	1	2	2a	2b	2c	3	3a	4
1 Organizational Performance	3.98	0.58								
2 Process Alignment	3.89	0.44	0.45**							
2a Horizontal Structure Alignment	3.63	0.58	0.34**	0.78**						
2b IT Alignment	3.93	0.61	0.31**	0.78**	0.34**					
2c Strategic Alignment	4.13	0.48	0.43**	0.79**	0.48**	0.46**				
3 People Involvement	3.82	0.45	0.42**	0.71**	0.58**	0.46**	0.66**			
3a Executive Commitment	3.75	0.61	0.32**	0.65**	0.51**	0.43**	0.61**	0.90**		
3b Employee Empowerment	3.88	0.45	0.41**	0.56**	0.49**	0.35**	0.49**	0.80**	0.44**	
4 Size	797.04	1299.16	0.19**	0.04				0.11		
5 Industry	1.34	0.48	−0.06	0.002				−0.08		0.04

* ≤0.05 ** ≤0.01

Table 8. Results of Regression Analysis of first order independent variable (PALI), second order independent variable (PINV) and dependent variable (OPER)

	Model 1		Model 2a		Model 3	
Variable	β	t	β	t	β	t
Industry	−0.09	−1.28	−0.09	−1.47	−0.08	−1.26
Size	0.19	2.87**	0.17	2.89**	0.16	2.71**
Process Alignment			0.44	7.33**	0.35	3.99***
People Involvement					0.13	1.44
R^2	0.04		0.24		0.25	
F	4.81**		21.91***		17.03***	
ΔR^2			0.20		0.01	
F			53.72***		2.06	

* ≤0.05 ** ≤0.01 *** ≤0.001
All two-tailed tests

and Hypothesis 2 are well supported by our data. In sum, the results indicate that each of Process Alignment and People Involvement is significantly related to organizational performance, and provide evidence of the robustness of this relationship.

Table 8 shows that when we use PALI as a first-order independent variable, PINV as second-order variable to the set of regression, from Model 2a to Model 3, the associated F-tests displayed for the change in the multiple squared correlation statistic in Model 3 is not with respect to OPER. However, Table 9 shows when we use PINV as a first-order independent variable, PALI as a second-order variable to the set of regression, from Model 2b to Model 3, the associated F-tests displayed for the change in the multiple squared correlation statistic in Model 3 are explained with respect to OPER.

Compare both Model 3 in Table 8 and Table 9, it seems there is the mediating impact of PALI with respect to OPER. As a further exploration, we found the effect of PINV on PALI is significant at the level of 0.01. Hence, the PINV is positively related to PALI.

Table 9. Results of regression analysis of first order independent variable (PINV), second order independent variable (PALI) and dependent variable (OPER)

	Model 1		Model 2b		Model 3	
Variable	β	t	β	t	β	t
Industry	−0.09	−1.28	−0.05	−0.82	−0.08	−1.26
Size	0.19	2.87**	0.15	2.40*	0.16	2.71**
People Involvement			0.38	6.12***	0.13	1.44
Process Alignment					0.35	3.99***
R^2	0.04		0.19		0.25	
F	4.81**		16.26***		17.03***	
ΔR^2			0.15		0.06	
F			37.50***		15.90***	

* ≤0.05 ** ≤0.01 *** ≤0.001
All two-tailed tests

Table 10. Conditions of mediating effect

Condition 1
(a) OPER $= 0.38$ PINV***
(b) OPER $= 0.44$ PALI**
Condition 2
PALI $= 0.71$ PINV**
Condition 3
OPER $= 0.13$ PINV $+ 0.35$ PALI***

$^{**} \leq 0.01$ $^{***} \leq 0.001$

According to Baron & Kenny (1986), the mediating effects should fulfil the following three conditions. The first condition stipulates that the independent variable and the proposed mediator must each be significantly related to the dependent variable when considered separately. The second condition requires the independent variable to be significantly related to the proposed mediator. The last condition stipulates that the relationship between the independent variable and the dependent variable should be weaker or non-significant when the proposed mediator is in the regression equation than when the proposed mediator is not in the equation. In our research, we examined three conditions and the results are showed in Table 10.

As Table 10 illustrates, condition 1 demonstrates that the independent variable (PINV) was significantly related to the dependent variable (OPER). The independent variable of the proposed mediator (PALI) was also significantly related to dependent variable (OPER). Condition 2 shows the independent variable (PINV) to be significantly related to the proposed mediator (PALI). Condition 3 shows that although PINV and PALI both have contributions to OPER, the impact of PINV to OPER is decreased. Therefore, we confirm that PALI has a mediating impact between PINV and OPER. Thus, Hypothesis 3 is supported by our data.

In Tables 8 and 9, the control variable – firm size – is significantly positively associated with OPER. We further explore the relationship between firm size and OPER. We use the means of sales volume to divide firm size into high and low groups. From an ANOVA test, we found the firm size added significantly to the proportion of performance variance explained ($F = 7.41$, $p < 0.01$). As an ANOVA result, firm size was not a predictive power for either PALI ($F = 0.00$, $p = 0.98$) or PINV ($F = 0.71$, $p = 0.40$) to OPER. The results were also confirmed from Table 7, and indicated the coefficient correlations among firm size to OPER, PALI and PINV are 0.19**, 0.04 and 0.11 respectively.

Discussions and Conclusions

The findings of this research have implications for research and practice of TQM/BPM. This research contributes to a better understanding of the field of Business Process Management. The results provide a useful insight for organization that considers implementing Business Process Management as a strategy for gaining competitive advantage. This study has supported existing knowledge. From the results, we also concluded that BPM contained two constructs – Process Alignment and People Involvement – that are positively associated with organizational performance. These findings support previous studies (Arthur, 1994; Lee & Dale, 1998; Ostroff, 1999; Powell, 1995; Zairi, 1997).

To date, numerous studies have been conducted based on the individual components of PALI (Horizontal Structure Alignment, Strategy Alignment and IT Competency Alignment) and its relations to organizational performance. No empirical evidence, except this present study, found in the literature has supported the relationship between organizational performance and the aggregate of all three concepts in one study. This research also found that People Involvement, as an aggregate concept, is positively associated with Organizational Performance. Stronger leadership and commitment from top management towards managing core processes, and providing more authority to employees to manage their work, tend to achieve the better performance of an organization. No empirical evidence in the literature supports the relationship between these two factors as an aggregate concept and business performance in one study.

When we implement Business Process Management, Process Alignment as a core concept then the various parts of the company work in harmony and also involve empowering employees in pursuit of common organizational goals. From these results, we also confirmed that Process Alignment as an aggregate concept plays an important mediating impact relationship between People Involvement and organizational performance. This finding provides in-depth interpretation of previous studies regarding the relationship between the effects of alignment on organizational effectiveness (Roth *et al*., 1991).

This research also provides empirical evidence for guiding principles that current Business Process Management literature advocates: namely, get your strategic objectives aligned with business processes, demonstrate executive commitment and empower your employee. Concentrate on these and while there is no certainty; the chances of achieving successful Business Process Management will be amplified. The implications for manages in this research are, when an organization seeks to sustain its competitive advantage, the process alignment, which includes Horizontal Structure Alignment, IT Alignment and Strategic Alignment, and PINV, which includes Executive Commitment and Employee Empowerment, should be considered. Semler (1997) also claimed that successful implementation of Business Process Management involves executive commitment and empowering employees to perform strategic action under suitable structure and use of IT.

However, the result of this research should be viewed with some caution. Our methodology adopted a cross-sectional survey type research and we can only prove association not causality. As the dependent and independent variables were collected using the same survey instrument, there is a danger that common method variance might influence the result. We have attempted to minimize the limitations by testing for common method variance (Podsakoff & Organ, 1986; Tsui *et al*., 1995).

Finally, based on the result and discussion of our research, one of our future research avenues is to extend the research model by adding a different set of independent variables and examine its impact on organizational performance.

References

Ahire, S.L. *et al.* (1996) Development and validation of TQM implementation constructs, *Decision Sciences*, 27(1), pp. 23–56.

Andrews, K.R. (1980) *The Concept of Corporate Strategy* (Homewood, IL: Irwin).

Arthur, J. (1994) Effects of human resources systems on manufacturing performance and turnover, *Academy of Management Journal*, 37, pp. 670–687.

Bailey, K.D. (1994) *Sociology and the New Systems Theory* (New York: State University of New York Press).

Baron, R.M. & Kenny, D.A. (1986) The moderator-mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations, *Journal of Personality and Social Psychology*, 51, pp. 1173–1182.
Beer, S. (1966) Decision and control: the meaning of operational research and management cybernetics, *Knowledge and Process Management*, 4(2), pp. 31–36.
Bounds, G. *et al.* (1994) *Beyond Total Quality Management: Toward The Emerging Paradigm* (New York: McGraw Hill).
Budros, A. (2000) Organizational types and organizational innovation: downsizing among industrial, financial, and utility firms, *Sociological Forum*, 15(2), pp. 273–306.
Burns, L. & Wholey, D. (1993) Adoption and abandonment of matrix management programs: effects of organizational characteristics and inter-organizational networks, *Academy of Management Journal*, 36(1), pp. 106–138.
Byrne, J.A. (1993) The horizontal corporation, *Business Week*, 20 December, pp. 76–81.
Carpinetti, L. *et al.* (2003) Quality management and improvement: a framework and a business–process reference model, *Business Process Management Journal*, 9(4), pp. 543–554.
Checkland, P.B. (1981) *Systems Thinking Systems Practice* (Chichester: Wiley).
Cooper, D.R. & Emory, C.W. (1995) *Business Research Methods*, 5th edn (Chicago, IL Irwin).
Daft, R.L. (1998) *Organization Theory and Design* (Cincinnati, OH: South-Western College Publishing).
Davenport, T.H. (1993) Need radical innovation and continuous improvement? Integrate process reengineering and TQM, *Planning Review*, 21(3), pp. 6–12.
Deming, W.E. (1986) *Out of the Crisis* (Cambridge, MA: MIT Centre for Advanced Engineering Study).
Deming, W.E. (1993) *The New Economics for Industry, Government, Education* (Cambridge, MA: MIT Centre for Advanced Engineering Study).
DeToro, I. & McCabe, T. (1997) How to stay flexible and elude fads, *Quality Progress*, 30(3), pp. 55–60.
Elzinga, J. *et al.* (1995) Business process management: survey and methodology, *IEEE Transactions on Engineering Management*, 42(2), pp. 119–128.
Gagnon, Y.C. & Dragon, J. (1998) The impact of technology on organizational performance, *Optimum*, 28(1), pp. 19–31.
Gates, W.H. & Hemingway, C. (1999) *Business @ The Speed of Thought: Using A Digital Nervous System* (England: Penguin Books).
Grover, V. *et al.* (1995) The implementation of business process reengineering, *Journal of Management Information Systems*, 12(1), pp. 109–136.
Guimaraes, T. & Bond, W. (1996) Empirically assessing the impact of BPR on manufacturing firms, *International Journal of Operations and Production Management*, 16(8), pp. 5–28.
Hall, M.J. (2002) Aligning the organization to increase performance results, *Public Manager*, 31(2), pp. 7–10.
Hammer, M. (1990) Reengineering work: don't automate, obliterate, *Harvard Business Review*, 68(4), pp. 104–112.
Hammer, M. (1996) *Beyond Reengineering: How the Process–Centered Organization is Changing Our Work and Our Lives* (New York: Harper Collins).
Hammer, M. & Champy, J. (1993) *Reengineering the Corporation: A Manifesto for Business Revolution* (New York: Harper Business).
Hinterhuber, H. (1995) Business process management: the European approach, *Business Change and Reengineering*, 2(4), pp. 63–73.
Hung, R.Y. (2001) An empirical examination of the relationship between BPM and business performance: a study of Australia's top 1000 companies (unpublished PhD dissertation, The University of Sydney).
Juran, J.M. (1993) Made in USA: a renaissance in quality, *Harvard Business Review*, July–August, pp. 42–50.
Kallenberg, A. & Moody, J. (1994) Human resource management and organizational performance, *Academy of Management Journal*, 37, pp. 948–962.
Keen, P. (1997) *The Process Edge: Creating Value Where It Counts* (Boston, MA: Harvard Business School Press).
Kilmann, R. (1995) A holistic program and critical success factors of corporate transformation, *European Management Journal*, 13(2), pp. 175–186.
Kirkham, A. (1996) Mitel Telecom business process reengineering: notes from the leading edge, in: C. Armistead & P. Rowland (Eds) *Managing Business Processes: BPR and Beyond*, pp. 7–20 (Chichester: Wiley).
Lawler III, E. *et al.* (1998) *Strategies for High Performance Organization: The CEO Report on Employee Involvement, TQM, and Reengineering Programs in Fortune 1000 Corporations* (San Francisco, CA: Jossey–Bass).

Lee, R.G. & Dale, B.G. (1998) Business process management: a review and evaluation, *Business Process Management Journal*, 4(3), pp. 214–225.

Lewin, A.Y. (1999) Prolegomena on coevolution: a framework for research on strategy and new organizational forms, *Organization Science*, 10(5), pp. 519–534.

McAdam, R. (1996) An integrated business improvement methodology to refocus business improvement efforts, *Journal of Business Process Reengineering and Management*, 2(1), pp. 63–71.

McKay, A. & Radnor, Z. (1998) A characterization of a business process, *International Journal of Operational and Production Management*, 18(9/10), pp. 924–936.

Mezias, S. (1990) An institutional model of organizational practice: financial reporting at the Fortune 100, *Administrative Science Quarterly*, 35, pp. 431–457.

Mintzberg, H. (1991) The effective organization: forces and forms, *Sloan Management Review*, 54(Winter), pp. 54–69.

Monge, P.R. (1990) Theoretical and analytical issues in studying organizational processes, *Organizational Science*, 1(2), pp. 23–34.

Nunnally, M.J. (1978) *Psychometric Theory* (New York: McGraw-Hill).

Oakland, J.S. & Sohal, A. (1996) *Total Quality Management: Text with Cases* (Melbourne: Butterworth-Heinemann).

Ostroff, F. (1999) *The Horizontal Organization* (New York: Oxford University Press).

Pickering, J.W. & Matson, R.E. (1992) Why executive development programs (alone) don't work, *Training and Development*, 46(5), pp. 91–95.

Podsakoff, P. & Organ, D. (1986) Self-reports in organizational research: problems and prospects, *Journal of Management*, 12(4), pp. 531–544.

Porter, M. (1985) *Competitive Advantage: Creating and Sustaining Superior Performance* (New York: Free Press).

Powell, T.C. & Dent-Micallef, A. (1997) Information technology as competitive advantage: the role of human, business, and technology resources, *Strategic Management Journal*, 18(5), pp. 375–405.

Powell, T.C. (1995) Total quality management as competitive advantage: a review and empirical study, *Strategic Management Journal*, 16(1), pp. 15–37.

Powell, W. (1991) Expanding the scope of institutional analysis, in W. Powell & P. Dimagio (Eds) *The New Institutionalism in Organizational Analysis*, pp. 183–203 (Chicago, IL: University of Chicago Press).

Pritchard, J.P. & Armistead, C. (1999) Business process management – lessons from European business, *Business Process Management Journal*, 5(1), pp. 10–35.

Pyburn, P.J. *et al.* (1983) Linking the MIS plan with corporate strategy: an exploratory study, *MIS Quarterly*, 7(2), pp. 1–14.

Robbin, S.P., *et al.* (1994) *Organizational Behavior: Concepts, Controversies, and Application* (Melbourne, Prentice-Hall of Australia).

Ross, J.E. (1995) *Total Quality Management: Text, Cases and Readings* (Delray Beach: St. Lucie Press).

Roth, K. *et al.* (1991) Global Strategy implementation at the business unit level, *Journal of International Business Studies*, 22, pp. 369–402.

Scott-Morton, M.S. (1991) *The Corporation of the 1990s: Information Technology and Organizational Transformation* (Oxford: Oxford University Press).

Semler, S.W. (1997) Systematic agreement: a theory of organizational alignment, *Human Resource Development Quarterly*, 8, pp. 23–40.

Senge, P.M. (1990) *The Fifth Discipline* (London: Century Business).

Shabana, A.A. (1995) Business process reengineering: testing an integration model of successful implementation. Unpublished PhD Dissertation, University of California at Los Angeles.

Short, J.E. & Venkatraman, N. (1992) Beyond business process redesign: redefining Baxter's business network, *Sloan Management Review*, 34(1), pp. 7–20.

Snow, C.C. & Miles, R.E. (1983) The role of strategy in the development of a general theory of organizations, *Advances in Strategic Management*, 2, pp. 231–259.

Sohal, A. & Ng, L. (1998) The role and impact of information technology in Australian businesses. Working Paper 46/98 (Faculty of Business and Economics, Monash University).

Spector, B.A. (1999) The horizontal organization: what the organization of the future actually looks like and how it delivers value to customers, *Academy of Management Executive*, 13(2), pp. 97–98.

Thompson, A. & Strickland, A.J. (1999) *Strategic Management: Concepts and Cases* (Boston, MA: Irwin-McGraw-Hill).

Tichy, N. & Cohen, E. (1997) How Leaders develop leaders, *Training and Development*, 51(5), pp. 58–70.
Tsui, A.S. *et al.* (1995) Dealing with discrepant expectations: response strategies and managerial effectiveness, *Academy of Management Journal*, 38(6), pp. 1515–1543.
Ulrich, S. (1997) Beyond leadership: shifting perspectives in management practice, *European Journal of Work and Organizational Psychology*, 6(4), pp. 431–451.
Venkatraman, N. *et al.* (1993) Continuous strategic alignment: Exploiting information technology capabilities for competitive success, *European Management Journal*, 11(2), pp. 139–148.
Vogt, J.F. & Murrell, K.L. (1990) *Empowerment in Organizations: How to Spark Exceptional Performance* (San Diego, CA: San Diego California University Associates).
Waldman, D.A. (1994) The contributions of total quality management to theory of work performance, *Academy of Management Review*, 19(3), pp. 510–536.
Wang, T.W. (2004) From general system theory to Total Quality Management, *Journal of American Academy of Business*, 4(1/2), pp. 394–400.
Watkins, J. (1998) *Information Technology, Organizations, and People: Transformations in the UK Retail Financial Services Sector* (London: Routledge).
Weiser, J.R. (2000) Organizational alignment: are we heading in the same direction?, *The Kansas Banker*, 90(1), pp. 11–15.
Wladawsky-Berger, I. (1999) Turning points in information technology, *IBM Systems Journal*, 38(2/3), pp. 449–452.
Yeung, A.K. & Berman, B. (1997) Adding value through human resources: reorienting human resource measurement to drive business, *Human Resource Management*, 36(3), pp. 321–335.
Zairi, M. & Sinclair, D. (1995) BPR and process management: a survey of current practice and future trends in integrated management, *Business Process Re-engineering and Management Journal*, 1(1), pp. 8–29.
Zairi, M. (1997) Business process management: a boundaryless approach to modern competitiveness, *Business Process Management Journal*, 3(1), pp. 64–80.

Chapter 1

Understanding BPM

In This Chapter

- Seeing BPM as the agile business model
- Recognizing how BPM benefits your business
- Switching to BPM
- Going beyond existing tools

Better processes produce lower cost, higher revenues, motivated employees, and happier customers. Business Process Management (BPM) is an approach that's designed to produce better processes through the combination of technology and expertise. BPM is a collaborative effort between business units and the IT world, and this effort fosters a new paradigm of efficient and logical business processes.

In this chapter, you get an introduction to BPM and see how it can benefit your business. You also see why now is a good time to implement BPM and why your existing tools just won't do the job.

BPM: The Flexible, Agile, and Scalable Business Model

In today's dynamic business environment, organizations need to be agile and flexible so they're ready to respond to whatever challenges come their way. BPM provides that agility and flexibility by giving you more direct control over your operational processes. You can make better use of technology and your entire enterprise becomes far more responsive, helping you meet your goals.

BPM helps create value for the enterprise through growth, improved performance, better productivity, higher staff effectiveness, and better customer service. All of these improvements result directly from improved processes.

TIP

BPM can help your organization become more agile in a number of different ways:

- **Increased productivity:** In today's economy you need to do more with fewer resources. Applying BPM principles helps your enterprise increase its productivity.
- **Speed to market:** When a new idea or product comes along, effective BPM helps you be one of the leaders, not one of the followers who were too late to take advantage of the new market.
- **Reaching the global market:** BPM can help you streamline and scale your supply chain operations, so you can take advantage of opportunities no matter where they may exist.
- **Achieving compliance:** Keeping up with complex compliance, regulatory, and corporate governance requirements can be very costly and time-consuming. Using BPM, you can keep these costs under control.
- **Accelerating innovation:** You need a business environment where innovation isn't only encouraged but also where innovation is a normal part of daily operations. Here, too, BPM can help make that possible.

REMEMBER

The agile and flexible organization has the ability to meet the needs of the customer and be the winner at the end of the day.

TECHNICAL STUFF

Practices such as Six Sigma and Lean Six Sigma as well as the work of quality control experts such as Deming are fully incorporated in BPM methodology. BPM actually enables you to leverage these practices to provide even greater benefits to your enterprise. For more information on Six Sigma, check out its full retail title, *Six Sigma For Dummies*.

ANEXO 14: LINDSAY, DOWNS, & LUNN, 2003 - BUSINESS PROCESSES—ATTEMPTS TO FIND A DEFINITION

Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT®

ELSEVIER

Information and Software Technology 45 (2003) 1015–1019

INFORMATION AND SOFTWARE TECHNOLOGY

www.elsevier.com/locate/infsof

Business processes—attempts to find a definition

Ann Lindsay, Denise Downs, Ken Lunn*

School of Computing and Engineering, University of Huddersfield, UK

Abstract

Definitions of business process given in much of the literature on Business Process Management are limited in depth and their related models of business processes are correspondingly constrained. After giving a brief history of the progress of business process modeling techniques from production systems to the office environment, this paper proposes that most definitions are based on machine metaphor type explorations of a process. While these techniques are often rich and illuminating it is suggested that they are too limited to express the true nature of business processes that need to develop and adapt to today's challenging environment.

Keywords: Business process; Business process management; Business process reengineering

1. Introduction

In the 21st century we are led to believe that the most valuable characteristic of an organization is its ability to adapt to the dynamic environment in which it operates. Much of the literature produced by the business process management and reengineering (BPM/R) community would suggest that implementing process orientated structures will help organizations to be more responsive to an increasingly changing environment. It is essential then that those utilizing such techniques understand adequately the nature of business processes. This paper suggests that not only are there few attempts to define process adequately, but those definitions and corresponding models most widely used are confined to a mechanistic viewpoint of process. After briefly discussing business process definitions and models the paper will propose why these models are inadequate and where research should be focused to enhance BPM techniques.

Definitions of process are usually short and succinct. Ould's book Business Processes [17] had still not attempted to define the term 'business process' by the end of the first chapter but instead lists a few key features; it contains purposeful activity, it is carried out collaboratively by a group, it often crosses functional boundaries, it is invariably driven by outside agents or customers. The rest of his book expands on these features. Jacobson [12] on the other hand quickly describes a business process as; 'The set of internal activities performed to serve a customer', Bider [3] suggests that the BPR community feel there is no great mystery about what a process is - they follow the most general definition of business processes proposed by Hammer and Champy [11] that a process is a 'Set of partially ordered activities intended to reach a goal'.

So if a business process is so easily defined that it can be described in a simple sentence and most businesses understand what a business process is, why do we need so many different ways to illustrate and model them? Why the shift towards BPR/M of the 90's? Jacobson though admitting processes are nothing new to companies explains the confusion surrounding them by their invisible nature, that they are neither named nor described.

2. Definitions and views of processes

2.1. A brief history

Communities began when individuals specialized in order to trade or sell their skills and wares in exchange for other produce or currency. Industrialization took specialization a step further, with machinery automating those specialist activities that were well defined and repetitive. As the business world evolved it was no longer adequate for companies to merely offer their goods for sale, in order to stay viable they had to keep their competitive advantage.

- in the '60s industry concentrated on how to produce more (quantity),

* Corresponding author.
E-mail address: k.lunn@hud.ac.uk (K. Lunn).

0950-5849/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0950-5849(03)00129-0

- in the '70s how to produce it cheaper (cost)
- in the '80s how to produce it better (quality)
- in the '90s how to produce it quicker (lead time)
- in the 21st century how to offer more (service) [19]

Except for the last shift with its emphasis on service the above chronology is about improvements to aspects of production. From the start of the industrial revolution the onus had been on automating and improving production efficiency and costs. The previously quoted definitions roughly describing business processes along the lines of, 'A sequence of activities which transform inputs into outputs' and their corresponding models have their routes in the scientific management techniques of the early twentieth century. A production process was seen as a linear progression taking raw material and transforming it into a finished product. Activities were studied, broken down, standardized and those activities conducive to automation were transferred to machine production. Activities too complicated or variable remained in the hands of the human operators. The characteristics of the processes studied and the descriptions proposed are, however, more indicative of production processes than a generic description of a business process. This is not to say they do not have value and their 'limited' view of a process is a substantial area of research in itself. These techniques have been used in many fields from work study and operations management, process control, business modeling and systems engineering to name but a few and are central to the business process reengineering (BPR) and business process management (BPM). The advent of automation in the office meant that the spotlight in search for efficiency and cost reduction was to encompass not only the shop floor but the office too with the same principles applied to office work.

2.2. Shifting the emphasis to office environments

The principle separating office process from production process is often based on the hypothesis that production workflow is traditionally analyzed by attention to the activities being performed whereas office systems are more goal based and people do whatever is necessary to attain a goal. The Workflow Management Coalition [20] proposed classifying processes as manual process activities and workflow process activities. The work of Gulla and Lindland [10] distinguishes between production processes and coordination processes, with traditional approaches of modeling, input—process—output, suited to modeling the chain of production processes but lacking concepts for modeling coordination activities which involve actors, information exchange and coordination structures. Ellis and Wainer [7], Bussler and Jablonski [4], Yu and Mylopoulos [21] have all suggested that analysis of activities is not appropriate for modeling office workflows or management decision-making. Yu [21] separates office processes from processes executed by machines; the latter are simply a progression of tasks whereas office workflows are about actors in social systems collaborating to achieve a goal. Joosten [13] extends this for workflow to 'a system whose elements are activities, related to one another by a trigger relation, and triggered by external events, which represents a business process starting with a commitment and ending with the termination of that commitment'.

ActionWorkflow [9] divides process into material process, information process and business process. Informational models include ERM (Entity Relationship Model) and OO (Object Orientated) models, material modeling is dataflow and process models, business models are viewed as networks referred to as workflow with interactions the main focus of attention. ActionWorkflow have worked on merging these three types of process definition into one workflow architecture. Continuing in the workflow arena other dimensions that have been analyzed are methods to help organizations specify, execute, monitor and coordinate flows of work. Finally Yu's paper [21] distinguishes between process as executed by machine and process as performed by humans. Curtis et al. [5] also made this distinction, noting that mathematically structured process programs could be used to described machine based processes whereas more flexible process scripts are used for manual tasks performed by humans who are able to interpret and enact ambiguous process descriptions. They looked at the processes employed in the software development industry and their interest was to extend traditional software modeling of data flows and transformations to the challenge of process representations that incorporated communication and coordination. Looking at a software development environment extends our enquiry of process structures into another discipline where the nature of processes is a creative, problem-solving environment. This still omits a great deal of the activities that go on within organizations to do with creative problem solving and decision making, to mention just two.

Many of the differences distinguishing office processes from production processes focus on actors, their coordination and communication roles in the activities of a process. Typical representations of these agent-related views of process are used by Role-Activity Diagrams [17] and UML collaboration diagrams, or the DEMO methodology using Business Oriented Petri Nets [1].

2.3. BPM and BPR additions to business process definitions

The fields of BPM/R strive to better understand a business's key mechanisms in order to improve, and in some cases radically change, the business performance by identifying opportunities for new business opportunities, for outsourcing, for improving business efficiency and for areas within the business where technology can be used to support business processes.

BPM/R methods have developed over the last 20 years from a variety of disciplines and as with software modeling techniques before them, have suffered from a lack of standard techniques and notations, with every BPM method using its own notation. The argument of this paper is that this lack of standardization in representing a business process is in part caused by the absence of an adequate definition of a business process.

Champions of BPR, Hammer and Champy [11] state, 'A business process is a collection of activities that takes one or more kinds of input and creates an output that is of value to the customer. A business process has a goal and is affected by events occurring in the external world or in other processes'. This contradicts slightly a later definition given by Eriksson and Penker's [8] who say that a business process emphasizes how work is performed rather than describing products or services that are a result of a process. The confusion here is whether a description of a process includes any information about the end product or service; surely in the customer focused ethos ascertaining whether a customer's requirements have been served needs repeated reference back to what is required in the first place. Another founder of the BPR movement, Davenport [6] seems to support Eriksson and Penker's view when he describes a process as; 'simply a structured set of activities designed to produce a specified output for a particular customer or market'. It implies a strong emphasis on how work is done within an organization, in contrast to a products focus on what is done. A process is thus a specific ordering of work activities across time and place with a beginning, an end and clearly identified inputs and outputs: a structure for action.

Jacobson [12] describes processes as cutting across traditional hierarchies and in order for them to succeed they involve collaboration between individuals or groups to achieve a goal. Jacobson says a customer-orientated process is expressed in terms of meeting an individual customer's needs, by concentrating on processes that provide value to customers and not merely to other parts of the business we arrive at what the organization should be doing.

When Taylor developed his theories of scientific management and refined the work activities in order to simplify, deskill and specialize, workers still carried out a process to achieve a goal. That goal was just a subgoal of BPR/M movements customer orientated goal. What was lost was the holistic view of the process from commencing with a customer and returning to the same customer when complete with someone responsible for achieving that goal for a specific individual. Thus we lost the person or group responsible for delivering the objectives of the process. Scientific management techniques segmented the process into the functional departments in the call for economies of scale and work standardization.

Perhaps here we can stop briefly to point out what the BPR/M movement adds to the scientific management view of a process. The characteristics of a process in a traditionally structured organization share many of the same attributes, but BPR/M by always keeping explicit the customer focus supports changing the organizational structure to cross-functional working practices or team structure. What is reflected here is that in the modern business world some organizations find that functional structure impedes their efficiency. This may be caused by the operators of the system, or the parts of the machine, illustrating the inadequacy of the machine metaphor when related to them, i.e. they don't work efficiently when treated as machine parts. Alternatively, it may be that the rapidly changing environment that modern companies find themselves in cannot respond adequately with a functionally structured organization.

To continue, Jacobson's [12] discussion of process explains the need for the different structure for a customer-focused process to deliver to a customer; the various players operating parts of the process collaborate to achieve the goal. Jacobson maintains traditionally structured organizations are not conducive to this way of working, and that issues such as internal politics and personal goals hamper the smooth running of the horizontal nature of a customer-orientated process. Jacobson's description concentrates on the interface between the internal business process and the customer. This gives an external view of the business, how the external actors use the business.

BPR in essence represents a repackaging of traditional techniques, the objectives of which are improvements in the measurable performance of identifiable work activities. The true nature of the BPR frame of reference is indicated by its 'slash and burn' reputation, in essence the same as Taylor's scientific management technique. Attempts have been made to soften the often heavily criticized BPR with BPM adding a humanist, team working, shared values and quality of work-life approach. However, the models capturing the detail of a BPR/M project, illustrating decomposition of tasks, linear progression, concurrent processing, are the stalwarts of IDEF3, UML activity diagrams, etc.. these are essentially flowcharts developed to show machines/computer operations can be utilized to shape and structure human activities. The literature uses the language and techniques of costs, profitability and competitive advantage failing to address the complexity and non-linear nature of much of the work carried out in organizations. BPM additions to the definition of process attempt to recognize the involvement of humans in the execution of processes but they assume the existence of crucial notions such as perfect knowledge, human participants as rational decision makers cooperating together to achieve agreed and clearly defined goals. These methods concentrate on the internal structure of the process and organization and although they often mention the need for a holistic approach the models used often do not reflect this. These models are concerned with past knowledge and promoting standardized or 'best' practice.

It has been suggested here that the definition of a business process has been constrained by objectives that seek to determine the most efficient way to carry out work, whether in a production or office environment. Not surprisingly perhaps the view held of a process, by those aiming to exact such improvements, is similar to that of a machine, especially as many of the operations subjected to such study might eventually be automated. Metaphors are often used to describe characteristics of something we don't fully understand in terms of something we are more familiar with. In describing different aspects of organizations Morgan [14] distinguishes eight metaphors to illustrate organizations. These metaphors can be further categorized into three groups: the machine group, the organism group and the mind group.

2.4. Further dimensions to process

Melao and Pidd [15] recognize the limitations of Morgan's machine metaphor to describe business processes. Morgan used metaphors to describe organization structure, Melao and Pidd use metaphors to specifically describe business processes and give four perspectives on business process to develop a conceptual framework with which to understand business processes more fully. The four categories:

- Business processes as deterministic machines
- Business processes as complex dynamic systems
- Business processes as interacting feedback loops
- Business processes as social constructs.

Most of the techniques mentioned in this paper fall into the first categories given by Melao and Pidd. The second two concepts deal with a process's interaction with its environment and ways to capture the variable or unpredictable nature of these interactions. Finally Melao and Pidd cover the human aspects of business processes offered from soft systems methods. While some of the above techniques often mention human factors of perceptions and motivation they are rarely illustrated in the corresponding models.

So far we have traveled from the production process to the office process, and we've briefly touched on the more complicated nature of processes shown in Melao and Pidd's classification. Ould tries to show that processes cover the whole spectrum of business with his core, support and management processes and his area of interest is the interactions between the people operating the process, but he fails to make the leap needed for the true complexity of what business processes could entail.

The characteristics of Morgan's machine metaphor closely correspond to those characteristics of a process modeled by such techniques as: input/output flow diagrams (IDEF0), workflow type diagrams (IDEF3, UML Activity Diagrams). That is 'machines are made up of interlocking parts that each play a clearly defined role in the functioning of the whole' [14]. Even the agent-related view of Ould's RAD diagrams demonstrate the same principles. These models are also strongly rooted in the premise that processes have a clearly defined start and the process progresses in a step wise fashion towards a define end point. Modeling methods have tended to concentrate on production and office systems. Davenport [6] spoke of a process having start and end points. In a production environment, it might be correct that we start with raw materials and end with a finished product, but this is often not the case in none production environments. Bider [2,3] recognized that this production line approach is not always satisfactory. Bider considers a view of process as a trajectory in a multi-dimensional state space where progression can be up and down, back and forward. Bider's view helps us to understand the flexibility that we must build into many of the business processes that we study today but it still assumes known start and end points and a measurable goal. Mintzberg's [16] early work on the nature of managerial practices does not fit easily into the structure of a process that has been described so far. Mintzberg's discussion of management practices speaks of the open-ended nature of management. Management processes according to Mintzberg are never ending tasks without clear milestones at which a goal could be said to have been reached or concluded. It would be a mistake to assume that management processes alone suffered from this issue. The process structures of software development have long been realized to not fit well to the linear structure often attributed to them by software development practices. Some of societies most crucial processes: in health, policing and education will not so easily fit our start and end point, flow type process description. However, they probably more than even most production or office processes need to conform to a set of auditable steps and procedures. Their goal may be indefinable; must we cure all patients, lock up all law-breakers and ensure full examination passes for all students for the measurable goal to be achieved? These processes illustrate more vividly perhaps than office processes a characteristic that most processes exhibit but which we fail to demonstrate in the modeling techniques described so far. Effective processes not only adhere to planned actions to deal with known goals, they also encompass another essential element that deals with not only the unpredictable but the unknowable.

3. Conclusion

Meloa and Pidd's [15] conceptualizations of process were efforts to deal with some of its other aspects that restrict the effectiveness of current modeling techniques. Models are simplifications in order to bring clarity and understanding to some aspect of a problem where there is complexity, uncertainty, change or assumptions. The problem in reality is complex and more variability exists than can be modeled. Both the environment the process must operate in and the process itself are not static, and changes in either one could affect changes in the other.

Underlying the mechanistic view of process is also a number of assumptions, such as perfect knowledge of the human actors involved, that humans work in a rational and logical way and that problems have a solution. Melao and Pidd's conceptualization recognizes that whilst the mechanistic/deterministic view of process gives a rich opportunity to model its tangible aspects failure to appreciate the limitations of these models can be dangerous.

Modeling gives a snapshot of what is perceived at a point in time. Sustainable business processes carried out by human operators are a balancing act between learning from the past and experimenting with and adapting to the future, and between rules and constraints versus freedom and flexibility. BPM definitions and models described here have all depended on past knowledge, what is already known, and the attempt to impart this knowledge through best practice. Process models are currently best used to represent the internal elements of business processes; for example the activities needed and their dependencies, the dataflow, the roles and actors involved, and the goals. None of the techniques described above can incorporate what Senge [18] refers to as 'The learning organization' but if organizations are to survive in the long term their processes have to be responsive and adaptable.

References

[1] J. Barjis, J. Dietz, T. Galatonov, Language Based Requirements Engineering Combined with Petri Nets, EMMSAD, Toronto, Canada, 2002 (2002).
[2] Bider, Business process modelling—concepts, Proceedings of PBPM'00, 2000.
[3] Bider, et al., Goal-Oriented Patterns for Business Processes, GBPM'02 London, 2002.
[4] C. Bussler, S. Jablonski, An approach to integrate workflow modelling and organization modelling in an enterprise, Proceedings of third Workshop on Enabling Technologies. Morgantown West Virginia (1994) 81–95.
[5] B. Curtis, Kellner, I. Marc, J. Over, Process modelling, Communication of the ACM 35 (9) (1992).
[6] T. Davenport, Process Innovation, Harvard Business School Press, Cambridge, MA, 1993.
[7] C.A. Ellis, J. Wainer, Goal-based models of collaboration, Collaborative Computing 1 (1) (1994) 61–86.
[8] H. Eriksson, M. Penker, Business Modeling with UML, Wiley, 2000, ISBN 0-471-29551-5.
[9] Flores, R.F., The value of a methodology for workflow, http://www.actiontech.com/market/papers/method5.html, 1995.
[10] J.A. Gulla, O.I. Lindland, Modelling Cooperative Work for Workflow Management, CaiSE'94, Utrecht, The Netherlands, 1994.
[11] M. Hammer, J. Champy, Re-engineering the Corporation; A Manifesto for Business Revolution, Harper Business, New York, 1993.
[12] I. Jacobson, The Object Advantage, Addison-Wesley, 1995, ISBN 0-201-42289-1.
[13] S. Joosten, Trigger Modelling for Workflow Analysis: Workflow Management—Challenges, Paradigms and Products, Con'94 Linz, Austria, 1994.
[14] G. Morgan, Images of Organization, Sage Publications, 1986.
[15] N. Melao, M. Pidd, A conceptual framework for understanding business processes and business process modelling, Information Systems Journal 10 (2000) 105–129.
[16] H. Mintzberg, The Structuring of Organizations, Prentice-Hall, 1979.
[17] Ould, A. Martyn, Business Processes: Modelling and Analysis for Re-engineering, Wiley, 1995, ISBN 0-471-95352-0.
[18] P.M. Senge, The Fifth Discipline: The Art and Practice of the Learning Organisation, Century Business, 1990, ISBN 0-7126-5687-1.
[19] Tersine, R., Open University, www.ou.edu/class/tersine.
[20] Workflow Management Coalition, Glossary—A Workflow Management Coalition Specification, Brussels Belgium, 1994.
[21] E. Yu, J. Mylopoulos, From E-r to A-R modelling strategic actor relationships for business process reengineering, Proceedings of Conference on Entity Relationship Approaches ER'94, 1994, 548–565.

ANEXO 15: WEINZIERL, ZILKER, DUNZER, & MATZNER, 2024 - MACHINE LEARNING IN BUSINESS PROCESS MANAGEMENT: A SYSTEMATIC LITERATURE REVIEW

Expert Systems With Applications 253 (2024) 124181

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Machine learning in business process management: A systematic literature review

Sven Weinzierl [a,*], Sandra Zilker [b], Sebastian Dunzer [a], Martin Matzner [a]

[a] Institute of Information Systems, Friedrich-Alexander-Universität Nürnberg-Erlangen, Germany
[b] Professorship for Business Analytics, Technische Hochschule Nürnberg Georg Simon Ohm, Germany

ARTICLE INFO

Keywords:
Business process management
BPM lifecycle
Machine learning
Deep learning
Literature review

ABSTRACT

Machine learning (ML) provides algorithms to create computer programs based on data without explicitly programming them. In business process management (BPM), ML applications are used to analyse and improve processes efficiently. Three frequent examples of using ML are providing decision support through predictions, discovering accurate process models, and improving resource allocation. This paper organises the body of knowledge on ML in BPM. We extract BPM tasks from different literature streams, summarise them under the phases of a process's lifecycle, explain how ML helps perform these tasks and identify technical commonalities in ML implementations across tasks. This study is the first exhaustive review of how ML has been used in BPM. We hope that it can open the door for a new era of cumulative research by helping researchers to identify relevant preliminary work and then combine and further develop existing approaches in a focused fashion. Our paper helps managers and consultants to find ML applications that are relevant in the current project phase of a BPM initiative, like redesigning a business process. We also offer – as a synthesis of our review – a research agenda that spreads ten avenues for future research, including applying novel ML concepts like federated learning, addressing less regarded BPM lifecycle phases like process identification, and delivering ML applications with a focus on end-users.

1. Introduction

Organisations have made great strides in digitising their business processes (Beverungen et al., 2021), so information systems now produce large amounts of process data (van der Aalst, 2016). BPM research offers approaches to create value from such data (van der Aalst, 2016), including process mining (e.g. van der Aalst et al., 2011), business activity monitoring (e.g. McCoy, 2002), predictive business process monitoring (e.g. Grigori et al., 2004), and anomaly detection (e.g. Bezerra, Wainer, & van der Aalst, 2009). ML is popular as an analytical capability at the core of BPM approaches. Because of the increased availability of event data, the emergence of off-the-shelf ML libraries, and advances in hardware, ML approaches are increasingly used to solve BPM tasks. According to Mitchell (1997, p. xv), "[t]he field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience". ML provides algorithms that identify and learn patterns from data (e.g. examples or observations) and capture those in mathematical models (e.g. functions) through automated improvement procedures (Bishop, 2006).

Before ML can be used in applications tailored to specific BPM tasks (e.g. anomaly detection), underlying ML models are developed by completing the following four steps: Data input, feature engineering, model building, and model assessment (Goodfellow, Bengio, & Courville, 2016; Janiesch, Zschech, & Heinrich, 2021; Liu & Motoda, 1998). Once these models are developed, ML applications can create value from process data, such as cost reduction or risk mitigation (Márquez-Chamorro, Eduardo and Resinas, & and Ruiz-Cortés, 2017). In short, ML supports organisations in improving their business processes (Mendling, Decker, Hull, Reijers, & Weber, 2018).

While ML in BPM bears considerable potential for improving organisational operations, we believe that an overview of ML applications in BPM can help scholars propose contributions in this research area and practitioners find existing ML applications that address a problem they are facing. This systematic literature review contributes to research and practice in the following five ways:

1. We describe BPM lifecycle phases and define BPM tasks in view of ML applications.

* Corresponding author.
E-mail addresses: sven.weinzierl@fau.de (S. Weinzierl), sandra.zilker@th-nuernberg.de (S. Zilker), sebastian.dunzer@fau.de (S. Dunzer), martin.matzner@fau.de (M. Matzner).

https://doi.org/10.1016/j.eswa.2024.124181
Received 29 June 2023; Received in revised form 20 April 2024; Accepted 7 May 2024
Available online 22 May 2024

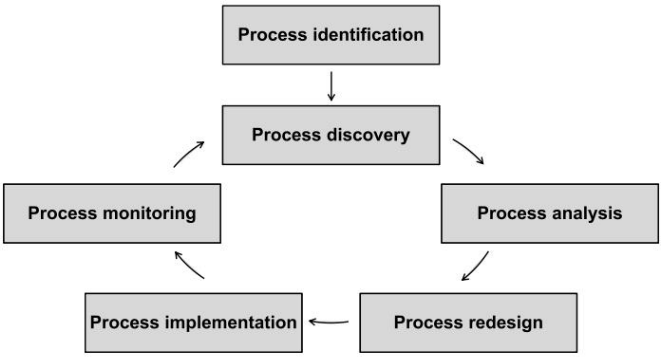


Fig. 1. BPM lifecycle based on Dumas et al. (2018).

2. We present an overview of ML applications in the field of BPM.
3. We provide a technical summary of ML applications of each BPM lifecycle phase.
4. We derived findings from our literature review.
5. We set up a future research agenda for advancing research on ML applications in BPM.

The remainder of this paper is structured as follows. After we present the core concepts of BPM and ML in Section 2, we provide an overview of related literature reviews in Section 3. Then, in Section 4, we describe our applied research method, after which we provide an overview of ML applications in BPM, and describe BPM lifecycle phases and define BPM tasks in the view of ML applications (Section 5). Subsequently, in Section 6, we synthesise the results from the previous section and present findings, which we derived from the synthesised results. In Section 7, we present directions for future research, implications for research and practice, and limitations of our literature review.

2. Background

This paper focuses on applications based on ML models in the BPM domain. Therefore, in this section, we provide overviews of BPM and its lifecycle, as well as the process of ML model development in general.

2.1. BPM and its lifecycle

BPM is a management discipline aiming to increase a business's competitive advantage by facilitating continuous improvement in organisational operations (Trkman, 2010). To achieve this, BPM proposes management practices that cultivate an end-to-end view of customer-oriented processes across functional boundaries (Trkman, 2010). BPM tasks are classified along lifecycle models to analyse, redesign, implement, and monitor business processes continuously (De Morais & Kazan, 2014). For this paper, we align with Recker and Mendling (2016) in using the BPM lifecycle of Dumas, La Rosa, Mendling, and Reijers (2018) to classify existing ML applications in BPM tasks. We also consider definitions from other BPM lifecycle models in describing each of the phases (e.g. De Morais & Kazan, 2014; Houy, Fettke, & Loos, 2010; zur Muehlen & Ho, 2005). BPM lifecycle model of Dumas et al. (2018) has six phases through which (except *process identification*, the entry point to BPM) business processes pass multiple times to facilitate constant improvement (see Fig. 1).

Process identification. The process identification phase identifies, delimits, and interrelates processes and their stakeholders. Taking a cross-functional perspective allows processes' architectures to be set up (Dumas et al., 2018) that aid in identifying the processes that should be improved along the lifecycle (Houy et al., 2010; Trkman, 2010). An activity at this stage includes analyses of process environment and organisation (zur Muehlen & Ho, 2005).

Process discovery. The process discovery phase documents the current state of processes in the form of as-is process models.[1] To enhance documentation, organisations may, for instance, specify a process in detail, model it in formal process-modelling languages, and conduct process walkthroughs (Houy et al., 2010; zur Muehlen & Ho, 2005; Zairi, 1997).

Process analysis. In the process analysis phase, the resulting process documentations and as-is models are assessed to identify the issues related to a process (Dumas et al., 2018). Examples of activities in this phase are simulating a process, calculating cost and cycle-time, applying process mining, and defining target metrics (De Morais & Kazan, 2014; zur Muehlen & Ho, 2005; Zairi, 1997).

Process redesign. Based on the results of the process analysis phase, the process redesign phase elaborates changes to processes that will improve them (Dumas et al., 2018), resulting in redesigned to-be process models that include information for process operation (De Morais & Kazan, 2014; Houy et al., 2010). Process analysis tools can help to determine process changes that should be implemented (Dumas et al., 2018).

Process implementation. Redesigned processes must be embedded in organisations' information systems (zur Muehlen & Ho, 2005) by using organisational change mechanisms to facilitate work according to redesigned processes and adjusting the IT systems required for the redesigned processes to be executed. In short, the process implementation phase yields executable process models (Dumas et al., 2018).

Process monitoring. During the execution of a redesigned process, conformance and performance are controlled using data about process executions (Houy et al., 2010). Corrective actions can be taken if errors occur or a process faces performance issues. However, if new issues arise, the lifecycle must be repeated (Dumas et al., 2018).

2.2. ML model development

The process of developing an ML model consists of four phases, as depicted in Fig. 2, enriched with relevant concepts (Goodfellow et al., 2016; Janiesch et al., 2021; Liu & Motoda, 1998). In what follows, these four phases are described.

[1] Other lifecycle models refer to this phase as process design (De Morais & Kazan, 2014).

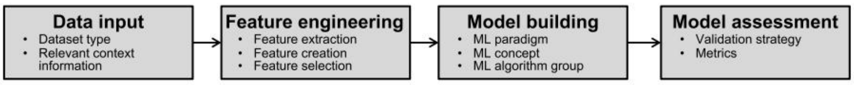


Fig. 2. ML model development phases based on Janiesch et al. (2021), Goodfellow et al. (2016), and Liu and Motoda (1998).

Data input. Data of various types can be inputs to ML models. For example, a dataset can consist of tabular, event log, or text data. The perspectives to which the event characteristics refer, such as time, resources, or data flow, can also be included for event logs (de Leoni, van der Aalst, & Dees, 2016).

Feature engineering. Feature engineering, which consists of feature extraction, creation, and selection, is performed to transform and simplify data, thus facilitating model building (Liu & Motoda, 1998). Generally, a feature is an independent variable in a dataset (Hastie, Tibshirani, & Friedman, 2009). Feature extraction retrieves features from the data input and encodes them (e.g. categorical features as one-hot encoded features), (manual) feature creation adds additional features (e.g. temporal features created based on the control-flow and timestamp information of an event log), and feature selection removes features that, for example, have little effect on the learning target.

Model building. ML algorithms are applied to build mathematical models based on data (Bishop, 2006). As the research field ML provides many ML algorithms, we differentiate them based on three dimensions (Janiesch et al., 2021): The *ML paradigm* they follow, a dimension that distinguishes among the approaches supervised, unsupervised, and reinforcement learning and those in-between (i.e. semi- and self-supervised learning); the *ML concept*, which refers to the approach (e.g. deep learning) an ML algorithm uses to address a learning problem in terms of such aspects, as the expected input data, the number of ML models, or the learning target; and the *ML algorithm group*, referring to how the algorithm builds models from data. Algorithms of the same group follow similar principles for model building, including model structure and parameter fitting. Table 1 shows an overview of the ML dimensions.

Model assessment. Several validation strategies can be used to test model generalisability, and metrics can be used to measure model properties (e.g. Hastie et al., 2009). While validation strategies include split validation and cross-validation, metrics can refer to quality (e.g. accuracy for classification models (Ferri, Hernández-Orallo, & Modroiu, 2009)) and runtime (e.g. training time for model training (Wang, Wong, Ding, Guo, & Wen, 2012)).

3. Related literature reviews

Previous literature reviews in BPM consider ML an approach to gain insights from process data. These reviews investigate the application of certain ML types for certain BPM lifecycle phases (e.g. process monitoring) and approaches (e.g. process mining). One group of literature reviews focuses on predictive business process monitoring, where ML is an approach for building predictive models. Márquez-Chamorro, Eduardo and Resinas, and and Ruiz-Cortés (2017) provide a global overview of the domain, while Di Francescomarino, Ghidini, Maggi, and Milani (2018) investigate algorithms' tasks, input data, families, and tools. Verenich, Dumas, La Rosa, Maggi, and Teinemaa (2019), Teinemaa, Dumas, La Rosa, and Maggi (2019), and Stierle, Brunk, et al. (2021) focus on techniques for predicting remaining time, techniques for classification-based process outcome prediction, and techniques that use explainable artificial intelligence (XAI) approaches, while Neu, Lahann, and Fettke (2021) and Rama-Maneiro, Vidal, and Lama (2021) conduct a systematic literature review on deep-learning approaches for predictive business process monitoring.

Previous literature reviews also address process mining. Tiwari, Turner, and Majeed (2008) perform a literature review on process mining, specifically process discovery. Maita, Martins, Paz, Peres, and Fantinato (2015) conduct a systematic literature review on applying NNs and support vector machines (SVMs) for data-mining tasks in process mining. Taymouri, La Rosa, Dumas, and Maggi (2021) conduct a systematic literature review on analysis methods for process variants, and consider ML a family of algorithms used in these methods. Folino and Pontieri (2021) perform two systematic literature reviews, one for AI -based process-mining approaches exploiting domain knowledge; second, for process-mining approaches that address auxiliary AI tasks jointly with target process-mining tasks. Both consider ML a subset of AI. Herm, Janiesch, Reijers, and Seubert (2021) conduct a literature review on intelligent robotic process automation (RPA), in which the authors consider ML an approach for transforming symbolic RPA into intelligent RPA. Wanner, Wissuchek, and Janiesch (2020) focus on the combination of ML and complex event processing in their literature review. Finally, Ko and Comuzzi (2023) perform a systematic literature review on anomaly detection for business process event logs. Table 2 provides an overview of related (systematic) literature reviews and positions our work in this context.

Our review has a more general scope and a broader focus than the other reviews we mention. The scope of our review is general, as we do not limit the time horizon because we want to consider all BPM papers that propose ML applications, regardless of when they were published. The focus of our review is broad, as we consider ML applications for BPM that rely on any type of ML and can tackle any BPM task, regardless of the phase of the BPM lifecycle to which they belong. Conducting a literature review with this scope and focus allows us to give an overview of how ML has been applied in BPM, describe BPM lifecycle phases and define BPM tasks in the view of ML applications, provide a technical summary of ML applications of each BPM lifecycle phase, derive findings, and set up an agenda that can advance research on ML applications in BPM.

4. Research method

We conducted a systematic literature review using a descriptive perspective and provide guidance based on the review's results (Paré, Trudel, Jaana, & Kitsiou, 2015). We follow vom Brocke et al. (2015) and the concept-centric notion of Webster and Watson (2002). Further, in line with Beese, Haki, Aier, and Winter (2019), we carried out the review in two steps, first applying a structured *search process* to gather relevant papers before developing a coding scheme and using a *coding process* following Recker and Mendling (2016).

4.1. Search process

We chose a systematic, sequential procedure to find a representative set of papers for our review (Cooper, 1988; vom Brocke et al., 2015). We used the databases *Scopus*, *IEEE Xplore*, and *Web of Science* to conduct the search process, as the combination of these databases covers a wide range of academic papers in the BPM domain. We defined the keywords for our search string based on our addressed research gap and the dimensions described in Section 2.1 and Section 2.2 combining BPM-related keywords with ML-related keywords (vom Brocke et al., 2015). The search string is depicted in Fig. 3 (see Appendix A for a detailed description of the creation of the search string).

We also screened the main proceedings of topical conferences (i.e. the *International Conference on Business Process Management (BPM)* and the *International Conference on Process Mining (ICPM)*). Our search and screening, executed in July 2022, retrieved 2,270 papers. Next, we

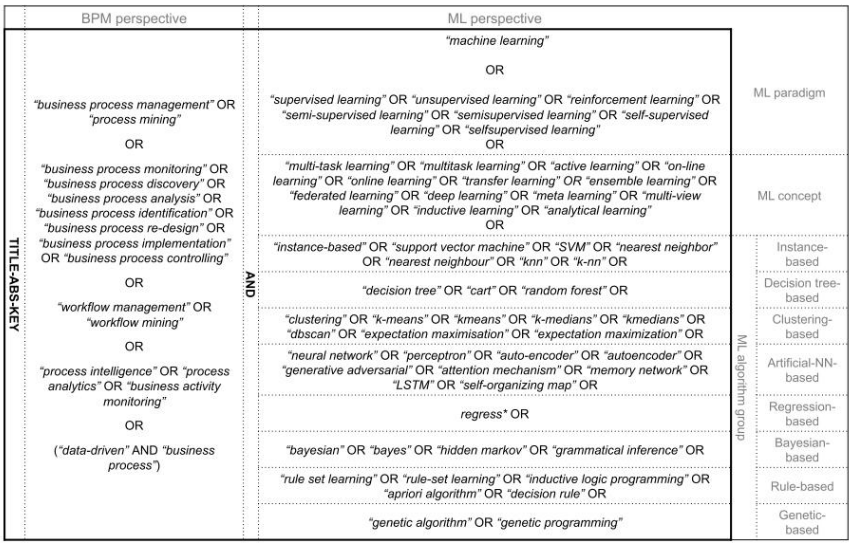


Fig. 3. Search string (in Scopus syntax).

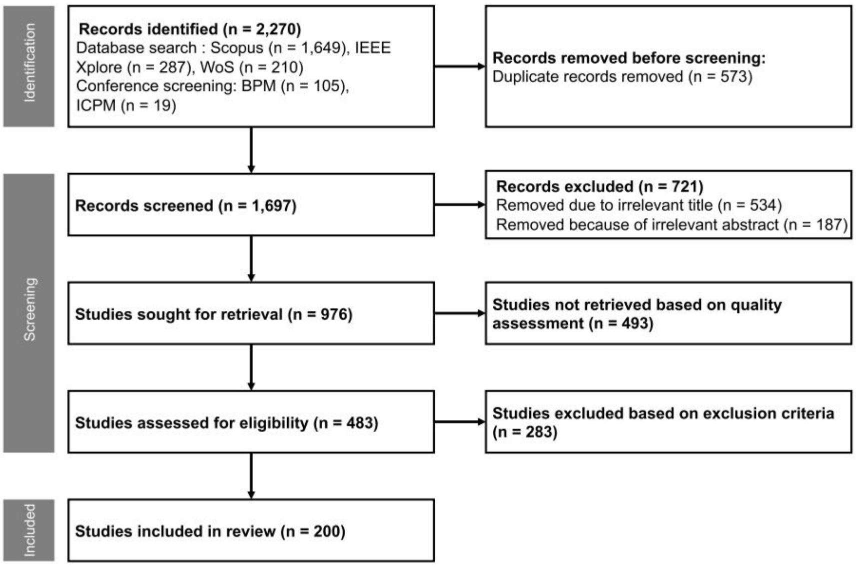


Fig. 4. Phases of the search process following the PRISMA framework.

Table 1
Overview of ML dimensions.

	Dimension	Description
ML paradigm	Supervised learning	The algorithm learns from examples (Alpaydin, 2014); that is, it learns from data instances that are labelled with their correct output (Kotsiantis, Zaharakis, & Pintelas, 2006).
	Semi-supervised learning	The algorithm uses labelled input data and unlabelled input data for model training (Zhu, 2005), which is more effective than unsupervised learning and needs less human effort than supervised learning.
	Reinforcement learning	The input data for reinforcement learning algorithms is generated through the algorithm's interactions with its environment and changes over time (Alpaydin, 2014). Hence, the algorithm relies on feedback that is received during training and is rewarded for steps towards or achieving a desired objective and punished otherwise (Brynjolfsson & McAfee, 2017; Kaelbling, Littman, & Moore, 1996; Russel & Norvig, 2016). Therefore, the algorithm is like an agent that interacts with a dynamic environment via perception and action (Kaelbling et al., 1996).
	Self-supervised learning	Using unlabelled input data (Jaiswal, Babu, Zadeh, Banerjee, & Makedon, 2021; Jing & Tian, 2021) and training an neural network (NN) model are based on pseudo-labels created from input data to solve a "pretext task'' (e.g. predicting a word based on its surrounding words in a sentence). Then the model is used to solve a "downstream task" using supervised or unsupervised learning.
	Unsupervised learning	Identifying and learning inherent patterns in the input data is based on structural properties (Jordan & Mitchell, 2015). In contrast to supervised learning, there are no labelled examples for model training (Russel & Norvig, 2016).
ML concept	Multi-task learning	The algorithm learns a model that addresses multiple related problems (e.g. Zhang & Yang, 2021).
	Active learning	An accurate model can be learned with a low volume of labelled training instances if an ML algorithm is allowed to choose the training data from which it learns (e.g. Settles, 2009).
	Online/ incremental learning	The algorithm uses the available data and updates the model before a prediction is required or after the last observation is made (e.g. Blum, 1998).
	Transfer learning	The algorithm first learns a model on a task. Then part or all of the model is used as the starting point for a related task (e.g. Pan & Yang, 2009).
	Ensemble learning	The algorithm learns two or more models on the same training set and the outputs from all models are combined (e.g. Dietterich, 2002).
	Deep learning	The algorithm learns multiple representations instead of a single representation to identify complex structures in data (e.g. LeCun, Bengio, & Hinton, 2015).
	Meta learning	The algorithm learns from the output of other ML algorithms that learn from a training set (e.g. Nichol, Achiam, & Schulman, 2018).
	Federated learning	Multiple data owners collectively learn and use a shared model while keeping all of the local training data private (e.g. Yang, Liu, Chen, & Tong, 2019).
	Multi-view learning	The algorithm learns one or more models from multiple views of training data (e.g. Xu, Tao, & Xu, 2013).
ML algorithm group	Instance-based	ML algorithms build models directly from the training instances (e.g. support vector machine (Cortes & Vapnik, 1995)).
	Decision tree-based	ML algorithms build tree-structured models, in which leaves represent class labels and branches represent conjunctions of the features that lead to these labels (e.g. classification and regression trees (Breiman, Friedman, Olshen, & Stone, 1984)).
	Clustering-based	ML algorithms build models to discover natural groups (i.e. clusters) in the data's feature space (e.g. k-means (Forgy, 1965)).
	Artificial-neural-network-based	ML algorithms build models consisting of one or more neurons that are connected via edges and structured into one or more layers from input to output (e.g. multi-layer perceptron (Rumelhart, Hinton, & Williams, 1985)).
	Regression-based	ML algorithms build models by separating data points through a fitted regression line (e.g. logistic regression (e.g. Hastie et al., 2009)).
	Bayesian-based	ML algorithms build models using Bayesian statistics (e.g. naïve Bayes (e.g. Lewis, 1998)).
	Rule-based	ML algorithms build models as a set of relational rules that represent knowledge (e.g. RIPPER (Cohen, 1995)).[a]
	Reinforcement-learning-based	ML algorithms build models on how an agent should behave in a particular environment by performing actions and observing the results (e.g. Q-learning (Watkins, 1989)).
	Genetic-based	ML algorithms build models by iteratively updating a population (pool of hypotheses), evaluating each population member based on a fitness measure, and selecting the best fitting members to produce the next generation (e.g. genetic algorithm (e.g. De Jong, Spears, & Gordon, 1993)).

[a] In this study, we associate the term ''rule-based'' with rule-based ML (Fürnkranz, Gamberger, & Lavrač, 2012), where some form of a learning algorithm is applied to automatically identify a set of relational rules for pattern detection or prediction, and not traditional rule-based artificial intelligence (AI) systems (Hayes-Roth, 1985), where a human expert with domain knowledge manually constructs a set of rules for knowledge representation or reasoning.

proceeded in a structured manner to determine the final set of relevant papers. The search process is depicted in Fig. 4.

After removing duplicates, we scanned the title and abstract of each paper for relevance to our topic and checked the papers that remained for whether they met our quality criteria (Okoli, 2015), for example, whether they meet a length requirement of at least four pages. Lastly, to assess eligibility, we screened the full text of each remaining paper in terms of our defined exclusion criteria (vom Brocke et al., 2015). The quality and exclusion criteria can be found in Appendix A. At the end of this process, we had 200 papers.

4.2. Coding process

We followed Hruschka et al. (2004) adjusted to three coders by first developing an initial coding scheme, coding the entire sample second, and, third, reconciling the coders' results. In the first step, we adapted Recker and Mendling's (2016) coding scheme and Janiesch et al.'s (2021) ML-model development process to set up the BPM-related and ML-related dimensions of our coding matrix (Webster & Watson, 2002). For the ML-related dimensions, we used the categories as defined in the search string (see Appendix A). Two authors conducted two coding iterations of ten randomly sampled papers per iteration per person. After each iteration, the authors reflected on the coding scheme's sufficiency and completeness and added or adapted dimensions and concepts to better fit the review's objectives (Webster & Watson, 2002). The coders also defined each concept textually to ensure a common understanding, which finalised the coding scheme.

In the second step, three authors conducted the final coding, analysing and coding each of the 200 papers in the final set. The papers were distributed equally among the coders to deal with the large number of papers. To clarify each paper's positioning along the BPM lifecycle and its goal, the paper's BPM aim, such as predicting the next activity or predicting remaining time, was extracted as free text. After all papers were coded, one researcher clustered the tasks (e.g. predictive business process monitoring) per lifecycle phase (Webster & Watson, 2002).

In the third step, the different codes were reconciled in a single table. Since there were multiple coders involved, we conducted an inter-coder reliability analysis using 25 randomly selected papers to ensure consistency among the coders. We calculated the percentage

Table 2
Overview of related literature reviews.

Reference	Scope of review	Focus of review
Tiwari et al. (2008)	**General:** Papers published (1998) – 2005	**Broad:** Process mining, specifically process discovery
Maita et al. (2015)	**General:** Journal and conference papers published 2004–2014	**Specific:** NNs and SVMs in process mining
Márquez-Chamorro, Eduardo and Resinas, and and Ruiz-Cortés (2017)	**Specific:** Journal and conference papers, and book chapters published 2010–2017	**Specific:** Predictive business process monitoring
Di Francescomarino et al. (2018)	**General:** Journal and conference papers published (2005) – 2018	**Specific:** Predictive business process monitoring
Verenich et al. (2019)	**General:** Papers published 2005–2017	**Very specific:** Remaining-time-prediction task in predictive business process monitoring
Teinemaa et al. (2019)	**General:** Papers published (2005) – 2017	**Very specific:** Outcome prediction task in predictive business process monitoring
Stierle, Brunk, et al. (2021)	**Specific:** Journal and conference papers published (2014) – 2020	**Very specific:** XAI approaches used in predictive business process monitoring
Neu et al. (2021)	**Very specific:** Papers published (2017) – 2020	**Very specific:** Deep-learning approaches used in predictive business process monitoring
Rama-Maneiro et al. (2021)	**Very specific:** Papers published (2017) – 2020	**Very specific:** Deep-learning approaches used in predictive business process monitoring
Taymouri, La Rosa, Dumas, and Maggi (2021)	**General:** Papers published 2003–2019	**Specific:** Business process variant analysis
Folino and Pontieri (2021)	**General:** (i) Journal and conference papers, and book chapters published (2009) – 2020; (ii) journal and conference papers, and book chapters published (2008) – 2020	**Specific:** (i) AI-based process-mining approaches exploiting domain knowledge; (ii) process-mining approaches addressing auxiliary AI tasks jointly with target process-mining tasks
Herm et al. (2021)	**Specific:** Papers published (2015) – 2020	**Specific:** AI-based RPA
Wanner et al. (2020)	**General:** Papers published (2007) – 2018	**Specific:** Combining complex event processing and ML
Ko and Comuzzi (2023)	**General:** Journal and conference papers published (2000) – 2021	**Very specific:** Anomaly-detection task in pattern detection
Our work	**General:** Journal and conference papers published (1998) – 2022	**Broad:** ML applications for tasks in all BPM lifecycle phases

The year in round brackets indicates the year of the oldest paper in the final coding table.

agreement, Krippendorff's α (Krippendorff, 2018), and Fleiss' κ (Fleiss & Cohen, 1973), which are established metrics for three or more coders (Lombard, Snyder-Duch, & Bracken, 2006), and found high inter-coder reliability. Further information on the analysis and the results can be found in Appendix B. The coders pointed out any ambiguities in the coding table during the coding process, which could be resolved later in open discussions to ensure uniformity.

For greater transparency, our coding table can be found in Appendix C and in an interactive concept matrix.

4.3. Interactive concept matrix

One aim of our systematic literature review is to make the data coded from the collected papers transparent and accessible in a useful manner. While the coding table in Appendix C fulfils the first aspect, its usability is limited because of its size (number of papers coded and the number of coding categories). Therefore, we created an interactive concept matrix (including a filterable table and two exemplary visualisations of the results)[2] that synthesises the concepts the literature considers from a descriptive perspective. The interactive concept matrix is structured in three sheets, each allowing the user to filter certain concepts to gain insights. On the first sheet "Concept development over time", the user sees how the concepts the literature considers evolve, for example, per BPM lifecycle phase. Using the second sheet, "Intersection of concepts", the user can gain insights into how two concepts intersect. One example can be the intersection of the BPM lifecycle phases and the ML paradigms, where the heatmap colour indicates the number of papers per intersection category, as shown in Fig. 5.

From the third sheet, "Complete coding table", the user can glean a holistic perspective of all concepts the literature considers and filter the complete concept matrix for certain values, such as specific authors, publication years, or lifecycle phases.

[2] The interactive concept matrix can be found here: https://literaturedashboardbpm.herokuapp.com/.

5. Results

This section presents an overview of ML applications in BPM along the six phases of the BPM lifecycle and the BPM tasks derived from our literature review. Fig. 6 shows the BPM lifecycle by Dumas et al. (2018), enriched with tasks we identified. In addition, descriptions of BPM lifecycle phases and definitions of BPM tasks in the view ML applications are provided.

5.1. Process identification

In the process identification phase, organisations collect information about their business processes. Tasks in this phase are identifying processes in the organisational landscape, determining which business processes can be optimised and whether an optimisation should happen within or outside the organisation. ML applications can support different tasks in this phase, especially by unlocking unstructured data, repairing missing data, and supporting outsourcing decision-making. Moreover, creating data representations facilitates thorough analyses in later phases of the BPM lifecycle. Fig. 7 summarises the tasks in the process identification phase.

5.1.1. Business process outsourcing

A strategic task of process identification is to decide whether a business process or a part of a business process can be optimised regarding time or cost when outsourced to another organisation. ML applications can help identify process indicators that propose a high probability of successful outsourcing. Therefore, Ciasullo et al. (2018) propose a reinforcement-learning-based framework that supports outsourcing decisions of business processes.

5.1.2. Event log creation

After adding process mining to many organisations' business intelligence portfolios, event logs gained importance quickly. Event logs are the standard data form for process-mining applications. While structured data in information systems can directly be compiled into an

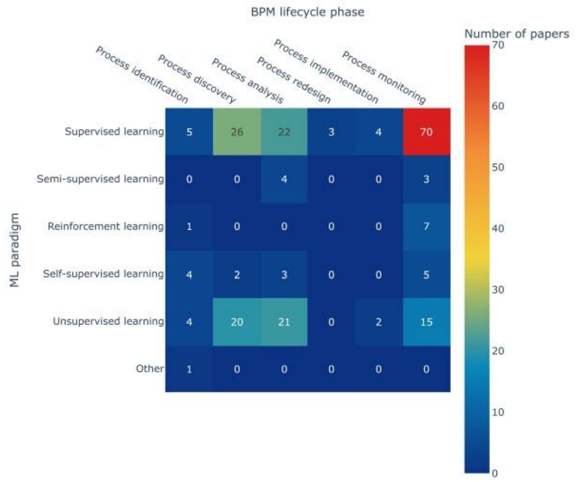


Fig. 5. Exemplary figure of the interactive concept matrix for the intersection of the BPM lifecycle phases and the ML paradigms.

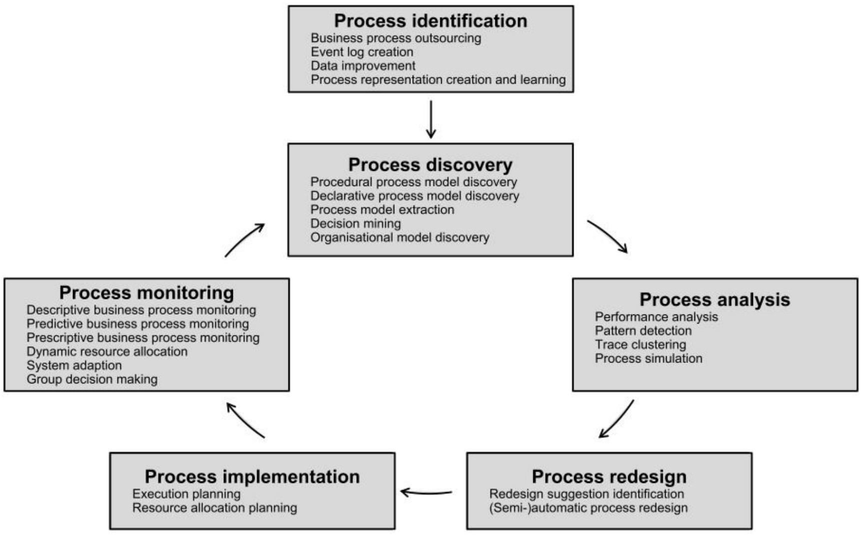


Fig. 6. BPM lifecycle of Dumas et al. (2018) enriched with the identified tasks.

event log, ML applications can also turn raw data (e.g. video, image, or plain text) into event logs. Therefore, Sim, Sutrisnowati, Won, Lee, and Bae (2022) propose a method to create an event log from raw event data using convolutional neural network (CNN) models. Tello, Gianini, Mizouni, and Damiani (2019) provide a framework that uses supervised-learning models for mapping low-level event data onto high-level process activities. An event log can also be created for further analysis by detecting process instances (i.e. sequences of events) in e-mails with a clustering model (Jlailaty, Grigori, & Belhajjame, 2017) or processing video data with a computer-vision-based approach (Kratsch, König, & Röglinger, 2022).

5.1.3. Data improvement

As part of process identification, process analysts identify process data in sources and sinks. Process data is often represented by event logs that hold a potential value but can lack quality when they include missing or incorrect data (Bose, Mans, & van der Aalst, 2013). ML applications can reconstruct missing values in event data or identify and replace incorrect event data. As ML applications for data improvement increase event-log quality, they leverage accurate analyses. Missing activities in event log data are repaired using a self-organising map model for trace clustering (Xu & Liu, 2019). Nolle, Seeliger, Thoma, and Mühlhäuser (2020) correct anomalies using gated recurrent unit

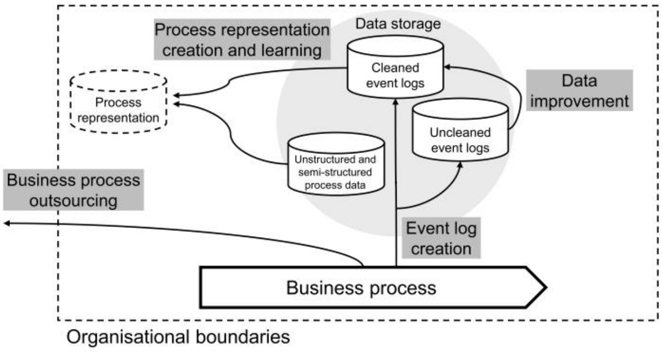


Fig. 7. Schematic representation of the tasks in the process identification phase.

(GRU) models with bi-directional beam search. Nguyen, Lee, Kim, Ko, and Comuzzi (2019) clean event log data using an autoencoder model.

5.1.4. Process representation creation and learning

Process identification also deals with finding compact, abstract, and computational process representations on which ML can build applications for addressing BPM tasks in subsequent phases of the BPM lifecycle. Therefore, BPM research has developed ML applications to create and learn process representations from process data.

Supervised-learning approaches can produce dense and accurate process representations (Seeliger, Luettgen, Nolle, & Mühlhäuser, 2021) and event abstractions (Tax, Sidorova, Haakma, & van der Aalst, 2016). De Koninck, vanden Broucke, and De Weerdt (2018) propose self-supervised learning techniques to create representations of activities, traces, event logs, and process models. Guzzo, Joaristi, Rullo, and Serra (2021) also use self-supervised learning to create process representations considering multi-dimensional aspects of event log traces.

5.2. Process discovery

Process discovery focuses on generating as-is models that describe a process as it is executed according to event data. Process models can become complex depending on the number of variants and the length of traces in a process. Unsupervised learning applications can focus on process variants to improve process model comprehension. While ML applications can detect general process model structures in data, they can also mine specific types of process models, such as procedural and declarative models. Moreover, ML applications can optimise discovered process models regarding specific process model quality criteria. Consequently, ML applications support many tasks in the process discovery phase and automate it to a certain extent. Fig. 8 overviews the process discovery tasks.

5.2.1. Discovery of procedural process model

Discovering procedural process models is a task to gain transparency over a process as it is executed. Discovered procedural process models can be complex, so they are often simplified using heuristics or filtering options. ML applications can identify process structures or optimise parameters and event-log sampling for discovery algorithms to create accurate and comprehensible procedural process models. Additionally, ML applications can choose the best-suited discovery algorithm based on the input data and produce specific procedural models (e.g. reference models). Early works in ML-based BPM, such as Herbst (2000), Herbst and Karagiannis (2000, 2004), focus on the discovery of process model structures using Bayesian algorithms.

Other studies address trace-clustering-based process model discovery. Greco, Guzzo, Pontieri, and Sacca (2004) propose an algorithm that incorporates a k-means model to cluster workflow executions, where a cluster is a global constraint of a discovered process. In a later work (Greco, Guzzo, Pontieri, & Sacca, 2006), the same authors present an extension of their algorithm for discovering conforming process models. García-Bañuelos, Dumas, La Rosa, De Weerdt, and Ekanayake (2014) propose a trace-clustering-based process discovery technique that allows users to control process model complexity. Closely connected to these studies, Qiao, Akkiraju, and Rembert (2011) present a clustering-based approach for retrieving business process models.

Process discovery applications that rely on genetic algorithms use target metrics like fitness (van der Aalst, De Medeiros, & Weijters, 2005), either fitness, replay, precision, generalisation, or simplicity (Buijs, van Dongen, & van der Aalst, 2012), or completeness, precision, and simplicity together (Vázquez-Barreiros, Mucientes, & Lama, 2014, 2015), to create optimised process models. Genetic algorithms can also be combined with graph-based representations to analyse complex processes (Turner, Tiwari, & Mehnen, 2008).

Other studies that consider discovery of procedural process model address various discovery tasks, such as the creation of process models for unseen event log data using graph neural network (GNN) models (Sommers, Menkovski, & Fahland, 2021), the creation of process models using hidden Markov models containing equations and rules (Sarno & Sungkono, 2016a), and, building on the latter, the inclusion of invisible prime tasks and parallel control-flow patterns (Sarno & Sungkono, 2016b).

Moreover, probabilistic process models using a Bayesian algorithm (Silva, Zhang, & Shanahan, 2005) are discovered. Lu, Zeng, and Duan (2016) discover synchronisation-core-based process models, where a decision-tree model is used to group traces into clusters, from which synchronisation cores of activity dependencies are derived. Ferreira and Gillblad (2009) discover process models from event stream data by estimating model parameters and case assignments using an expectation–maximisation procedure. Ferreira, Szimanski, and Ralha (2013) discover hierarchical process models that capture the relationships between macro-level activities and micro-level events using a hierarchical Markov model with expectation–maximisation-algorithm-fitted parameters.

Another group of studies discovers certain forms of process models, such as process trees using a k-means model to handle context data in addition to control-flow data (Shraga, Gal, Schumacher, Senderovich,

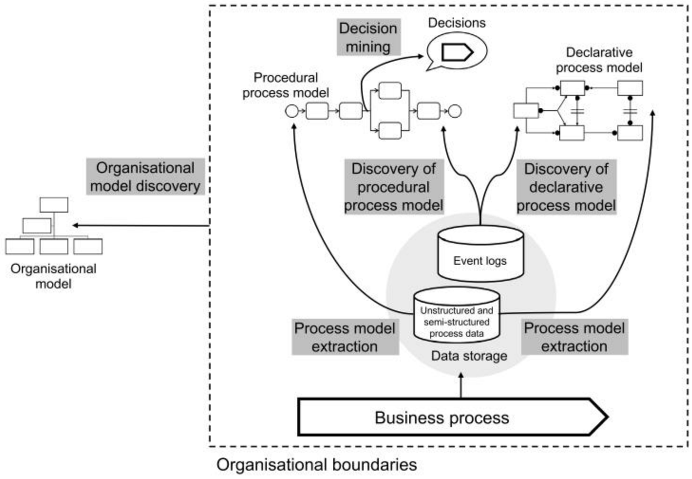


Fig. 8. Schematic representation of the tasks in the process discovery phase.

& Weidlich, 2020) or connections between events using a logistic regression model (Măruşter, Weijters, van der Aalst, & Van Den Bosch, 2002). Building on the latter, Măruşter, Weijters, van der Aalst, and Van Den Bosch (2006) propose a method that uses the RIPPER algorithm to create rule sets representing process-activity relations.

Finally, specific procedural process discovery applications are meta-process discovery for process discovery algorithm selection using an SVM model for reference model selection (Wang et al., 2012), reference model discovery using a genetic algorithm (Martens, Fettke, & Loos, 2014) or a clustering algorithm (Li, Reichert, & Wombacher, 2010), behavioural pattern model discovery using a clustering algorithm (Diamantini, Genga, & Potena, 2016), and discovery of configurable process model using a genetic algorithm (Buijs, van Dongen, & van der Aalst, 2013).

5.2.2. Discovery of declarative process model

In contrast to procedural process models, declarative models express specific rules, which should not be infringed during execution. The discovery of declarative process models finds constraints in an event log. ML applications support both discovering declarative models and consistency checking of such constraints. Leno, Dumas, and Maggi (2018) facilitate the discovery of multi-perspective declarative process models using a clustering model for selecting traces and a classification model to identify declarative constraints. Leno, Dumas, Maggi, La Rosa, and Polyvyanyy (2020) additionally present a redescription-mining-based approach for declarative constraint identification that uses two decision-tree models. Inductive logic programming allows for the discovery of declarative models (Chesani et al., 2009; Lamma, Mello, Montali, Riguzzi & Fabrizio and Storari, 2007) and learning integrity constraints for declarative modelling (Lamma, Mello, Riguzzi, Storari, Sergio, 2007). Maggi, Bose, and van der Aalst (2012) use the Apriori algorithm to find declarative constraints, and Maggi, Di Ciccio, Di Francescomarino, and Kala (2018) combine the Apriori algorithm with a sequence analysis algorithm to discover declarative constraints.

5.2.3. Process model extraction

Process models can be modelled in diagram software or paper drawings, or described in plain text. Interpreting these process models and translating them into computationally interpretable process models can be done with ML applications. Plain-text process descriptions can serve as foundations for extracting procedural process models using a hierarchical multi-grained deep neural network (DNN) (Qian et al., 2020) and declarative process models using a discovery algorithm that incorporates a BERT model (López, Strømsted, Niyodusenga, & Marquard, 2021). Polančič, Jagečič, and Kous (2020) extract digital process models from hand-drawn process diagrams using a CNN model. Kim, Suh, and Lee (2002) propose a document-based discovery approach that applies case-based reasoning to effectively reuse the design outputs.

5.2.4. Decision mining

When discovering processes, it is essential to understand decision points where process flows split. Decision mining comprises identifying decision points and conditions, which lead to a course of action. ML applications can automate identifying decision points and decision dependencies. Process models can be enhanced by using decision-tree models to detect dependencies in the data at the decision points in the process model, as Rozinat and van der Aalst (2006) do in one of the first approaches. Bazhenova, Buelow, and Weske (2016) use decision trees to semi-automatically derive decision models from event logs. Both Mannhardt, de Leoni, Reijers, and van der Aalst (2016) and Effendi and Sarno (2017) present a technique for discovering overlapping rules in event logs using decision-tree models. de Leoni, Dumas, and García-Bañuelos (2013) propose a technique for discovering branching conditions by combining decision-tree models with invariant discovery techniques. Lastly, Sarno, Sari, Ginardi, Sunaryono, and Mukhlash (2013) propose a decision-mining approach that uses a decision-tree model and considers multi-choice workflow patterns.

5.2.5. Organisational model discovery

Processes are not executed in isolation but in the context of an organisation. Consequently, event data contains information about organisational functions. ML applications can create models of the executing organisation based on event logs using clustering models (Yang, Ouyang, Pan, Yu, & ter Hofstede, 2018).

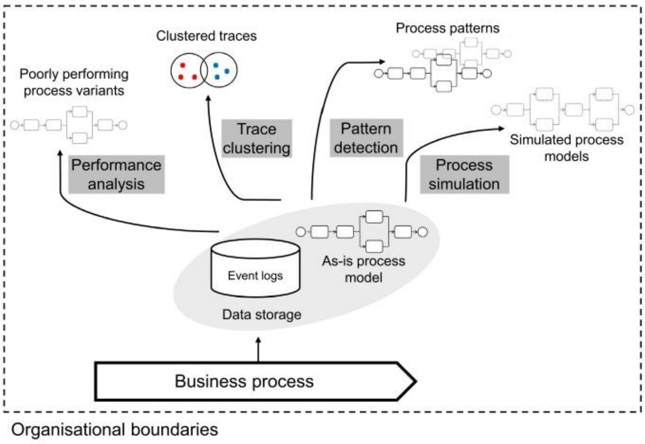


Fig. 9. Schematic representation of the tasks in the process analysis phase.

5.3. Process analysis

Process analysis assesses as-is processes using process models and process data. Process analysis takes a retrospective perspective in analysing a process. Process analysts determine key performance indicators (KPIs), assess process performance, look for problematic instances, and simulate processes to evaluate process redesigns. All these tasks can be supported with ML applications' pattern recognition capabilities, unsupervised learning capabilities for identifying similar process instances regarding control flow or performance, and predictive capabilities to determine realistic simulation parameters. Fig. 9 shows a schematic representation of the tasks in the process analysis phase.

5.3.1. Performance analysis

When assessing the performance of processes and process instances, organisations determine KPIs to look for poorly performing processes and the cause of their performance failure. In doing this, analysts find starting points for process redesign. ML applications support performance analysis, for example, by repairing scarce performance data, aggregating KPIs, and identifying causal relations between data and process outcome. Berkenstadt, Gal, Senderovich, Shraga, and Weidlich (2020) propose an approach using techniques from queuing theory and supervised learning to predict process performance indicators. Wetzstein et al. (2009) use rule-based learning to recompose low-level process indicators into high-level process metrics. To enable fairness-aware process mining and performance analysis, Qafari and van der Aalst (2019) add situation-specific discrimination to event logs and train decision-tree models. Es-Soufi, Yahia, and Roucoules (2016) propose an approach that detects process patterns using process mining and creates performance-related predictions per activity of a process pattern using a supervised-learning model.

Training logical decision-tree models based on logical event log encoding (Ferreira & Vasilyev, 2015) and uplift tree models (Bozorgi, Teinemaa, Dumas, La Rosa, & Polyvyanyy, 2020) can reveal causal relationships in process performance and specific control-flow actions. Savickas and Vasilecas (2018) use belief network models to analyse process performance with domain-specific data. In contrast to the previously mentioned approaches, Theis and Darabi (2020) compute the generalisability of process models using generative adversarial network (GAN) models. To support performance analyses in general, de Leoni, van der Aalst, and Dees (2014), de Leoni et al. (2016) present a process analysis framework to correlate different process characteristics using regression and decision-tree models.

5.3.2. Pattern detection

Process analysis further strives to identify favourable and problematic process patterns. Before pattern recognition capabilities of ML applications could be leveraged, process analysts manually compared process variants to find workarounds, anomalous behaviour or time series patterns, which could be improved or transferred to other instances. Process analysts can thereby provide input for redesigning process variants. Analysts can use ML applications with autoencoder models to identify anomalies even when they lack domain knowledge, as Nolle, Luettgen, Seeliger, and Mühlhäuser (2018) do without, and Krajsic and Franczyk (2020) do within, a lambda architecture designed for event steams. Another approach uses a word2vec model to encode activities in event log data and a (one-class) SVM model to detect anomalies (Junior, Ceravolo, Damiani, Omori, & Tavares, 2020). Rogge-Solti and Kasneci (2014) detect temporal anomalies in the activity execution duration using Bayesian models.

Cuzzocrea, Folino, Guarascio, and Pontieri (2015, 2016a) detect deviant process instances using multi-view, multi-dimensional ensemble learning in event logs. Building on their techniques, Cuzzocrea, Folino, Guarascio, and Pontieri (2016b) present a probabilistic-based framework for robust detection of deviance. Stierle, Weinzierl, Harl, and Matzner (2021) derive activity relevance-scores from an attention-layer of a GNN model that detects deviant process instances.

Weinzierl, Wolf, Pauli, Beverungen, and Matzner (2022) present a method for detecting workarounds using an autoencoder model that removes noisy process instances and a CNN model that maps process instances to workaround classes. Yeshchenko, Di Ciccio, Mendling, and Polyvyanyy (2019) detect concept drifts in event log data using visual analytics and time-series-based clustering. Finally, supervised learning and conformance checking (Valdés, Céspedes-González, & Pou, 2022a) or supervised learning and graph kernels (Valdés, Céspedes-González, & Pou, 2022b) can prove useful in identifying time-series patterns in event log data.

5.3.3. Trace clustering

Clustering approaches can automatically group process instances or instance variants to analyse processes quickly. Improvements, problems, and advantages identified in one variant may apply to other instances or instance variants in the same cluster. Therefore, clustering traces can support process analysts.

The most popular trace-clustering approach in BPM is distance-based trace clustering. Distance-graph models (Ha, Bui, & Nguyen, 2016), graph similarity metrics (De Weerdt, vanden Broucke, Vanthienen, & Baesens, 2012), and similarity in heterogeneous information networks (Nguyen, Slominski, Muthusamy, Ishakian, & Nahrstedt, 2016) allow for distance-based trace clustering. Delias, Doumpos, Grigoroudis, and Matsatsinis (2019) take the outranking relations theory into account for trace clustering. Other studies address trace clustering using a co-training-based strategy with multiple trace profiles (Appice & Malerba, 2015), frequent-item-set mining (Seeliger, Nolle, & Mühlhäuser, 2018), and a Levenshtein-distance-based approach (Bose & van der Aalst, 2009) to consider the process context. High-level process analysis is achieved by employing hierarchical clustering (Jung, Bae, & Liu, 2008, 2009) or clustering using graph partitioning (Sarno, Ginardi, Pamungkas, & Sunaryono, 2013) based on graph similarity of different business process models.

Previous research also uses model-based trace clustering and approaches that include expert and domain knowledge via process model metrics (De Koninck, Nelissen, Baesens, Snoeck & Monique and De Weerdt, 2017), the extension with must-link and cannot-link relationships between process instances (De Koninck, Nelissen, Baesens, Snoeck, & De Weerdt, 2021), and active learning (De Weerdt, vanden Broucke, Vanthienen, & Baesens, 2013). Super-instance-based trace clustering enables combining distance- and model-based clustering (De Koninck & De Weerdt, 2019). Boltenhagen, Chatain, and Carmona (2019) use distance-based quality criteria in model-based trace clustering.

Prior research also describes certain applications of trace clustering. Folino, Guarascio, and Pontieri (2015) use a logical trace-clustering model to create comprehensible process models. Varela-Vaca, Galindo, Ramos-Gutiérrez, Gómez-López, and Benavides (2019) apply trace clustering to mine configuration flows optimised for a particular type of user. Yang et al. (2017) use distance-based clustering to set up a recommender system. Ferreira, Zacarias, Malheiros, and Ferreira (2007) apply a mixture of Markov chains learned with the expectation-maximisation algorithm to create traces from identified tasks. De Koninck, De Weerdt, and vanden Broucke (2017) present a trace clustering approach with instance-level explanations.

5.3.4. Process simulation

Process analysis assesses potential process changes by simulating redesigned process models. ML applications can aid in creating realistic simulation conditions, as the conditions are learned from a mass of data instead of hard-coded rules. While Camargo, Dumas, and González-Rojas (2022) facilitate data-driven process simulation using two long short-term memory (LSTM) models, one for processing time prediction and one for waiting time prediction, Khodyrev and Popova (2014) use decision and regression tree models to predict KPI values for short-term process simulation.

5.4. Process redesign

Process redesign's goal is to improve an as-is process by applying process changes to the model and creating a to-be process model. Process designers use, for example, experience-based redesign heuristics, describing concrete measures to make a process efficient or less prone to error. ML applications can support process redesigners by automatically suggesting process redesign measures based on data or by performing (semi-)automatic process redesign, where more appropriate or optimised process designs are identified. Fig. 10 provides an overview of these tasks.

5.4.1. Redesign suggestion identification

A task in process redesign is to elaborate redesign suggestions based on ideas, heuristics, or new technology. ML applications can identify unfavourable patterns in process instances and how to redesign them, for example, geared towards a certain target (e.g. customer satisfaction or throughput time). Therefore, research in this area addresses redesign suggestion identification. Mustansir, Shahzad, and Malik (2022) use language models and DNN models to detect suggestions for process redesign in textual customer feedback data.

5.4.2. (Semi-)automatic process redesign

A subset of ML applications in process redesign focuses on semi-automatic process redesign. More concretely, ML applications addressing this task use genetic algorithms to perform multi-objective optimisation of business processes (Vergidis, Tiwari, Majeed, & Roy, 2007) or to determine an objective prioritisation of process redesigns (Afflerbach, Hohendorf, & Manderscheid, 2017).

5.5. Process implementation

Redesigned process models must be integrated into an information system before they can be executed. Resources need to be assigned to redesigned tasks, staff needs to be trained, and execution plans need to be rolled out. ML-supported process implementation tasks include creating execution and staffing plans from process models and data and resource allocation planning. Fig. 11 shows the tasks of the process implementation phase.

5.5.1. Execution planning

Transferring from a to-be process model to process implementation requires setting up execution plans. Process models, including decision points and process semantics, can be used to create execution plans. This is tedious work, but ML applications can approximate an optimal execution plan from a detailed model. Therefore, Bae, Lee, and Moon (2014) propose a method to enrich BPM structures with alternative process paths and semantics, after which a genetic-based algorithm can find an execution plan that enhances performance.

5.5.2. Resource allocation planning

After having set execution plans, resources, such as machines, employees, storage spaces or production lines, must be assigned to cases. Process owners define who or what is employed in which process variant. To make decisions during planning, ML applications can provide suggestions that optimise resource utilisation or throughput time. Process owners can then pick a suggested resource allocation plan and adjust it if needed. ML applications thereby eliminate trial and error approaches. Resource allocation planning is realised based on the prediction of execution routes using a naïve Bayes model (Kazakov, Novikov, Kulagina, & Shlapakova, 2018), the resource decision mining from events using classification models and heuristics (Senderovich, Weidlich, Gal, & Mandelbaum, 2014), the examination of process instance scheduling using a genetic-based algorithm (Xu, Liu, Zhao, Yongchareon, & Ding, 2016), or the recommendation of resource allocation candidates using classification models (Liu, Cheng, & Ni, 2012). Finally, Delcoucq, Dupiereux-Fettweis, Lecron, and Fortemps (2022) apply a bi-dimensional clustering approach for resource allocation planning, where resources are clustered based on similar behaviour in the process, and activities are clustered based on the executing resource.

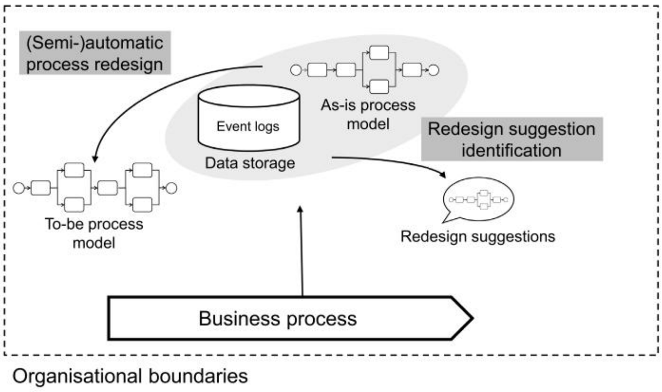


Fig. 10. Schematic representation of the tasks in the process redesign phase.

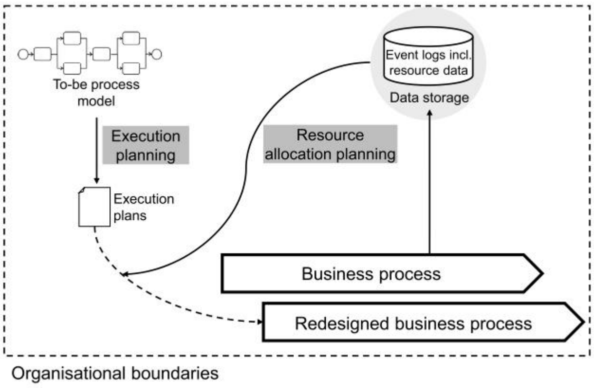


Fig. 11. Schematic representation of the tasks in the process implementation phase.

5.6. Process monitoring

After the redesigned processes are implemented, they are executed, and how well they perform in terms of their performance and objectives must be determined. Process monitoring challenges analysts because parallel cases must be overseen. Additionally, cases, which look fine momentarily, can turn out problematic later on. ML applications can recognise patterns in event data that indicate problematic outcomes, even when a case is in a good state. ML applications can help analysts identify cases for further analysis in large process data sets or provide decision support to process users, enabling them to execute processes with higher performance. Executing business processes can also be optimised regarding KPIs or process conformance. ML applications in process monitoring can further automate noticing an appropriate point for starting another BPM lifecycle iteration by detecting concept drifts in processes. Fig. 12 provides a schematic overview of the tasks and their relationship.

5.6.1. Descriptive business process monitoring

Descriptive business process monitoring determines KPIs of running process instances and provides real-time information about reactive actions that could increase process performance. It can also detect concept drifts in processes and identify anomalous behaviour in process instances. ML applications can recognise patterns in high-volume data. In doing that, they help to find entry points for reiterating the BPM lifecycle to adapt to concept drifts or positive anomalies, or to counteract poor process behaviour.

Montani and Leonardi (2012) retrieve traces via a k-nearest-neighbour model and cluster them using a hierarchical clustering model to detect changes in event log data.

Online detection of anomalies is addressed using graph autoencoder models (Huo et al., 2021) or supervised learning, specifically, a random forest model, extreme gradient-boosting model, or an LSTM model (Lee, Lu, & Reijers, 2022). Other studies apply a DBSCAN model to cluster process instances into common and anomalous behaviour (Junior, Tavares, da Costa, Ceravolo, & Damiani, 2018), or a DenStream model (Cao, Estert, Qian, & Zhou, 2006) to cluster process instances with a graph-based representation including control flow and time (Tavares, Ceravolo, da Costa, Damiani, & Junior, 2019) and additional attributes (Ceravolo, Damiani, Schepis, & Tavares, 2022). Krajsic and Franczyk (2021) create a latent space representation of event data with a variational autoencoder, cluster the representation with a k-means model, and use either an isolation forest, a local outlier factor, or a one-class SVM model for each cluster to detect anomalies. For

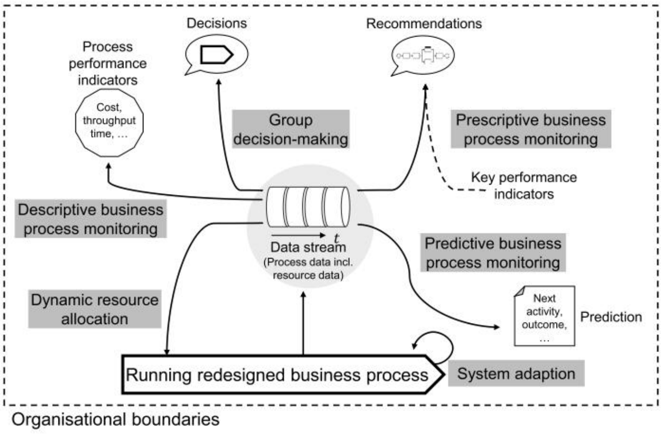


Fig. 12. Schematic representation of the tasks in the process monitoring phase.

their part, Nolle, Seeliger, and Mühlhäuser (2018) and Nolle, Luettgen, Seeliger, and Mühlhäuser (2019) determine anomaly scores based on the output of a GRU model and map them onto types of anomalies, while Lee, Burattin, Munoz-Gama, and Sepulveda (2021) check process conformance online using a hidden Markov model.

5.6.2. Predictive business process monitoring

In contrast to descriptive monitoring, predictive business process monitoring shifts to an ex-ante perspective by predicting different process properties of running process instances. With these predictions, process users can take proactive or corrective actions to mitigate risks or improve the performance of process executions. Predictive business process monitoring pursues multiple objectives including future activities, outcomes, remaining time, or a combination of objectives. Those approaches facilitate early warning systems.

A common prediction task is the next activity prediction, which deep-learning-based techniques address. For the next activity prediction, studies employ LSTM models (Evermann, Rehse, & Fettke, 2017; Pasquadibisceglie, Appice, Castellano, & Malerba, 2021), CNN models (Di Mauro, Appice, & Basile, 2019; Heinrich, Zschech, Janiesch, & Bonin, 2021; Pasquadibisceglie, Appice, Castellano, & Malerba, 2019, 2020), multi-layer perceptron (MLP) models (Mehdiyev, Evermann, & Fettke, 2020; Theis & Darabi, 2019), and transformer models (Heinrich et al., 2021; Philipp, Jacob, Robert, & Beyerer, 2020). Some works use deep-learning models and provide explainable next-activity predictions. Wickramanayake et al. (2022) (and their earlier version (Sindhgatta, Moreira, Ouyang, & Barros, 2020)) propose LSTM models that differ in the attention mechanism they use for the model explanation, Hsieh, Moreira, and Ouyang (2021) use an NN architecture computing dynamic features through an LSTM and static features through an MLP and generate explanations for predictions via the model-agnostic counterfactual algorithm DiCE (Mothilal, Sharma, & Tan, 2020), Hanga, Kovalchuk, and Gaber (2020) use an LSTM model and infer from it business process models to explain the model's decision-making process, and Gerlach, Seeliger, Nolle, and Mühlhäuser (2022) use a GRU model and infer from it multi-perspective likelihood graphs for the same purpose. A smaller group of studies proposes Bayesian approaches, that is, regularised probabilistic finite automata models (Breuker, Matzner, Delfmann, & Becker, 2016) or Bayesian network models (Brunk et al., 2021; Pauwels & Calders, 2020). Other studies compute transition probabilities in process models using decision-tree models (Lakshmanan, Shamsi, Doganata, Unuvar, & Khalaf, 2015) or propose factorisation machines to predict next activities (Lee, Parra, Munoz-Gama, & Sepulveda, 2018). Pauwels and Calders (2021) introduce incremental learning strategies for updating next-activity-prediction models. Finally, instead of only predicting the next activity, Bernard and Andritsos (2019) address the remaining activity sequence prediction (or suffix prediction) using clustering and SVM models to provide transparent predictions.

Another prediction task is outcome prediction, whose outcome can be binary (e.g. the violation of a business rule), multi-class (e.g. multiple conditions), or continuous (e.g. a KPI value). To predict binary process outcomes in the form of conditions, studies use a CNN model (Pasquadibisceglie, Appice, Castellano, Malerba, & Modugno, 2020), an LSTM model (Wang, Yu, Liu, & Sun, 2019) or a random forest or logistic regression model including features created with text models (Teinemaa, Dumas, Maggi, & Di Francescomarino, 2016). Other studies with the same aim use a gated GNN model, from which attention-based explanations are extracted (Harl, Weinzierl, Stierle, & Matzner, 2020), a model built with an evolutionary algorithm and decision rules generated to explain the model's predictions (Márquez-Chamorro, Resinas, Ruiz-Cortés, & Toro, 2017), or a neuro-fuzzy model, from which explainable rules are directly extractable (Pasquadibisceglie, Castellano, Appice, & Malerba, 2021). Rizzi, Di Francescomarino, and Maggi (2020) predict the violation of business rules with a random forest model and use post-hoc explanations created with LIME (Ribeiro, Singh, & Guestrin, 2016) to manipulate data to improve model accuracy. Di Francescomarino, Dumas, Maggi, and Teinemaa (2016) use a DBSCAN model as a density or a model-based clustering approach to cluster partial process instances. Then they apply a decision tree or a random forest model per cluster to predict whether a running process instance violates a business rule. Folino, Guarascio, and Pontieri (2012) and Bevacqua, Carnuccio, Folino, Guarascio, and Pontieri (2013) predict violations of service-level agreements. Folino et al. (2012) first cluster running process instances. Then, with a cluster-dependent finite state machine model, the remaining time is predicted, and a rule is applied to the predictions to determine the violations. Bevacqua et al. (2013) follow a similar approach but apply a rule to remaining-time or remaining-step predictions generated with regression models to determine the violations. Cuzzocrea, Folino, Guarascio, and Pontieri (2019) predict violations of aggregated performance constraints. First, running process instances of a window are clustered into variants. Afterwards, with a cluster-dependent linear regression or a k-nearest neighbour model, a performance indicator is predicted for each running process instance. Then, a Gaussian-process regression model predicts an aggregated performance indicator over future window slots, and a rule is applied to the predictions to determine the violations. As multi-class process outcomes, Folino, Greco, Guzzo, and Pontieri (2011) predict behavioural classes of running process instances using a decision-tree model. Lastly, continuous process outcomes are predicted as customer delays with non-linear regression or regression tree models (Senderovich, Weidlich, Gal, & Mandelbaum, 2015), as delivery times with various regression models and concept-drift handling (Baier, Reimold, & Kühl, 2020), and as process performance at the process-model level using a CNN, LSTM, or long-term recurrent CNN model (Park & Song, 2020).

Prediction of the remaining time is another common prediction task. The remaining time of running process instances is predicted using shallow ML models considering inter-case features (Klijn & Fahland, 2020; Senderovich, Di Francescomarino, Ghidini, Jorbina, & Maggi, 2017), shallow ML models considering news sentiments as external process context (Yeshchenko, Durier, Revoredo, Mendling, & Santoro, 2018), data-aware transition systems annotated with shallow ML models without (Polato, Sperduti, Burattin, & de Leoni, 2014) and with concept-drift adaption (Firouzian, Zahedi, & Hassanpour, 2019b), Bayesian NN models (Weytjens & De Weerdt, 2021), LSTM models (Alves, Barbieri, Stroeh, Peres, & Madeira, 2022), and MLP models with entity-embedding (Wahid, Adi, Bae, & Choi, 2019).

Some studies do not focus on a specific prediction task but address multiple prediction tasks. Appice, Di Mauro, and Malerba (2019) predict next activities, next timestamps, and remaining time using various feature encodings, learning approaches, and shallow ML models. Extending their earlier work (Polato et al., 2014), Polato, Sperduti, Burattin, and de Leoni (2018) present three methods that rely on naïve Bayes and support vector regression models to predict the remaining time, one of which can also be used to predict suffixes. In a pioneering work using deep learning in the domain, Tax, Verenich, La Rosa, and Dumas (2017) predict next activities, next timestamps, remaining times, and suffixes using a multi-task LSTM model. This approach (Tax et al., 2017) has been be extended in various ways, such as with embedding layers for efficient computation (Camargo, Dumas, & González-Rojas, 2019), linear temporal logic rules in the post-processing for accurate suffix predictions (Di Francescomarino, Ghidini, Maggi, Petrucci, & Yeshchenko, 2017), or text models in the pre-processing to compute textual data (Pegoraro, Uysal, Georgi, & van der Aalst, 2021). Galanti, Coma-Puig, de Leoni, Carmona, and Navarin (2020) use an LSTM model to generate remaining time, activity occurrence, and cost predictions, and apply SHAP (Lundberg & Lee, 2017) to create post-hoc explanations for the predictions. Other deep-learning-based approaches predict process failure and next activities using a DNN model with convolutional and recurrent layers (Borkowski, Fdhila, Nardelli, Rinderle-Ma, & Schulte, 2019), or next activities and suffixes using an LSTM-based encoder–decoder model with an attention layer (Jalayer, Kahani, Beheshti, Pourmasoumi, & Motahari-Nezhad, 2020). Pfeiffer, Lahann, and Fettke (2021) encode different event information as gramian angular fields (i.e. 2D-images), learn a representation based on this encoding using a CNN model, and fine-tune the trained model towards various prediction tasks. Taymouri, La Rosa, Erfani, Bozorgi, and Verenich (2020) use a GAN model to predict next activities and next timestamps, and Taymouri, La Rosa, and Erfani (2021) combine these with beam search to predict suffixes and remaining time.

Other studies propose solutions for certain prediction tasks — predicting activity ordering (Verenich, Dumas, La Rosa, Maggi, & Di Francescomarino, 2016), enterprise social networks (Pham, Ahn, Kim, & Kim, 2021), or process outcomes based on scarcely labelled event logs (Folino, Folino, Guarascio, & Pontieri, 2022).

5.6.3. Prescriptive business process monitoring

Prescriptive business process monitoring aims at steering processes towards specific optimisation objectives. To achieve this, such ML applications recommend taking action in process instances. Other approaches warn process owners when a process instance requires human attention; for instance, if a negative outcome is expected. ML applications point analysts' attention to important process instances and provide instructions to process users on how they can carry out their work effectively and efficiently.

One stream of research investigates warnings that are based on process predictions. One indicator of a triggered warning can be the prediction of a process outcome, as Teinemaa, Tax, de Leoni, Dumas, and Maggi (2018) propose using a random forest model or a gradient-boosted tree model in combination with a cost model. Another indicator can be the prediction of a process performance indicator, as suggested by Kang, Kim, and Kang (2012) using an SVM model. Bozorgi, Teinemaa, Dumas, La Rosa, and Polyvyanyy (2021) use an orthogonal random forest model to recommend whether and when to apply an intervention in a process to decrease running process instances' cycle times.

Another stream of research investigates recommendations for actions that need to be taken in response to process predictions. The recommendation of the next best actions is addressed by predicting process risk using a decision-tree model and applying integer linear programming (Conforti, de Leoni, La Rosa, van der Aalst, & ter Hofstede, 2015). Next-best-action recommendations are also determined based on next activity predictions generated by an LSTM model, KPI information, and declarative process model simulation (Weinzierl, Dunzer, Zilker, & Matzner, 2020). Alternatively, they are based on KPI value predictions obtained using a random forest, SVM, or decision-tree model and a transition-system abstraction (de Leoni, Dees, & Reulink, 2020). Khan, Ghose, and Dam (2021) model the next-best-action recommendation as a Markov decision process and apply deep Q-learning to learn the optimal policy for solving this decision problem.

5.6.4. Dynamic resource allocation

While resource allocation planning is an important task in process implementation, dynamic resource allocation is part of daily operations. Because of changes in order situations or resource availability beyond expectation, organisations need to reallocate their resources to activities during process execution. This task can be formulated as an optimisation problem. Therefore, ML applications optimise resource allocation taking run-time constraints into account, and assist process owners in finding optimal resource utilisation. Some papers apply shallow ML models to map activities to appropriate workers (Liu, Wang, Yang, & Sun, 2008), use naïve Bayes models to predict the performance of human resources before incoming jobs are assigned to them based on the predictions (Wibisono, Nisafani, Bae, & Park, 2015), or apply a k-means model to group process instances at critical activities before these are mapped to available resources (Pflug & Rinderle-Ma, 2016). Other works use a DBSCAN model and an ensemble of MLP models to allocate human resources based on team faultiness (Zhao, Pu, & Jiang, 2020), or an agglomerative hierarchical clustering model and a k-means model to recommend resources' task preferences (Zhao, Liu, Dai, & Ma, 2016). Studies applying reinforcement learning, specifically the Q-learning algorithm, allocate resources via policies that consider time and costs (Huang, van der Aalst, Lu, & Duan, 2011) or time and workload balancing (Firouzian, Zahedi, & Hassanpour, 2019a) as optimisation objectives. Park and Song (2019) use an LSTM model to predict the next activities and processing time and use the predictions to build a bipartite graph, laying the basis for solving a minimum-cost and maximum-flow problem for resource allocation.

5.6.5. System adaption

Process and system adaptations during execution may be required to adjust to changing process environments. Such adjustments do not necessarily need to trigger an iteration of the full BPM lifecycle but can be implemented instantaneously. Self-adaptive systems approaches find adaptations, which improve process performance in the monitoring phase. In contrast to prescriptive business process monitoring, self-adaptive systems apply permanent changes to a process instead of improving a process instance. Studies address self-adaptive systems that increase process reliability using an ensemble of LSTM models (Metzger, Neubauer, Bohn, & Pohl, 2019) and reinforcement learning (Metzger, Kley, & Palm, 2020). Saraeian, Shirazi, and Motameni (2019) use an MLP model to estimate uncertain characteristics to control and optimise an autonomous BPM system architecture. Huang, van der Aalst, Lu, and Duan (2010) adapt work distributions based on reinforcement learning, consider process performance goals as optimisation objectives, and learn work distribution policies during process condition changes. Samiri, Najib, El Fazziki, and Boukachour (2017) combine reinforcement learning and forecasting techniques to adapt workflows automatically.

5.6.6. Group decision-making

Group decision-making is only addressed by De Maio, Fenza, Loia, Orciuoli, and Herrera-Viedma (2016), who use reinforcement learning to learn the weighting of decision-makers (e.g. heterogeneous experts) for a decision activity (e.g. select a supplier) based on past process executions considering context and business process performances.

6. Synthesis of results and derived findings

In this section, we synthesise the results presented in the previous section for each BPM lifecycle phase. Based on the synthesised results along with the interactive coding table, we derived findings, which we present in Section 6.7.

6.1. Process identification

As data input, most ML applications in process identification use event logs, including time or resource information. Unlike other BPM lifecycle phases, input data types include unstructured raw event, video, and text data for event log creation.

Moreover, ML applications that address process representation creation and learning employ language models trained with NNs such as word2vec (e.g. De Koninck et al., 2018) or use embedding layers in DNN models (e.g. Guzzo et al., 2021) for feature extraction.

In process identification, ML applications follow the supervised- and unsupervised-learning approach. Additionally, ML applications deal with self-supervised learning because NNs are employed to create and learn process representations. These NNs are designed to extract a label from underlying data automatically (e.g. Nguyen et al., 2019). While around half of the ML applications in this phase adopt deep learning, almost all use NN models. This is because of the flexibility of designing NNs used in process representation creation and learning, event log creation, and data improvement applications.

To assess the performance of models in the tasks process representation creation and learning and event log creation, clustering models are trained with the created process data instances, and clustering metrics are calculated for the clustered instances (e.g. rand index or mutual information score). For this, mainly publicly available event logs are used such as the ones from the BPI challenges.[3]

[3] For example, see Business Process Intelligence Challenge: https://www.tf-pm.org/resources/logs.

6.2. Process discovery

In process discovery, most ML applications use event logs that only include control-flow information as data input to discover process models. Additionally, unstructured data, such as text from e-mails or image data, are used because ML applications can extract process models from unstructured data.

ML applications addressing process model extraction use language models trained with NNs, such as a BERT model (López et al., 2021) for feature extraction from unstructured data. For other tasks of this phase, like discovery of procedural process models, feature extraction refers to transforming the control flow of event logs into instance types, such as activity sequences or activity graphs.

Concerning the ML paradigm, ML applications of this phase follow the supervised-learning and unsupervised-learning approaches at about the same rate. Because of the lack of label information, unsupervised learning, however, is more common. For example, in discovery of procedural process model, process activity relations are mined in an unsupervised manner using the RIPPER algorithm (Măruşter et al., 2006). Regarding ML concepts, deep learning and transfer learning are adopted in process model extraction applications. While deep learning facilitates learning process model patterns from unstructured input data, transfer learning supports learning such patterns through pre-trained models. Moreover, clustering, Bayesian, decision-tree, genetic, and rule-based algorithms are used in process discovery. Clustering algorithms facilitate generating comprehensible process models for further analysis, as, for example, Greco et al. (2006) do. Bayesian-based algorithms can induce process structures from data (e.g. Herbst, 2000). Decision-tree-based algorithms are applied to create decision-tree models in decision-mining applications. Genetic-based algorithms are used to find optimal process models according to one or more selected process model quality criteria, as, for example, Vázquez-Barreiros et al. (2015) do. Rule-based algorithms are applied to learn rules from event logs; such rules represent a declarative process model in process discovery (e.g. Maggi et al., 2012).

The model assessment uses process discovery metrics from process mining, such as fitness, precision or generalisation, indicating how good the ML-created process models are. Often, own event logs are used for model assessment, but also publicly available logs.

6.3. Process analysis

Most ML applications use event logs, including time, resource, or data-related context information. Text data or process models are only occasionally used as data input. Some approaches also integrate domain knowledge.

ML applications in process analysis propose individual feature engineering techniques for event log data, and commonly used techniques cannot be observed, even when considering single tasks.

Like in process discovery, the supervised-learning and unsupervised-learning approach is addressed at about the same rate. This ratio can be explained by trace clustering applications, which are prominent in process analysis and address unsupervised-learning problems (e.g. Bose & van der Aalst, 2009). Analysis applications adopt deep, ensemble, active, meta, and multi-view learning. Deep learning is used to learn accurate models for process analysis tasks by considering intricate structures in event log data, as Camargo et al. (2022) do for process simulation. Ensemble learning is adopted to detect patterns based on the outcome of models, which are trained with different supervised ML algorithms (e.g. Cuzzocrea et al., 2016a). Active learning is used to increase ML models' accuracy by selecting process instances from an event log based on a metric, as in De Weerdt et al. (2013) for trace clustering, or human judgement, as in Cuzzocrea et al. (2016b) for pattern detection. A meta-model is learned based on the outcome of several base models to detect patterns in event log data (e.g. Cuzzocrea et al., 2016a). Multi-view learning is adopted to train multiple shallow

ML models based on event logs, where each model refers to a view for accurate pattern detection (e.g. Cuzzocrea et al., 2015) or complexity-reduced trace clustering (Appice & Malerba, 2015). ML applications in process analysis use unsupervised clustering algorithms for trace clustering. Further, as in process discovery, we observe that rule-based algorithms are employed here to learn models as a set of explainable rules from event log data; these rules reveal the cause of delayed processes (e.g. Ferreira & Vasilyev, 2015).

Process discovery metrics (e.g. fitness, precision, or generalisation) or clustering metrics (e.g. cluster set entropy or mutual information score) are calculated to assess the performance of clustering models employed in trace clustering applications. For applications addressing other tasks in process analysis, common supervised ML metrics are calculated for model assessment. Mainly publicly available event logs, but sometimes also own event logs are used.

6.4. Process redesign

ML applications for (semi-)automatic process redesign receive business process designs consisting of activities, connections, and routing decisions (e.g. Afflerbach et al., 2017) as data input, while ML applications for redesign suggestion identification can compute event log data (Mustansir et al., 2022). In ML applications for process redesign, feature engineering is not explicitly described. ML applications for process redesign follow the supervised-learning approach as both tasks in this phase, redesign suggestion identification (e.g. Mustansir et al., 2022) and (semi-)automatic process redesign (e.g. Afflerbach et al., 2017), require guidance for model training. Concerning ML algorithms, NNs are used for detecting redesign suggestions, while (semi-)automatic process redesign is addressed using genetic algorithms. Common supervised ML metrics and KPIs (e.g. customer satisfaction) are used with publicly available event logs for model assessment.

6.5. Process implementation

ML applications for resource allocation receive event logs with resource information as data input (e.g. Liu et al., 2012). Other dataset types can also be found in this phase, such as business process problems (e.g. Bae et al., 2014). Regarding feature engineering, ML applications for process implementation only selectively describe the extraction of features. However, common techniques for feature extraction cannot be observed. ML applications in this phase follow the supervised- and unsupervised-learning approach. However, there is a tendency towards the supervised-learning approach. Resource allocation planning can explain this ratio (e.g. Xu et al., 2016), which is addressed equally using supervised- and unsupervised-learning algorithms. Further, compared to other ML algorithm groups, genetic-based algorithms tend to be used more for process implementation. Resource-specific or time-related measures are calculated using mainly publicly available event logs for model assessment of resource allocation planning applications, such as r-precision (e.g. Delcoucq et al., 2022) or processing time (e.g. Xu et al., 2016).

6.6. Process monitoring

While other BPM lifecycle phases work with event logs of fixed sizes, process monitoring applications receive continuous data streams in practice. Still, most of the recent process monitoring approaches are developed based on fixed-sized event logs (e.g. Brunk et al., 2021; Evermann et al., 2017; Heinrich et al., 2021). Because of their public availability for research purposes and their static properties (e.g. a static number of activities), event logs are preferred over event streams in research. These event logs are generally not limited to the control flow and include time, resource, or data-related context information. Additionally, some approaches integrate domain knowledge.

Monitoring applications rely on established techniques to extract features from event log data. Examples of such encoding techniques are the boolean, frequency, or index-based technique (e.g. Leontjeva, Conforti, Di Francescomarino, Dumas, & Maggi, 2016). Monitoring applications incorporate features that are created manually from event log data that describe certain aspects of a business process, facilitating the learning of accurate ML models. For instance, temporal features can be created based on timestamps of events, such as the time from the first event of a process instance to the current event of the same process instance (e.g. Tax et al., 2017). Another example is creating inter-case features, such as the number of resources that perform a certain activity (e.g. Senderovich et al., 2017). Feature selection is applied in this phase. For example, to learn accurate models by choosing an event log's most relevant context features (e.g. Alves et al., 2022).

Regarding model building, monitoring applications cope primarily with supervised learning, perhaps because ML is used to learn mappings from process instances or prefixes (i.e. subsequences of process instances) to targets (e.g. Maggi, Di Francescomarino, Dumas, & Ghidini, 2014). Monitoring applications adopt deep learning, ensemble learning, multi-task learning, online learning, and transfer learning. In predictive and prescriptive business process monitoring, process instances from event logs are converted into prefixes to create predictions or prescriptions as early as possible, increasing the volume of data in an event log. As deep learning profits from large data volume, ML applications in this area often use deep-learning models, such as for predicting the next activity, as Evermann et al. (2017) do. Ensemble learning is adopted for addressing all tasks of the monitoring phase using the outcome of several models, which are trained with supervised-learning algorithms (e.g. Metzger et al., 2019). The concept of multi-task learning is used when several process indicators should be monitored, leading to the use of, for instance, DNN models that are designed to approximate multiple learning targets simultaneously (e.g. Tax et al., 2017). The concept of incremental learning assumes a continuous data stream as input, so the concept is present in the process-monitoring phase. Concept-drift detection, in particular, appears with incremental learning (e.g. Baier et al., 2020). Because concept-drift detection aims to identify changes in labels or distributions in event stream data, it is a good match for these approaches. Only a few studies use transfer learning and all these studies are realised with DNNs that distinguish between representation learning and fine-tuning (e.g. Pfeiffer et al., 2021). After learning a generic representation from scratch, the resulting model can be fine-tuned for a desired task. A pre-trained model can also be fine-tuned to transfer from the general representation to the task. Instance, tree, regression, and NN-based algorithms are often employed in the process-monitoring phase because they include popular supervised ML algorithms that can be used for a variety of prediction tasks (e.g. the tree-based algorithm C4.5). Bayesian algorithms are also common in predictive business process monitoring. This is because transition probabilities are often learned with a Bayesian algorithm for a given process model and a certain prediction task before the process model is augmented with the learned probabilities. In doing so, the process model is transformed into a predictive model. Reinforcement-learning algorithms are used for a few process-monitoring tasks, such as next-best-action recommendations in prescriptive business process monitoring, as in Khan et al. (2021), or dynamic resource allocation, as in Firouzian et al. (2019a). Because of the required mapping between actions and a certain state or change of state in a system or an environment, reinforcement learning is particularly well suited to these tasks. Clustering algorithms are also used in predictive business process monitoring to train accurate predictive models per cluster, as Di Francescomarino et al. (2016) do. Additionally, some studies focus on the explanation of ML applications, as Galanti et al. (2020) do by applying SHAP for predictive business process monitoring.

For some tasks, such as predictive business process monitoring, the quality of ML models can be directly assessed. Because of the availability of ground-truth labels in event logs (e.g. next activities), established metrics from supervised ML research can be calculated. For other tasks, model assessment must consider metrics other than these standard metrics. For example, prescriptive monitoring applications take KPIs, often defined based on context features, like cost savings (e.g. Bozorgi et al., 2021) or expected throughput-time (e.g. Weinzierl, Dunzer, et al., 2020) into account for model assessment. Also, measuring time-related metrics is important in this phase as models are applied in running business processes. For example, prediction time (Zhao et al., 2016) or earliness (Teinemaa et al., 2016) are such metrics. For model assessment mainly publicly available event logs are used. Little approaches use own event logs. Further, monitoring applications use a split validation strategy more often than a cross-validation strategy compared to other BPM lifecycle phases. This preference towards split validation is due to retaining the process data's natural structure (Weytjens & Weerdt, 2021).

6.7. Derived findings

The results described in the previous sections are summarised in Table 3 for each BPM lifecycle phase and ML model development phase. Based on these synthesised results, we derived overarching findings.

Finding 1: ML applications mainly address the event-data-intensive BPM lifecycle phases process discovery, analysis, and monitoring. A look into the number of reviewed papers shows that most ML applications address process monitoring, analysis and discovery. Due to the availability of input data, especially process data in the form of event log data, ML applications prevail in these phases. Therefore, various prediction and prescription tasks have been addressed in predictive and prescriptive business process monitoring, various pattern detection and trace clustering tasks in process analysis, and various model discovery tasks in process discovery. On the contrary, for example, it is more challenging to develop ML applications in non-event-data-intensive phases, where process identification relies on documentation, process redesign on redesign heuristics and practices, and process implementation on work plans.

Finding 2: ML applications along the BPM lifecycle mainly use event logs with limited process context. Most ML applications in BPM are developed based on publicly available event logs. These event logs document a business process or its sub-types and include a limited amount of context data. However, in organisations, information systems (e.g. ERP systems) can add more context data to event logs than publicly available event logs. That is important for ML applications of BPM lifecycle phases, where context information plays a role, such as process monitoring or analysis.

Finding 3: New forms of process data enable new use cases. New process data forms have led to the development of ML applications that address new BPM tasks. However, new BPM tasks can already be addressed if an event log includes a certain context attribute, such as the resource attribute for the tasks resource allocation planning in process identification or dynamic resource allocation in process monitoring. In the case of unstructured forms of data, event log creation in process identification is possible with video or text data, and process model extraction in process discovery can be addressed with text or image data. Other BPM tasks, such as (semi-)automatic process redesign, are enabled with more specific forms of data like business process designs.

Finding 4: Integrating domain knowledge can advance ML applications. Domain knowledge in various forms has been integrated into ML applications to fulfil different purposes. For example, process models are integrated into prescriptive-business-process-monitoring applications to ensure the conformance of next best actions (e.g. Weinzierl, Dunzer, et al., 2020), constraints integrated into trace-clustering applications to control the clustering of traces (e.g. De Koninck et al., 2021), or logical rules integrated into predictive-business-process-monitoring applications to improve the prediction performance (e.g. Di Francescomarino et al., 2017).

Finding 5: Common feature extraction techniques are only used for a few BPM tasks. Considering feature extraction techniques of ML applications across BPM tasks, we observe that ML applications apply the same feature extraction techniques for a few tasks. In contrast, specific feature extraction techniques were proposed for the other tasks. For example, ML applications for predictive business process monitoring often use the same techniques to realise feature extraction from event log data. These techniques facilitate the development of predictive-business-process-monitoring applications, as finding an appropriate technique to extract features from a given event log can be challenging. Examples of such techniques are boolean, frequency, or index-based sequence encoding (Leontjeva et al., 2016).

Finding 6: Deep learning is the dominant ML concept in BPM. Deep learning is the most adopted ML concept in BPM. The main reason for this is that the architecture of DNN models can be constructed flexibly depending on the desired tasks (Goodfellow et al., 2016). For example, autoencoder models reproducing the input are constructed in process identification to create process representations (e.g. Guzzo et al., 2021). In contrast, GNN models considering event dependencies in the form of edges are constructed in process discovery to detect generalisable process models (e.g. Sommers et al., 2021).

Finding 7: ML applications mainly focus on improving performance. ML applications in BPM are developed to achieve a performance improvement. A performance improvement can be a higher prediction or detection performance, a lower training or inference time, or a trade-off of various performance criteria.

Finding 8: Explainability of ML applications is primarily addressed in process monitoring. Although BPM research already develops explainable ML applications based on approaches from the field of XAI,[4] these ML applications are mainly developed for predictive business process monitoring (Stierle, Brunk, et al., 2021). Specifically, when DNN models are used for predictive business process monitoring, which are perceived as black boxes, ML applications aim to make model decisions transparent for process users or other process stakeholders.

Finding 9: Model assessment of ML applications is mainly done with data-based metrics. ML applications in BPM are typically assessed with data-based metrics. Common supervised ML metrics, refinements of such, or KPIs are used in BPM tasks, based on supervised learning. In contrast, model assessment is often realised indirectly for BPM tasks where no supervised learning is addressed. That is, a not-supervised-learned ML model is applied, and then clustering or process discovery metrics are calculated based on the model output. In doing that, model assessment can be data-based, even if no label information is available.

Finding 10: Most ML applications along the BPM lifecycle use benchmark event logs. Our reviewed papers show that ML applications are mostly evaluated using benchmark experiments. Further, most benchmark experiments use publicly available event logs. BPM research, in particular predictive business process monitoring research, has elaborated on many ways to address prediction tasks with those event logs. Therefore, identifying accurate ML models for a business process captured in one of the publicly available event logs is consequently quite simple. When addressing a different business process, it is hard to identify promising approaches based on those benchmarks, as they only cover a selection of possible process types and characteristics. Consequently, transferring the acquired findings and insights into practice is limited.

[4] The focus of this paper is on ML, a sub-field of AI. However, as some concepts or research areas are branded with the term "AI", we use "AI" rather than "ML".

Table 3
Overview of synthesised results and derived findings.

	Data input	Feature engineering	Model building	Model assessment
Process identifica-tion (12 paper)	• Mostly event logs with time- and resource-related context information • Other dataset types for *event log creation* (e.g. raw event, video, and text data)	• For *process representation creation and learning* language models based on artificial NNs (e.g. word2vec or doc2vec) or artificial NNs with embedding layer	• Besides supervised and unsupervised learning relatively often self-supervised learning for *process representation creation and learning* • Around half of the ML applications adopt deep learning • Almost all ML applications use an artificial NN	• For *process representation creation and learning* and *event log creation* task clustering models (k-Means) and clustering metrics (e.g. rand index or mutual information score) • Mainly publicly available event logs (e.g. BPI challenge event logs)
Process discovery (45 paper)	• Mostly event logs without context information • Sometimes other dataset types for *process model extraction* (e.g. text or image data)	• For *process model extraction* language models based on artificial NNs (e.g. word2vec or BERT) • For other tasks, feature engineering refers to the transformation of control-flow data into certain instance types (e.g. activity sequences or activity graphs)	• Supervised and unsupervised at about the same rate; unsupervised more common because of lack of label data (e.g. *discovery of declarative process model*) • Occurrence of deep learning and transfer learning for *process model extraction* more than once • Relatively often Bayesian-based, clustering-based, rule-based, and genetic-based	• Often process discovery metrics (e.g. precision, generalisation, or fitness) • Often own event logs, but also publicly available logs
Process analysis (44 paper)	• Mostly event logs with time-, resource-, or data-related context • Rarely other data types (e.g. text data or set of process models) • Sometimes domain knowledge integrated	• Different approaches; no standards regarding feature extraction, creation, or selection	• Supervised and unsupervised at about the same rate • Deep Learning (various tasks), ensemble learning (*pattern detection*), active learning (*pattern detection and trace clustering*), meta Learning (*pattern detection*), and multi-view learning (*pattern detection*) more than once • Often clustering; relatively often rule-based (as in process discovery)	• Often common performance metrics from supervised ML • Metrics for *trace clustering* often process discovery metrics (e.g. fitness, precision, or generalisation) or clustering metrics (e.g. cluster set entropy or mutual information score) • Mainly publicly available event logs, but sometimes also own event logs
Process redesign (3 paper)	• Event logs for *redesign suggestion detection* • Business process designs for *(semi-) automatic process redesign*	• Feature engineering generally not explicitly described	• Only supervised learning • Genetic-based for *(semi-) automatic process redesign* and artificial-NN-based for *redesign suggestion detection*	• Common performance metrics from supervised ML and KPIs (e.g. customer satisfaction) • Mainly publicly available event logs
Process implementa-tion (6 paper)	• Event logs with resource information for *resource allocation planning* • Other data set types can be found (e.g. set of business process problems)	• Feature extraction sometimes described, but common patterns cannot be observed	• Only supervised and unsupervised learning with a tendency to supervised learning as *resource allocation planning* is addressed about the same rate with supervised and unsupervised learning	• Resource-specific measures (e.g. r-precision) or time-specific measures (e.g. processing time) • Mainly publicly available event logs
Process monitoring (90 paper)	• Continuous data streams in theory, but in practice mostly event logs with fix size • Event log often with time-related, resource-related, or data-related context • Sometimes domain knowledge integrated	• Certain techniques for feature extraction from event logs (e.g. event- or sequence-encoding techniques) • Manual feature creation from event log data (e.g. temporal or inter-case features) • Feature selection sometimes applied (e.g. to select relevant context features)	• Mostly supervised learning • Occurrence of deep learning (various tasks), ensemble learning (various tasks), multi-tasking learning (various tasks), incremental learning (various tasks), and transfer learning (*predictive business process monitoring*) more than once • Often NN-based, instance-based, and tree-based, regression-based, Bayesian-based ML applications for *predictive business process monitoring*, reinforcement-learning-based for few process monitoring tasks (e.g. *prescriptive business process monitoring* or *dynamic resource allocation*), and clustering-based for *predictive business process monitoring* • Explainability of ML applications is sometimes focused	• Often established performance metrics from supervised ML research (for some tasks) • Often metrics defined based on context features for *prescriptive business process monitoring* (e.g. cost savings) • Relatively often time-related metrics (e.g. prediction time or earliness) • More often split validation than cross-validation compared to other phases • Mainly publicly available event logs; seldomly own event logs
Derived findings	**F1:** ML applications mainly address the event-data-intensive BPM lifecycle phases process discovery, analysis, and monitoring **F2:** ML applications along the BPM lifecycle mainly use event logs with limited process context **F3:** New forms of process data enable new use cases **F4:** Integrating domain knowledge can advance ML applications	**F5:** Common feature extraction techniques are used for a few BPM tasks	**F6:** Deep learning is the dominant ML concept in BPM **F7:** ML applications mainly focus on improving performance **F8:** Explainability of ML applications is primarily addressed in process monitoring	**F9:** Model assessment of ML applications is mainly done with data-based metrics **F10:** Most ML applications along the BPM lifecycle use benchmark event logs

Table 4
Future research agenda.

Research direction		Based on
Data input		
1	ML applications for event-data-intensive BPM lifecycle phases (i.e. process discovery, analysis and monitoring) received much attention. Exhibiting unstructured data enables the development of ML applications for less event-data-intensive BPM lifecycle phases, where especially process data in the form of event log data are not or only to a limited extent available (i.e. process identification, redesign, and implementation).	F1
2	Most ML applications along the BPM lifecycle mainly use event logs with limited process context. However, in real-world cases, event logs contain more information about processes; for example, machine parameters in manufacturing. To overcome that, future research should develop ML applications considering the enterprise process network and their data.	F2
3	New forms of process data enable new use cases. As object-thinking is gaining attention in practice and academia, process data is increasingly stored as object-centric event logs. Object-centric event logs include object dependencies and thus enable new use cases. Therefore, future research should investigate the development of ML applications for object-centric event logs.	F3
4	Business processes depend on their execution domain and integrating domain knowledge can advance ML applications. Process rules and heuristics may be known in advance. Therefore, in future research, the integration of domain knowledge into ML applications should be further investigated.	F4
Feature engineering		
5	Most feature extraction techniques used in ML applications for BPM are created paper for paper and common feature extraction techniques are only used for a few BPM tasks. Therefore, the development of general feature extraction techniques and the use of those is a direction for future research. For this, generative AI approaches, such as large language models, are promising.	F5
Model building		
6	Deep learning is the dominant ML concept in BPM research. However, beyond deep learning, there are ML concepts that have received little attention in BPM research but are promising to advance ML applications and unlock future research. For example, such an ML concept is transfer learning in the context of standard business processes.	F6
7	ML applications strongly emphasise improving process performance, which can be unsuitable for successful use in practice. Therefore, incorporating aspects of ethical AI in into ML applications, including transparency, justice, fairness, non-maleficence, responsibility and privacy, is a promising direction for future ML research in BPM.	F7
8	While explainability is primarily addressed in the development of ML applications for process monitoring tasks, future research should consider explainability in novel ML applications for other tasks along the BPM lifecycle. The use of XAI approaches and the symbiosis with domain knowledge are important aspects in this future research direction.	F8
Model assessment		
9	Model assessment in ML applications in BPM is mostly done using standard data-based metrics. Beyond data-based assessment, human-centric and economic metrics related to the BPM domain may fit the model assessment more accurately. Therefore, future research should develop such metrics for model assessment for BPM tasks (e.g. regarding efficiency and explainability of ML models).	F9
10	Even though experiments with benchmark event logs are the de facto standard for assessing the utility of ML applications in BPM, generalisability and transfer of findings and insights can be limited. Therefore, future research should propose new approaches to support assessing the utility of ML applications. For example, promising approaches include the elaboration of methodological and technical guidance for benchmark experiments or the application of qualitative and quantitative methods to gain further insights from practitioners.	F10

7. Discussion

This section discusses future research directions we derived from our findings, implications of our literature review for research and practice, and limitations of our study.

7.1. Future research directions

We propose ten future research directions to advance research on ML applications in BPM, as summarised in Table 4. We identified these research directions based on the findings from our literature review and structured them along the phases of the ML model development process.

Research direction 1: According to Finding 1, ML applications in BPM mainly address tasks in the event-data-intensive BPM lifecycle phases of process discovery, analysis and monitoring, where process data in the form of event log data is typically available. Therefore, we propose as future research direction the examination of other data sources from different enterprise systems and the investigation of approaches to induce structure into unstructured and semi-structured data. Consequently, future research can develop novel ML applications for the less event-data-intensive BPM lifecycle phases (i.e. process identification, redesign, and implementation).

To make unstructured process data in these lifecycle phases usable, techniques from the natural language processing (NLP) (e.g. Otter, Medina, & Kalita, 2020) or computer vision (e.g. Voulodimos, Doulamis, Doulamis, & Protopapadakis, 2018) field, which are deeply interconnected with ML approaches, could facilitate the development of novel ML applications. For example, setting up organisational business process landscapes in process identification with NLP techniques based on textual documentation (e.g. van der Aa, Carmona Vargas, Leopold, Mendling, & Padró, 2018), suggesting and testing redesign opportunities in process redesign with computer vision techniques based on video recordings (e.g. Mustansir et al., 2022), and supporting the implementation of business processes in suitable information systems with NLP techniques based on text-based work plans (e.g. Xu et al., 2016).

Research direction 2: In accordance with Finding 2, ML applications along the BPM lifecycle mainly use event logs with limited process context. Therefore, we propose as future research direction the development of novel ML applications, which consider the enterprise process network instead of isolated business processes (Oberdorf et al., 2023). In doing that, ML applications receive input data from multiple data sources, including control-flow information from different business processes and process context information related to the business processes. As more context is considered in such ML applications, the selection of valuable context information via feature selection increases in importance. Finally, as business processes can span multiple organisations, process networks are not limited to one organisation. ML applications could thus span across multiple organisations, providing insights and recommendations for supply chain processes.

Research direction 3: Based on new forms of process data (Finding 1), including object-centric event logs, future research should explore further use cases to develop novel ML applications. Currently, both process-mining vendors and academia are adjusting to object-thinking in BPM (e.g. Li, de Murillas, de Carvalho, & van der Aalst, 2018), which originated in artefact-centric workflow modelling (Nigam & Caswell, 2003). Consequently, process data are more and more stored as object-centric event logs. "Object-centric" is a process paradigm, after which an instance of a business process is not performed in isolation like in the process-instance-centric process paradigm, but interacts via objects with other instances of business processes (van der Aalst, 2019). According to the object-centric process paradigm, an object-centric event log stores object dependencies, which are neglected in traditional event logs (van der Aalst, 2016). Therefore, object-centric event logs reveal new use cases, such as predicting remaining object interactions in the next 24 h, detecting incorrect object interactions, or repairing missing object interactions in object-centric event log data.

Research direction 4: Derived from Finding 4, integrating domain knowledge into model building can advance ML applications. Therefore, the integration of domain knowledge into ML applications should be further investigated. For example, this includes (i) the integration of multiple entities of domain knowledge, (ii) the amount and form of integrated domain knowledge required for solving certain learning problems, and (iii) the integration of profound domain knowledge from BPM.

Concerning (i), each entity can express its domain knowledge as "logical rules, constraints, mathematical equations [...], probability distributions, similarity measures, knowledge graphs or ontologies" (Folino & Pontieri, 2021, p. 86). These entities can also be related to each other in diverse ways or be structured into various areas or levels. Predicting future process behaviour could benefit from integrating domain knowledge as employee-related rules and a company-related ontology into ML applications. Concerning (ii), ML research contains work that proposes approaches that allow controlling which domain knowledge should be integrated into a specific situation into an ML algorithm. For example, Maier et al. (2019) propose operators to incorporate prior knowledge into ML algorithms in a controlled manner. The development of future ML applications could include control mechanisms to handle intended exceptional process flows that, for example, mitigate risks. Concerning (iii), profound domain knowledge from the BPM domain could be present as redesign heuristics, or topical and organisational constraints (e.g. Dumas et al., 2018). The integration of BPM knowledge could aid in automatically suggesting improved process models and definitions.

Research direction 5: According to Finding 5, common feature extraction techniques are only used for a few BPM tasks. Therefore, future research should propose general feature extraction techniques for BPM. Particularly interesting are tasks in BPM lifecycle phases for which various specific techniques have been proposed (e.g. pattern detection in process analysis). The common feature extraction techniques should also consider the extraction of features from dataset types apart from event logs (e.g. text or video) (Kratsch et al., 2022).

To realise common feature extraction techniques, the use of language models seems to be a promising direction. In our reviewed papers, we observed that (large) language models (e.g. word2vec and BERT) are increasingly used in ML applications either as primary components to solve BPM tasks (e.g. process representation creation and learning (e.g. De Koninck et al., 2018) or process model extraction (e.g. Qian et al., 2020)) or as secondary components to support other BPM tasks through learning precise representations of event log or text data (e.g. predictive business process monitoring (e.g. Teinemaa et al., 2016) or pattern detection (e.g. Junior et al., 2020)).

When pre-trained large language models (e.g. the generative pre-trained transformer 4 (GPT-4) (OpenAI, 2023)) are fine-tuned and evolve from the pure ML model to a system (e.g. a conversational agent like ChatGPT (Guo et al., 2023)) or application (e.g. content generation like search engine optimisation) (Feuerriegel, Hartmann, Janiesch, & Zschech, 2023), they open up new possibilities for ML-based BPM research (Vidgof, Bachhofner, & Mendling, 2023). For example, for predictive business process monitoring, conversational agents can be used during process execution, and process users can ask questions to these agents about different aspects of a running process instance. Another example refers to redesign suggestion identification, for which conversational agents can be asked to reveal innovation opportunities for process improvement. Moreover, for process implementation, large-language-model-based applications solving the problem of routine task automation are a new opportunity for ML-based BPM research. Finally, as large language models refer to a type of model used in the context of generative AI (Feuerriegel et al., 2023), other model types, such as diffusion probabilistic models, GANs models, or variational autoencoder models, are promising for developing novel ML applications in BPM. Additionally, models considering data modalities beyond text (e.g. image, audio, and code) or combinations of them (e.g. image-to-text or audio-to-text) are promising for future BPM research on ML applications.

Research direction 6: In accordance with Finding 6, deep learning is the dominant ML concept in BPM. However, beyond deep learning, we propose as future research direction the development of novel ML applications considering ML concepts that have received little attention in BPM research but are promising to advance ML applications. This includes in particular the ML concepts transfer learning, federated learning, causal learning, and neuro-symbolic AI.

The concept of transfer learning (e.g. Pan & Yang, 2009) is promising for BPM research as BPM software provides reference models for many standard business processes (e.g. order to cash) (van Dongen, Jansen-Vullers, Verbeek, & van der Aalst, 2007), and such standard processes are similar in many organisations and organisational units. For example, in such a homogeneous scenario, a model for predicting remaining time could be trained in an organisational unit before this unit could pass the pre-trained model on to other units that do not have enough process data to train models themselves. First works in BPM research make use of transfer learning to address multiple prediction targets given an event log for comprehensive predictive business process monitoring (e.g. Pfeiffer et al., 2021) or to integrate additional context information for accurate discovery of declarative process models (e.g. López et al., 2021).

The concept of federated learning (e.g. Yang et al., 2019) is worth investigating for BPM research as processes can be distributed and go beyond the boundaries of one organisation. For such business processes, federated learning provides promising approaches to training shared ML models using data from multiple owners – while keeping all training data local and private – to train powerful ML models. Multi-organisational process optimisation could also be addressed using federated learning, and small companies with only a small amount of event data could federate to learn predictive models for common processes while keeping their data private.

The concept of causal learning (or causal ML) is promising for BPM research as it aids in better understanding business process improvement by formalising the data-generation process as a structural causal model and enabling to reason about the effects of changes to this process (intervention) and what would happened in hindsight (counterfactuals) (Kaddour, Lynch, Liu, Kusner, & Silva, 2022). Some research in process analysis (e.g. Bozorgi et al., 2020) and process monitoring (e.g. Bozorgi et al., 2021) already demonstrates the benefit of using causal supervised ML. However, the high number of additional ideas in causal ML, including causal generative modelling, causal explanations, causal fairness, and causal reinforcement learning (Kaddour et al., 2022), provide promising avenues for future research.

The concept of neuro-symbolic AI (or neuro-symbolic computation) is worth investigating for BPM research as it provides new principles, concepts, and methods for integrating domain knowledge into ML applications. In fact, neuro-symbolic AI emerged in AI research to overcome the challenge of integrating learning and reasoning (Garcez et al., 2015). For this purpose, neuro-symbolic AI combines robust connectionist machines (e.g. neural networks) with sound, logical abstractions (e.g. logical rules) (Garcez et al., 2022). For example, Pasquadibisceglie, Castellano, et al. (2021) do so as one of the first with a neuro-fuzzy model in predictive business process monitoring.

Research direction 7: According to Finding 7, ML applications mainly focus on improving performance. However, a pure consideration of performance can be insufficient for successful use in practice, and future BPM research should address the development of novel ML applications that incorporate aspects of ethical AI. The field of AI ethics, often called trustworthy AI, has emerged in response to growing concerns about the impact of ML applications among other issues (Kazim & Koshiyama, 2021). AI ethics comprises such aspects as transparency, justice, fairness, non-maleficence, responsibility, and privacy (Feuerriegel, Dolata, & Schwabe, 2020; Jobin, Ienca, & Vayena, 2019; van der Aalst, Bichler, & Heinzl, 2017), and its application regularises the impact of ML applications.

However, despite AI ethics' importance for ML applications on the individual level, BPM research has discussed AI ethics only in theory (Mendling et al., 2018). Further, BPM research addresses fairness-aware process mining (Qafari & van der Aalst, 2019) and demonstrates how process mining can be used to analyse and ensure AI's ethical compliance (Pery, Rafiei, Simon, & van der Aalst, 2022), but those approaches consider fairness and ethical compliance from a process-instance perspective in specific situations. Aspects of AI ethics on the individual level are not yet incorporated in developing ML applications for BPM.

One way to address this gap could be to consider fairness in developing an ML application predicting process outcomes, which can be robust to biases. For example, a credit application might be rejected because of discrimination in a loan application process. Another example refers to privacy, which should be considered in developing an ML application for detecting deviations. Deviations may present an incorrect behaviour of a specific worker to every process analyst without restrictions.

Research direction 8: In accordance with Finding 8, explainability is primarily addressed in the development of ML applications for process monitoring tasks. As explainability is useful and not restricted to process monitoring tasks, we propose as a future research direction the consideration of it in the development of novel ML applications for other tasks along the BPM lifecycle.

For example, trace clustering, to determine why certain traces are assigned to clusters. De Koninck, De Weerdt, et al. (2017) already propose one of the first approaches in this direction. Following Ko and Comuzzi (2023), explainability is also relevant for anomaly detection, as explanations aid in understanding why a detected instance is an anomaly. Organisations can also be interested to understand why an ML model for (semi-)automatic redesign decides that a process should change. As an additional point of this research direction, we subscribe to Bauer, Hinz, van der Aalst, and Weinhardt (2021)'s recommendation of merging XAI with domain knowledge using human-in-the-loop approaches (Wu et al., 2022; Zanzotto, 2019), as first studies in the BPM domain demonstrate (e.g. Junior et al., 2018).

Research direction 9: According to Finding 9, model assessment of ML applications is mainly done with data-based metrics. Therefore, we propose as future research direction the development of additional human-centric and economic metrics that measure BPM-specific aspects of ML applications, that are relevant for process stakeholders but cannot be directly obtained from the underlying data.

One of these aspects is the effectiveness of ML applications, that is, how useful the output of an ML application is for process users and other process stakeholders. For example, in prescriptive business process monitoring, some approaches recommend the next best actions and report KPI improvements (e.g. cost or time savings). However, these approaches do not consider how useful the recommended actions are for process users or other stakeholders (e.g. Weinzierl, Zilker, Stierle, Matzner, & Park, 2020).

Another aspect is the explainability of ML applications, that is, how explainable ML applications are for process users or other stakeholders. For predictive business process monitoring, where explainability of ML applications is mostly addressed in BPM research, the quality of explanation is often assessed via a demonstration of created explanations without any reference to process users or other stakeholders (Stierle, Brunk, et al., 2021). However, the focus of the explanations should be on the humans who execute the processes rather than those who develop them. This focus is important as developers may lack the knowledge to understand such explanations (Bauer et al., 2021). Therefore, metrics are required to measure the explainability of ML applications for process users or other stakeholders.

Research direction 10: According to Finding 10, most ML applications along the BPM lifecycle use benchmark event logs for assessing the utility. However, as generalisability and transfer of findings and insights can be limited using benchmark event logs, future research should investigate new approaches to support assessing the utility of ML applications in BPM.

One approach could be elaborating methodological and technical guidance for benchmark experiments, including event logs, which represent different process types and characteristics. While there is such guidance at pre-print stages for general ML benchmarks, including a framework and data sets (e.g. Romano et al., 2021), BPM requires such methodological approaches for developing ML applications.

Another approach could be assessing the utility of ML applications qualitatively or quantitatively. Involving practitioners in early-stage problem formulation and in the evaluation of an ML application's utility would improve both model assessment and transfer to practice. For example, the utility of an ML application for an organisation can be evaluated in case studies, like Stierle, Weinzierl, et al. (2021) do in the context of process analysis.

7.2. Implications

Our literature review has implications for both research and practice. For research, our literature review provides a knowledge base of ML applications structured into BPM lifecycle phases and BPM tasks. The phases and tasks of the knowledge base are described and defined in the view of ML applications. With this knowledge base, we structure the discipline of ML applications in BPM and show how, where, and when ML applications can be used in BPM. While researchers from the BPM domain can use and extend the knowledge base for their research purposes, researchers new to the field can use it as an entry point. Finally, with an interactive coding table, we give researchers a tool that supports them in analysing ML applications in BPM according to concepts, which are structured into the ML model development process phases.

We stimulate more ML research outside the predominant BPM lifecycle phases. In doing that, we contribute to ML applications developed and used in the field of BPM. Nevertheless, this review covers ML applications from all BPM lifecycle phases, as we found that a comparison of ML applications across all lifecycle phases is essential, as the obtained insights from predominant phases can be transferred to less investigated phases.

With our findings and research directions addressing the utility of ML applications, we propose shifting the focus from prediction-performance-oriented ML research in BPM to more business-value-oriented ML research. In doing that, we aim to improve the transfer of developed ML applications, concepts, or ideas from research into practice.

Our literature review also has implications for practice. We show practitioners various ways and potentials to develop and use ML applications to improve business processes and ultimately create value. For example, ML can be used in process identification to create event logs, which are then analysed with process mining applications to detect bottlenecks in business processes. As another example, ML can be used in process monitoring to predict the outcome of running business processes. This enables process users to intervene in running business processes if the predicted process outcomes cause problems. These examples illustrate the benefits, which ML applications from BPM research can have for practice.

Moreover, we recorded the existence of implementations of ML applications in our coding table. Therefore, practitioners can use our coding table to find the source code of an ML application that addresses a desired BPM task. As the source code from a paper can be used as a basis for implementation, practitioners can build on existing knowledge.

7.3. Limitations

Like all studies, this research project has limitations. First, as with any literature review, a natural limitation is the investigated time period. Thus, derived findings may change over time and new findings can evolve. To omit this issue, we provide the interactive coding table alongside this publication. The content can be updated periodically and new findings may be derived.

Second, the number of BPM-related and ML-related keywords in our search string is unbalanced, as many papers do not use the keyword "machine learning" itself but the name of a specific ML algorithm or ML concept. Including more specific BPM-related keywords (e.g. the BPM lifecycle phase process monitoring) resulted in many research papers retrieved.

Third, because of the high number of papers examined in this review, the papers were divided among three researchers for reading and coding. By applying an iterative approach with discussions, as described in Section 4, we tried to ensure that the researchers had a common understanding of the concepts to be coded. Our inter-coder reliability analysis results (Appendix B) confirm that the researchers used a common understanding in coding the concepts. Still, we cannot guarantee that the coding table is free of inconsistencies.

Fourth, we suppose that some papers do not report all of the concepts we are interested in even when those concepts are part of the underlying ML application.

Fifth, assigning each paper to a BPM lifecycle phase was not trivial. Identifying a clear assignment to a BPM lifecycle phase was challenging when the authors of a paper did not mention a particular phase. In addition, a few papers address more than one phase, and we assigned those papers based on the papers' primary focus.

Finally, we used the ML model development process as a basis on which to structure ML applications, but some ML applications use more than one ML model, which we considered in our coding. However, the composition of ML models in or across ML applications goes beyond the scope of this review.

CRediT authorship contribution statement

Sven Weinzierl: Conceptualization, Data curation, Investigation, Writing – original draft, Visualization. **Sandra Zilker:** Conceptualization, Data curation, Investigation, Writing – original draft, Visualization. **Sebastian Dunzer:** Conceptualization, Data curation, Writing – original draft, Visualization. **Martin Matzner:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The developed coding table is included in the paper.

Acknowledgements

We want to thank Ludwig Grobmeyer for this support in the eary stages of this research project.

Appendix A. Search process specifications

The search process consisted of three steps: First, we developed a *search string* (Appendix A.1). Then, we conducted the collection of papers based on different search activities depending on the used databases Appendix A.2. Lastly, we derived the final set of papers in the (*assessment phase*), applying various *quality and exclusion criteria* (Appendix A.3).

A.1. Search string development

The search string was constructed from two perspectives: the *BPM perspective* and the *ML perspective*.

To cover the *BPM perspective*, we used the search terms *"business process management"* and *"process mining"* (van der Aalst, 2016), as well as *"workflow management"* and *"workflow mining"* for a general view. Then, we used the phases of the BPM lifecycle by Dumas et al. (2018), combined with the term *"business"* to strictly consider business processes (e.g. *"business process monitoring"* or *"business process discovery"*). We also included related concepts like *"process intelligence"* (van der Aalst, 2016). Lastly, we combined the keywords *"data-driven"* and *"business process"* to include publications that address business processes from a data-driven perspective.

For creating the ML perspective of our search string, we first screened fundamental works on ML to get a comprehensive overview of ML approaches. However, none of the investigated works provided an overview with a suitable degree of detail for our work. Consequently, we set out to develop a new overview following fundamental works. First, we defined levels of the overview, that is, ML paradigm, ML concept, and ML algorithm group following Janiesch et al. (2021). Second, we initially defined the categories for the level *ML paradigm* following Mohri, Rostamizadeh, and Talwalkar (2018) and refined these based on further works (see listed works under ML paradigm in Table 1). Third, we initially defined the categories for the level *ML concept* following Goodfellow et al. (2016), Murphy (2012) and Mohri et al. (2018) and refined these based on further works (see listed works under ML concept in Table 1). Fourth, we initially defined the categories for the level *ML algorithm group* and the corresponding representatives for each category following Mitchell (1997), Bishop (2006) and Mohri et al. (2018) and refined these based on further works (see listed works under ML algorithm group in Table 1).

Further, concerning the ML perspective of our search string, we included, where applicable keywords using alternate spellings (e.g. *"semi-supervised learning"*, *"semi-supervised learning"*) or keywords and their abbreviations (e.g. *"support vector machine"* and *"SVM"*). For most keywords, we used the keyword as a fixed term, but for some, we used the word stem to cover a broader spectrum of papers (e.g. "regress*" to find results for *"regression"* and *"regressor"* among other).

We combined the search strings using Boolean operators. The keywords were combined with "OR" within one perspective, whereas two perspectives were combined with "AND" to ensure that both ML and BPM are covered in the results. The search was conducted for the title of the publication, its abstract, and its keywords.

Table A.5
Quality and exclusion criteria applied in the search process.

ID	Criterion
Q1	The paper meets a length requirement of four pages to ensure a profound contribution.
Q2	The paper is accessible to enable further screening and potential coding.
Q3	The paper has been published at a conference or journal (except predatory journals).
Q4	The paper meets the citation of at least two citations per year on average for papers published in 2021 or earlier, while younger papers were included regardlessly.
E1	The paper does not treat a business process, e.g. social network analyses.
E2	The paper presents an ML application for addressing a domain-specific problem, e.g. the prediction of a certain disease in the healthcare domain.
E3	The paper only considers optimisation algorithms (i.e. solvers) without a reference to optimising parameters of an ML model, e.g. linear programming.
E4	The paper uses classical AI approaches, e.g. logic programming.
E5	The paper uses a clustering algorithm without an automated improvement procedure, e.g. a rule-based clustering approach for process model discovery.
E6	The paper uses ML solely for an evaluation purpose, e.g. linear regression used in a quantitative study.
E7	The paper does not provide a novel approach, but has a comparative nature, e.g. comparison of different approaches or is a literature review.

A.2. Conducted search activities per database

To conduct the search and retrieve the relevant documents, three different databases including a wide range of academic publications were used.

Scopus For the identification of all relevant records in Scopus[5], the search string as depicted in Fig. 3 was used in the "Advanced document search". After the documents were found, we refined the search by only including records in English. The results were then exported as a csv file including all bibliometric information.

IEEE Xplore For the database IEEE Xplore[6], the search string in Fig. 3 was used with the "command search" function where only the data fields were updated to fit the syntax. As there is no possibility to search for a paper's title, abstract, and keywords in one search, three individual searches were conducted. Once all records were found for one of the three searches, the items per page was set to the maximum of 100 records. All results per page were then selected using the "select all on page" function. The results and the bibliometric information were exported as a csv file. The same was done for the remaining pages until all records were exported. The search procedure was then repeated for all three data fields (title, abstract, and keywords). The individual csv files were consolidated in one file and all duplicates were removed.

Web of Science For Web of Science[7], the "advanced search" was used for the search string shown in Fig. 3. To search for title, abstract and keywords, the "topic" field was applied. After the search was conducted, the records were filtered to only include English publications. The results were then exported with the full records as an Excel file.

All individual files retrieved from the three databases were then consolidated using MS Excel to all follow the same structure. Lastly, a unique identifier was assigned to each entry of the consolidated list. This list was also used to filter duplicates and to keep track of the papers during the screening phases.

[5] https://www.scopus.com/search/form.uri?display=advanced
[6] https://ieeexplore.ieee.org/search/advanced/command
[7] https://www.webofscience.com/wos/woscc/advanced-search

A.3. Quality and exclusion criteria

In line with Okoli (2015) and vom Brocke et al. (2015), we created quality and exclusion criteria for the full-text assessment of the papers, summarised in Table A.5.

Appendix B. Inter-coder reliability analysis

Table B.6 shows the results of the inter-coder reliability analysis for all binary and categorical dimensions. The dimensions that required free text were all checked by one coder at the end of the coding phase to ensure consistency, for which we used a random sample of 25 papers (approximately 12.5% of the set of papers). The coding was done independently by the three coders based on the concept matrix that we developed during the coding process.

We used three types of metrics to measure the inter-coder reliability. First, we calculated the percentage agreement representing the relative number of papers, for which all coders set the same code (Lombard et al., 2006). We also calculated Krippendorff's α, which is appropriate for two or more coders and can deal with diverse types of data and missing values (Krippendorff, 2018). Lastly, we calculated Fleiss' κ (Fleiss & Cohen, 1973), which is a common metric for three or more coders. We used R (version 4.2.1) with the *DescTools* library for the calculation.

The percentage agreement among the coders is high for all dimensions and concepts (greater than 0.8); only one dimension (*Number of data sets*) is below 0.8, so a single coder coded all papers for this dimension again to ensure consistency in the coding.

The percentage agreement does not account for agreements by chance (Lombard et al., 2006), so we calculated Krippendorff's α and Fleiss' κ, following Neuendorf (2017). Both metrics correct for the probability of agreement by chance (Landis & Koch, 1977). These metrics reveal that almost all concepts are within a very good range, as most values are above 0.8, which resembles a near-perfect agreement (Landis & Koch, 1977), or are between 0.61 and 0.8, resembling substantial agreement (Landis & Koch, 1977).

We observed divergent behaviour for only three concepts: The values for *Feature creation*, *Self-supervised learning*, and *Online/incremental learning* are between 0.45 and 0.5. The results for *Rule-based* and *Other* in the algorithm group section, are not representative due to strongly imbalanced data. However, to ensure the reliability of our coding, one of the researchers checked at the end again the complete coding.

Table B.6
Inter-coder reliability analysis results.

Dimension		% Agreement	Krippendorff's α	Fleiss' κ
Overall		**0.954**		
Evaluation	**Overall**	**0.867**		
	Reference	0.958	0.785	0.787
	Data type	0.917	0.884	0.883
	Number of data sets	0.625	0.686	0.684
	Implementation available	0.833	0.774	0.773
BPM lifecycle phase	**Overall**	**0.972**		
	Process identification	1.000	1.000	1.000
	Process discovery	0.917	0.840	0.839
	Process analysis	0.958	0.907	0.906
	Process redesign	1.000	1.000	1.000
	Process implementation	1.000	1.000	1.000
	Process monitoring	0.958	0.944	0.944
Data input	**Overall**	**0.958**		
	Type of dataset	0.958	0.740	0.738
Feature engineering	**Overall**	**0.900**		
	Feature creation	0.875	0.454	0.454
	Feature selection	0.917	0.639	0.636
ML paradigm	**Overall**	**0.937**		
	Supervised learning	0.833	0.761	0.759
	Semi-supervised learning	1.000	1.000	1.000
	Reinforcement learning	1.000	1.000	1.000
	Self-supervised learning	0.917	0.474	0.470
	Unsupervised learning	0.875	0.814	0.813
	Other	1.000	1.000	1.000
ML concept	**Overall**	**0.983**		
	Multi-task learning	1.000	1.000	1.000
	Active learning	1.000	1.000	1.000
	Online/incremental learning	0.958	0.489	0.486
	Transfer learning	1.000	1.000	1.000
	Ensemble learning	0.958	0.737	0.735
	Deep learning	0.958	0.937	0.936
	Meta learning	1.000	1.000	1.000
	Federated learning	1.000	1.000	1.000
	Multi-view learning	1.000	1.000	1.000
	Other	0.958	1.000	1.000
ML algorithm group	**Overall**	**0.954**		
	Instance-based	1.000	1.000	1.000
	Decision tree-based	0.917	0.840	0.839
	Clustering-based	1.000	1.000	1.000
	Artificial-neural-network-based	0.875	0.830	0.829
	Regression-based	0.958	0.737	0.735
	Bayesian-based	0.917	0.840	0.839
	Rule-based[a]	0.958	NA	NA
	Reinforcement learning-based	1.000	1.000	1.000
	Genetic-based	1.000	1.000	1.000
	Other[a]	0.917	NA	NA
Model assessment	**Overall**	**0.833**		
	Validation strategy	0.833	0.822	0.821

[a] Only the percentage agreement is reported due to strongly imbalanced data.

Appendix C. Coding table

The coding table is shown in Table C.7.

Table C.7
Coding table for ML applications in BPM.

Paper	Evaluation: Data type	Number of datasets	Name of datasets	Statistical method	Further	BPM lifecycle phase: Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Data input: Dataset type	Rel. context information	Feature engineering: Feature extraction
Afflerbach et al. (2017)	Real-world, Synthetic	2	Travel agent process, synthetic logs							x			(Semi-)automatic process redesign	Business process designs (activities, connections, routing decisions)		5 matrices (activity-attribute, object-attribute, the activity-input, activity-output, activity-process-attribute)
Alves et al. (2022)	Real-world	3	Incident, BPIC2013, WFM			x						x	Predictive business process monitoring	Event log	time	one-hot
Appice et al. (2019)	Real-world	2	Helpdesk, BPIC2012									x	Predictive business process monitoring	Event log	time	index-based, window-based, frequency of activities
Appice and Malerba (2015)	Real-world, Synthetic	8	BPIC2013, etm, hospital, isbpm, photo, review, repair, claims	ANOVA					x				Trace clustering	Event log		frequency-based
Bae et al. (2014)	Synthetic	5	Randomly generated processes								x		Execution planning	Set of BP problems		sequence priority, resource assignment
Baier et al. (2020)	Real-world	1	P2P									x	Predictive business process monitoring	Event log		one-hot
Junior et al. (2018)	Real-world	1	Hospital Billing			x						x	Descriptive business process monitoring	Event log	time	edit weighted distance, time-weighted distance, global time
Bazhenova et al. (2016)	Real-world	1	Bank system			x		x					Decision mining	Event log	data	
Berkenstadt et al. (2020)	Real-world, Synthetic	2	DFCI, MM1			x			x				Performance analysis	Event log		
Bernard and Andritsos (2019)	Real-world, Synthetic	4	Helpdesk, BPIC2012, BPIC2013, envPermit			x						x	Predictive business process monitoring	Event log		
Bevacqua et al. (2013)	Real-world	1	Transshipment process									x	Predictive business process monitoring	Event log	time, data	bag, set encoding
Boltenhagen et al. (2019)	Synthetic	3	Synthetic event logs			x			x				Trace clustering	Event log		pseudo-Boolean formula
Borkowski et al. (2019)	Real-world, Synthetic	4	Bosch, Cargo 2000, BPIC2017, CFS									x	Predictive business process monitoring	Event log	data	
Bose and van der Aalst (2009)	Real-world	1	Telephone repair process						x				Trace clustering	Event log		depending on clustering technique
Bozorgi et al. (2020)	Real-world	1	BPIC2017			x			x				Pattern detection	Event log		
Bozorgi et al. (2021)	Real-world	2	Own			x						x	Prescriptive business process monitoring	Event log	time, re-source, data, inter-case	aggregation encoding (activity, resource), last-state (data), one-hot encoding

Paper	Feature creation	Feature selection	Model building – ML paradigm: Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	ML concept: Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	ML algorithm group: Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf. learn.-based	Genetic-based	Other	Model assessment: Metrics	Validation strategy
Afflerbach et al. (2017)			x															x									x		Fitness	
Alves et al. (2022)	x	x	x											x								x							MAE	split
Appice et al. (2019)	x		x										x						x	x			x						MAE, Accuracy	split, cross
Appice and Malerba (2015)							x										x				x								Mean of silhouette width, Computation time, Number of learning iterations, Fitness, Complexity of model	
Bae et al. (2014)			x															x									x		Fitness	
Baier et al. (2020)		x	x								x								x	x		x		x					Accuracy	split
Junior et al. (2018)	x						x				x										x									
Bazhenova et al. (2016)			x															x		x										
Berkenstadt et al. (2020)			x																				x						RMSE, MAE, MARE	split
Bernard and Andritsos (2019)			x				x											x	x		x								Damerau similarity, Average execution time, Accuracy	split
Bevacqua et al. (2013)	x		x															x	x	x	x		x		x				RMSE, MAE, MAPE	cross
Boltenhagen et al. (2019)							x											x			x								Execution time, Maximum distance between a trace and the centroid of its cluster, Number of clusters, Ratio of clustered traces, Inter-cluster distance	
Borkowski et al. (2019)			x								x			x								x							Precision, Recall, Matthews correlation coefficient	split, cross
Bose and van der Aalst (2009)							x											x			x								Fitness, Avg weighted fitness, Complexity	
Bozorgi et al. (2020)			x															x		x									Kullback–Leibler divergence, Action rules	
Bozorgi et al. (2021)	x		x															x		x									Gini, Net value	split

(continued on next page)

S. Weinzierl et al.

Table C.7 (*continued*).

Paper	Evaluation: Data type	Number of datasets	Name of datasets	Statistical method	Further	BPM lifecycle phase: Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Data input: Dataset type	Rel. context information	Feature engineering: Feature extraction
Breuker et al. (2016)	Real-world, Synthetic	2	BPIC2012, BPIC2013	Kruskal–Wallis test		x						x	Predictive business process monitoring	Event log		n-gram
Brunk et al. (2021)	Real-world	2	BPIC2012, BPIC2013			x						x	Predictive business process monitoring	Event log	re-source, data	
Buijs et al. (2013)	Real-world, Synthetic	9	CoSeLoG, loan application process					x					Discovery of procedural process model	Collection of event logs		
Buijs et al. (2012)	Synthetic	3	al2, al2AllSpcNoise, HerbstFig6p34					x					Discovery of procedural process model	Event log		
Camargo et al. (2019)	Real-world	4	BPIC2012, BPIC2013, BPIC2015, Helpdesk			x						x	Predictive business process monitoring	Event log	time, re-source	embedding, normalisation
Camargo et al. (2022)	Real-world, Synthetic	9	BPIC2017 W, BPIC2012 W, INS, ACR, MP, CVS, CFM, CFS, P2P			x			x				Process simulation	Event log	time, re-source, data, inter-case	embedding
Ceravolo et al. (2022)	Real-world, Synthetic	271	Helpdesk, synthetic logs			x						x	Descriptive business process monitoring	Event log	time, cost, re-source	graph-based
Chesani et al. (2009)	Synthetic	NA	Synthetic event logs					x					Discovery of declarative process model	Event log	data	
Ciasullo et al. (2018)	Real-world	1	Italian footwear company				x						Business process outsourcing	NA		
Conforti et al. (2015)	Real-world	1	Claims handling process of insurance company	Person's X'2, Kruskal–Wallis test, Jonkheere's test, Kolmogorov–Smirnov Z two-samples test, Mann–Whitney test		x						x	Prescriptive business process monitoring	Event log	time, re-source, data	observation instance, features from decision point are mapped to risk
Cuzzocrea et al. (2016a)	Real-world	1	BPIC2011						x				Pattern detection	Event log	data	pattern-based encoding
Cuzzocrea et al. (2015)	Real-world	1	BPIC2011						x				Pattern detection	Event log		
Cuzzocrea et al. (2016b)	Real-world	1	BPIC2011						x				Pattern detection	Event log	data	index-based encoding, index-based encoding with HMM-based
Cuzzocrea et al. (2019)	Real-world	2	BPIC2013, Harbor									x	Predictive business process monitoring	Event log	re-source, data	event and trace abstraction functions, single trace vs. window across traces encoding

Paper	Feature creation	Feature selection	Model building – ML paradigm: Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	ML concept: Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	ML algorithm group: Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf. learn.-based	Genetic-based	Other	Model assessment: Metrics	Validation strategy
Breuker et al. (2016)							x											x			x			x					Cross entropy, Accuracy, TP, TN	split
Brunk et al. (2021)							x											x			x			x					Accuracy, F1-Score, Mean, StDev, AD	cross
Buijs et al. (2013)			x															x									x		Fitness, Precision, Simplicity, Generalisation, Similarity, Size, Number of configuration points	
Buijs et al. (2012)			x															x									x		Overall fitness, Replay fitness, Precision, Simplicity, Generalisation	
Camargo et al. (2019)	x		x						x					x								x							Accuracy, Damerau–Levenshtein distance, MAE	split
Camargo et al. (2022)	x		x											x								x							MAE, EMD, RMSE, SMAPE	split
Ceravolo et al. (2022)							x				x										x								Graph Distance, RMSLE, F-Score, Standard deviation	
Chesani et al. (2009)			x															x							x				Errors, Accuracy	split
Ciasullo et al. (2018)					x													x								x			Weights of decision-makers per context and over time	
Conforti et al. (2015)	x		x															x		x									Percentage of faulty instances (mean, median)	
Cuzzocrea et al. (2016a)			x										x		x		x		x	x	x	x		x	x				AUC, G-mean, F-measure	cross
Cuzzocrea et al. (2015)			x										x		x		x		x	x	x	x			x				AUC, G-mean, Recall, Precision	cross
Cuzzocrea et al. (2016b)			x	x						x			x		x		x							x					Accuracy, AUC, TP, FP, Precision, Recall	cross
Cuzzocrea et al. (2019)			x				x				x								x	x	x		x		x				Precision, Recall, F-measure	split

(continued on next page)

Expert Systems With Applications 253 (2024) 124181

Table C.7 (*continued*).

Paper	Evaluation					BPM lifecycle phase							BPM task category	Data input			Feature engineering		Model building: ML paradigm						ML concept										ML algorithm group										Model assessment	
	Data type	Number of datasets	Name of datasets	Statistical method	Further	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Dataset type	Rel. context information	Feature extraction	Feature creation	Feature selection	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf.-learn.-based	Genetic-based	Other	Metrics	Validation strategy
De Koninck and De Weerdt (2019)	Real-world	4	BPIC2012, BPIC2015, BPIC2017, BPIC2018			x			x				Trace clustering	Event log		activity profiles, pairs, 3-grams, PCA							x											x			x								Internal consistency, Process model quality, Computational performance	
De Koninck, De Weerdt, et al. (2017)	Real-world, Synthetic	10	Telecom, Volvo, insurance, admin, purchase, tender, incident, environment, reviewing, incman			x			x				Trace clustering	Event log		binary			x				x											x	x		x								Accuracy, Explanation length, Run time, Normalised mutual information	cross
De Koninck, Nelissen, et al. (2017)	Real-world	1	Newspaper reading data			x			x				Trace clustering	Text						x			x											x			x								F1-Score, Normalised mutual information, Relative improvement	
De Koninck et al. (2021)	Real-world	2	BPIC2015, TABREAD						x				Trace clustering	Event log									x											x			x								F1 score, Relative improvement score, Jaccard index, Percentage of violated must-link constraints	
De Koninck et al. (2018)	Real-world	1	BPIC2015			x	x						Process representation creation and learning	Event log		word2vec, doc2vec						x	x											x			x	x							Rand index, Normalised mutual information, Cosine distance, Behaviour profiles	
de Leoni et al. (2020)	Real-world	1	Dutch reintegration company			x						x	Prescriptive business process monitoring	Event log		frequency-based			x										x						x	x									KPI, Accuracy	split
de Leoni et al. (2013)	Synthetic	21	Synthetic logs					x					Decision mining	Event log	data	observation instances including values of variables in case prior to execution of an specific activity			x															x		x									Relative information gain, Execution time	
de Leoni et al. (2016)	Real-world	1	Illness management			x			x				Performance analysis	Event log	data, resource, time, conformance	trace manipulations and event filter	x		x															x		x									F-measure, Correlation rules, Number of reclamations, Fitness, Mean Absolute Percentage Error (MAPE)	split
de Leoni et al. (2014)	Real-world	1	Event log from employee insurance agency			x			x				Performance analysis	Event log	data, resource, time, conformance	independent variables mapped to dependent variable	x		x															x		x									Accuracy	
De Maio et al. (2016)	Synthetic	1	Own									x	Group decision-making	Event log	data	preference vector					x						x															x			Weights of decision makers over time	
De Weerdt et al. (2013)	Real-world, Synthetic	104	Helpdesk process, CRM process, second-line CRM process, incoming document handling			x			x				Trace clustering	Event log									x			x											x								Cluster set entropy, Weighted average f1-score, Place/transition connection degree	
De Weerdt et al. (2012)	Real-world	1	Incident management process			x			x				Trace clustering	Event log	data, text	frequent keyword features, type information feature			x	x														x		x	x								Accuracy	cross
Delcoucq et al. (2022)	Real-world, Synthetic	3	Hospital, Review, Repair								x		Resource allocation planning	Event log	resource								x											x			x								F1-Score, R-Precision, R-Recall	

(*continued on next page*)

S. Weinzierl et al.

Table C.7 (*continued*).

Paper	Evaluation					BPM lifecycle phase							BPM task category	Data input			Feature engineering		Model building: ML paradigm						ML concept										ML algorithm group										Model assessment	
	Data type	Number of datasets	Name of datasets	Statistical method	Further	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Dataset type	Rel. context information	Feature extraction	Feature creation	Feature selection	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf.-learn.-based	Genetic-based	Other	Metrics	Validation strategy
Delias et al. (2019)	Real-world, Synthetic	1	BPIC2011						x				Trace clustering	Event log	data	features grouped into 5 criteria	x						x											x			x								Accuracy, cyclomatic number (CN), coefficient of connectivity (CNC), Coefficient of network complexity (CNC_k), Density, Performance (Case duration)	
Di Francescomarino et al. (2016)	Real-world	1	BPIC2011	Two-tailed non-parametric Wilcoxon test								x	Predictive business process monitoring	Event log		frequency-based, sequence-based			x				x						x							x	x								F1, Failure rate, Earliness	split
Di Francescomarino et al. (2017)	Real-world	6	BPIC2011, WABO, CoSeLoG, Helpdesk, BPIC2012, BPIC2013			x						x	Predictive business process monitoring	Event log		one-hot			x											x								x							Average Damerau–Levenshtein similarity	split
Di Mauro et al. (2019)	Real-world	3	Receipt phase, BPIC2012, Helpdesk			x						x	Predictive business process monitoring	Event log		embedding, time diff between current and last event			x											x								x							Brier score, Accuracy	cross
Diamantini et al. (2016)	Real-world, Synthetic	2	BPIC2011, synthetic log					x					Discovery of procedural process model	Event log									x											x			x								Diversity, Completeness/Coverage, Representativeness, Frequency of subpatterns	
Effendi and Sarno (2017)	Real-world	1	Online book store management process					x					Decision mining	Event log	data				x															x		x										
Es-Soufi et al. (2016)	NA	NA	Past executions of design processes						x				Pattern detection	Event log					x															x		x										split
Evermann et al. (2017)	Real-world	2	BPIC2012, BPIC2013	Kolmogorov–Smirnov test		x						x	Predictive business process monitoring	Event log	resource	embedding			x											x								x							Mean and standard deviation of training precision and validation precision	cross
Ferreira et al. (2007)	Real-world, Synthetic	2	Own						x				Trace clustering	Event log									x											x			x									
Ferreira and Gillblad (2009)	Synthetic	1	Synthetic event stream			x		x					Discovery of procedural process model	Event stream		symbol sequence							x											x			x			x					G-Score, Execution time	
Ferreira et al. (2013)	Synthetic	1	Purchase process					x					Discovery of procedural process model	Event log	data								x											x			x			x						
Ferreira and Vasilyev (2015)	Real-world	2	BPIC2012, BPIC2013						x				Performance analysis	Event log	time, resource, data	predicates (e.g. performs, flow or user)	x		x															x							x					
Firouzian et al. (2019a)	Real-world	1	BPIC2012									x	Dynamic resource allocation	Event log	time, resource						x													x								x			Averaged cycle time, Entropy (sum of entropy of all work lists at a time t), Variance (of workloads of all resources at time t ->balancing of workload)	
Firouzian et al. (2019b)	Real-world, Synthetic	2	BPIC2012, Own									x	Predictive business process monitoring	Event log	data				x								x		x						x										Accuracy, F1-Score	
Folino et al. (2022)	Real-world	5	BPIC2012, hospital billing, road traffic fines	Friedman test, Nemenyi test, Critical Distance		x						x	Predictive business process monitoring	Event log	time, resource, data	embedding, one-hot	x			x								x		x								x							Accuracy, AUC	split

(*continued on next page*)

Expert Systems With Applications 253 (2024) 124181

Table C.7 (*continued*).

Paper	Evaluation					BPM lifecycle phase							BPM task category	Data input		Feature engineering			Model building ML paradigm						ML concept										ML algorithm group										Model assessment	
	Data type	Number of datasets	Name of datasets	Statistical method	Further	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Dataset type	Rel. context information	Feature extraction	Feature creation	Feature selection	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf.-learn.-based	Genetic-based	Other	Metrics	Validation strategy
Folino et al. (2011)	Real-world, Synthetic	4	Claim handling process, 2 logistic processes, collaboration process									x	Predictive business process monitoring	Event log	data				x															x		x									Accuracy, Confidence	cross
Folino et al. (2012)	Real-world	1	Logistic process			x						x	Predictive business process monitoring	Event log	time, data	list, set, bag	x	x	x															x		x	x				x				RMSE, MAE, MAPE, Precision, Recall, Computation time	cross
Folino et al. (2015)	Real-world	1	BPIC2013						x				Trace clustering	Event log	time, resource, data	t-view (context data and abstract activities)	x		x															x		x	x				x				Fitness, Behavioural precision	cross
Galanti et al. (2020)	Real-world	1	Italian bank account closure process			x						x	Predictive business process monitoring	Event log	resource, data	one-hot			x											x								x							MAE, F1-Score, AUC ROC, AUC PR	split
García-Bañuelos et al. (2014)	Real-world	3	BPIC2012a, insurance claims handling process			x		x					Discovery of procedural process model	Event log		depending on the trace clustering approach used							x											x			x								Size, Structural complexity (CFC, ACD, CNC and density), Time performance	
Gerlach et al. (2022)	Real-world, Synthetic	11	Small, wide, medium, huge, p2p, paper, BPIC2012, BPIC2013, BPIC2015, BPIC2017, BPIC2020			x						x	Predictive business process monitoring	Event log					x											x								x								
Greco et al. (2006)	Synthetic	NA	Order management process, review paper process					x					Discovery of procedural process model	Event log									x											x			x								Soundness	
Greco et al. (2004)	Synthetic	1	Own					x					Discovery of procedural process model	Event log									x											x			x								Soundness	
Guzzo et al. (2021)	Real-world	2	BPIC2015, BPIC2019			x	x						Process representation creation and learning	Event log	time, resource, data	embedding	x	x				x								x								x							Precision, Recall, F score, Adjusted rank index, Adjusted mutual information, Homogeneity score, Completeness score	cross
Ha et al. (2016)	Real-world	1	prBm6						x				Trace clustering	Event log									x											x			x								Precision, Fitness (compare vector with graph representation)	
Hanga et al. (2020)	Real-world	4	Helpdesk, BPIC2012, BPIC2013, RTFM									x	Predictive business process monitoring	Event log		one-hot			x											x								x							Accuracy, Similarity score	split
Harl et al. (2020)	Real-world	1	BPIC2012			x						x	Predictive business process monitoring	Event Log					x											x								x							Accuracy	split
Heinrich et al. (2021)	Real-world	11	BPIC2011, BPIC2012, BPIC2013, Helpesk, EnvLog									x	Predictive business process monitoring	Event log	data	embedding			x											x								x							Accuracy, Precision, Recall, F1 score	cross
Herbst (2000)	Real-world, Synthetic	NA	Own					x					Discovery of procedural process model	Event log		dependency graph							x											x						x					Nodes/edges of result model, No. splits, Time, Correct, Log-likelihood	
Herbst and Karagiannis (2000)	Synthetic	1	Simplified release process in car manufacturing					x					Discovery of procedural process model	Event log									x				x													x					Nodes/edges of result model, Search steps, Time spent, Log-likelihood	

(*continued on next page*)

Table C.7 (*continued*).

Paper	Evaluation					BPM lifecycle phase							BPM task category	Data input		Feature engineering			Model building ML paradigm						ML concept										ML algorithm group										Model assessment	
	Data type	Number of datasets	Name of datasets	Statistical method	Further	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Dataset type	Rel. context information	Feature extraction	Feature creation	Feature selection	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf.-learn.-based	Genetic-based	Other	Metrics	Validation strategy
Herbst and Karagiannis (2004)	Real-world, Synthetic	NA	Different Simulated/Two Industrial logs		Workflow model discovery with different experts and with different models			x					Discovery of procedural process model	Event log									x											x						x					Semantic equality between process models	
Hsieh et al. (2021)	Real-world	1	BPIC2012									x	Predictive business process monitoring	Event log	resource, data	one-hot			x											x								x							Sparsity as average Levenshtein distance, Proximity as euclidean distance	split
Huang et al. (2011)	Real-world	1	Radiology CT-scan examination process			x						x	Dynamic resource allocation	Event log	time, resource, data						x													x								x			Flow time, MSE (with 5% confidence interval), STD	
Huang et al. (2010)	Real-world	1	Radiology CT-scan examination process			x						x	System adaption	Other		process data objects					x													x								x			Reward value, Flow time, Confidence level (90%)	
Huo et al. (2021)	Real-world, Synthetic	6	BPIC2012, BPIC2013, BPIC2017, Loan, Large, Huge									x	Descriptive business process monitoring	Event Log		one-hot						x								x								x							F1-score	split
Jalayer et al. (2020)	Real-world	4	Helpdesk, BPIC2012, BPIC2012 W, BPIC2012O									x	Predictive business process monitoring	Event log					x											x								x							Accuracy, Demerau-Levenshtein	split
Jlailaty et al. (2017)	Real-world	1	250 E-Mails taken from a Ph.D student				x						Event log creation	Text		word2vec, tf-idf		x				x	x											x			x	x							F1-Score, Rand-index, Precision, Recall, Purity	cross
Jung et al. (2009)	Real-world	1	Insurance process						x				Trace clustering	Set of process models		process vector							x											x			x								Cosine similarity	
Jung et al. (2008)	Synthetic	1	Insurance Processes						x				Trace clustering	Set of business process models		process vector							x											x			x								Similarity	
Junior et al. (2020)	Real-world, Synthetic	11	BPIC2012, BPIC2013, BPIC2015, BPIC2017			x			x				Pattern detection	Event log		word2vec			x			x												x	x			x							F score	
Kang et al. (2012)	Real-world	1	Compressive strength process									x	Prescriptive business process monitoring	Tabular data					x															x	x										False warning, Missing warning, True warning, True no-warning, Error rate, Probability of target value, FP, FN, TP, TN	cross
Kazakov et al. (2018)	NA	2	NA								x		Resource allocation planning	NA		BP function features			x															x						x						
Khan et al. (2021)	Real-world	1	MIMIC									x	Prescriptive business process monitoring	Event Log	data						x									x								x				x			Estimated policy value	split
Khodyrev and Popova (2014)	Real-world	2	CarRep, CFactRep						x				Process simulation	Event log	time, data	petri-net-element-based	x	x	x															x		x									Confidence intervals	split
Kim et al. (2002)	Real-world	1	Order processing workflow					x					Discovery of procedural process model	Case base (set of cases in the form of is-a hierarchy), vocabulary base (task names and their nouns/verbs)					x															x	x										Similarity	
Klijn and Fahland (2020)	Real-world	2	BPIC2020, RTP			x						x	Predictive business process monitoring	Event log	inter-case	last state/aggregate/index-based encoding, prefix-length/cluster/single bucketing	x		x										x							x									MAE, Computation time	split
Krajsic and Franczyk (2020)	Real-world	1	BPIC2019						x				Pattern detection	Event log		one-hot						x												x				x							Accuracy, Misclassification, Precision, Recall, F1-Score, Specificity	split

(*continued on next page*)

Table C.7 (*continued*).

Paper	Evaluation					BPM lifecycle phase							BPM task category	Data input		Feature engineering			Model building: ML paradigm						ML concept										ML algorithm group										Model assessment	
	Data type	Number of datasets	Name of datasets	Statistical method	Further	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Dataset type	Rel. context information	Feature extraction	Feature creation	Feature selection	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Relat.-learn.-based	Genetic-based	Other	Metrics	Validation strategy
Krajsic and Franczyk (2021)	Synthetic	4	Loan Application Process Variants									x	Descriptive business process monitoring	Event log								x	x							x					x	x		x							Excess-Mass Score, Mass-Volume-Score	split
Kratsch et al. (2022)	Real-world	1	Own	Friedman test		x	x						Event log creation	Video data										x										x										x		
Lakshmanan et al. (2015)	Synthetic	1	Automobile Insurance Claims Handling Process									x	Predictive business process monitoring	Event log	data				x															x		x									Normalised Log Likelihood	split
Lamma, Mello, Montali, et al. (2007)	Real-world	2	Cervical cancer screening careflow, NetBill					x					Discovery of declarative process model	Event log					x															x							x				Accuracy	cross
Lamma, Mello, Riguzzi, et al. (2007)	Real-world	2	Electronic auction, NetBill					x					Discovery of declarative process model	Interaction log					x															x						x	x					
Lee et al. (2022)	Synthetic	6	Loan application process			x						x	Predictive business process monitoring	Event log	time	one-hot	x		x										x	x						x		x							F1-score	split
Lee et al. (2021)	Real-world	1	Hospital Billing			x						x	Descriptive business process monitoring	Event log		caseid-activity-pairs							x											x			x			x						cross
Lee et al. (2018)	Real-world	2	BPIC2012, BPIC2013									x	Predictive business process monitoring	Event log		frequency-based, one-hot			x															x										x	Precision, RMSE	split
Leno et al. (2018)	Real-world, Synthetic	12	BPIC2015, BPIC2011, synthetic log			x		x					Discovery of declarative process model	Event log	data	violation and fulfilment vectors	x		x				x											x		x	x				x				Support, Confidence, Execution time	
Leno et al. (2020)	Real-world, Synthetic	5	BPIC2011, BPIC2013, BPIC2017, Sepsis, RTFM			x		x					Discovery of declarative process model	Event log	data	violation and fulfilment vectors	x		x										x							x					x				Support, Confidence, Recall, Precision, F-Score	split
Li et al. (2010)	Real-world, Synthetic	2	Synthetic process models		Case study			x					Discovery of procedural process model	Set of business process models		Order matrix							x											x			x								Computation time, Average weighted distance	
Liu et al. (2012)	Real-world	10	Manufacturing processes								x		Resource allocation planning	Event log	time, re-source	frequent 3-itemset							x											x							x				Precision, Time elapsed, Correct prediction count, Strong rules count	cross
Liu et al. (2008)	Real-world	3	Own									x	Dynamic resource allocation	Event log	re-source				x															x	x	x				x					Accuracy	
López et al. (2021)	Real-world	1	NA			x		x					Process model extraction	Text		bert model			x			x						x		x								x							Precision, Recall, F1-Score	split, cross
Lu et al. (2016)	Real-world	1	CORDYS BPM					x					Discovery of procedural process model	Event log	data								x											x			x								Flow condition, Descriptive comparison to related work: Ability to model decomposable cyclic dependencies, to deal with noise	
Maggi et al. (2018)	Real-world, Synthetic	84	BPIC2012, BPIC2013, BPIC2014, RTFM, Sepsis, BPIC2017, synthetic logs					x					Discovery of declarative process model	Event log									x											x							x				Computation time, Memory consumption	
Maggi et al. (2012)	Real-world, Synthetic	NA	Synthetic logs, process from dutch municipality					x					Discovery of declarative process model	Event log		frequent activity sets							x											x							x				Computation time, Conditional-probability increment ratio	
Mannhardt et al. (2016)	Real-world	2	RTFM, Sepsis			x		x					Decision mining	Event log					x															x		x									Fitness, Precision (compare to different methods)	
Márquez-Chamorro, Resinas, Ruiz-Cortés, and Toro (2017)	Real-world	2	BPIC2013, IT Department of Health Service			x						x	Predictive business process monitoring	Event log	time, re-source, data	Window-based	x		x															x									x		Precision, Recall, Specificity, Mean, StDev, F-Measure, AUC-ROC	cross
Martens et al. (2014)	Real-world	1	CoSeLoG					x					Discovery of procedural process model	Set of business process models		graph-based			x															x									x		Graph-Edit Distance, Precision, Reall, F-Measure	
Măruşter et al. (2006)	Real-world, Synthetic	2	Simulated Log, Dutch governmental fine collection					x					Discovery of procedural process model	Event log					x															x							x				F-measure	cross
Măruşter et al. (2002)	Synthetic	5	Own					x					Discovery of procedural process model	Event log					x															x					x						Accuracy	cross
Mehdiyev et al. (2020)	Real-world	3	BPIC2012, BPIC2013, helpdesk									x	Predictive business process monitoring	Event log	re-source, data	n-gram, hashing			x			x								x								x							Accuracy, Precision, Recall, F-measure, MCC, AUC	cross
Metzger et al. (2020)	Real-world	3	BPIC2012, BPIC2017, Traffic			x						x	System adaption	Event log					x		x						x		x	x								x				x			Execution cost	
Metzger et al. (2019)	Real-world	4	Cargo2000, Traffic, BPIC2012, BPIC2017			x						x	System adaption	Event log		one-hot			x						x				x	x								x							Cost savings	
Montani and Leonardi (2012)	Real-world	1	Stoke management process			x						x	Descriptive business process monitoring	Event log	time		x		x															x	x		x								Cluster homogeneity	
Mustansir et al. (2022)	Real-world	3	Own							x			Redesign suggestion identification	Text					x											x								x							F1-Score	cross
Nguyen et al. (2019)	Real-world, Synthetic	2	BPIC2012, BPIC2013			x	x						Data improvement	Event log	time	one-hot encoding, standardising	x					x								x				x				x							Precision, Recall, F1-score	
Nguyen et al. (2016)	Real-world, Synthetic	3	BPIC2013, Receipt, bank						x				Trace clustering	Event log									x											x			x								Weighted average conformance fitness, Weighted average structure complexity	
Nolle, Luettgen, et al. (2018)	Real-world, Synthetic	700	BPIC2012, BPIC2017, P2P, Small, Medium, Large, Huge, Wide			x			x				Pattern detection	Event log	re-source	one-hot						x												x				x							F1-Score, DAE error heatmap	
Nolle et al. (2019)	Real-world, Synthetic	13	BPIC2012, BPIC2013, BPIC2015, BPIC2017, Paper, P2P, Small, Medium, Large, Huge, Gigantic, Wide, Anonymous	Nemenyi post hoc test, Friedman test		x						x	Descriptive business process monitoring	Event log	time, re-source, data	integer, embedding	x		x						x					x								x							Precision, Recall, F1-Score, Variance, Confusion matrix, Critical difference, Anomaly scores	
Nolle, Seeliger, et al. (2018)	Real-world, Synthetic	81	BPIC2012, BPIC2013, BPIC2015, BPIC2017, COMP, P2P, Small, Medium, Large, Huge, Wide			x						x	Descriptive business process monitoring	Event log	time, re-source, data	integer, embedding	x		x						x					x								x							F1-Score, Anomaly scores	
Nolle et al. (2020)	Synthetic	252	Own			x	x						Data improvement	Event log	time, re-source, data				x						x					x								x							F1-Scores, Correction Accuracy, Average Error, Alignment Optimality	
Park and Song (2020)	Real-world	3	Helpdesk, BPIC2012, healthcare service process of South Korean hospital			x						x	Predictive business process monitoring	Event log	time	transition-system-based	x		x											x								x							MAE, Mean absolute percentage error	cross
Park and Song (2019)	Real-world, Synthetic	2	BPIC2012, Own									x	Dynamic resource allocation	Event log	time, re-source	one-hot (activity, resource)	x		x						x					x								x							Accuracy	
Pasquadibisceglie et al. (2019)	Real-world	2	Helpdesk, BPIC2012									x	Predictive business process monitoring	Event log		frequency-based, temporal features			x											x								x							Accuracy	split

(continued on next page)

Table C.7 (*continued*).

Paper	Evaluation					BPM lifecycle phase							BPM task category	Data input		Feature engineering			Model building: ML paradigm						ML concept										ML algorithm group										Model assessment	
	Data type	Number of datasets	Name of datasets	Statistical method	Further	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Dataset type	Rel. context information	Feature extraction	Feature creation	Feature selection	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf.-learn.-based	Genetic-based	Other	Metrics	Validation strategy
Pasquadibisceglie, Appice, Castellano, and Malerba (2020)	Real-world	3	BPIC2012, Receipt Phase, BPIC2013			x						x	Predictive business process monitoring	Event log	time, re-source	rbg encoding	x		x											x								x							Accuracy, Precision, Recall, F1-Score, Standard deviation	cross
Pasquadibisceglie, Appice, et al. (2021)	Real-world	5	BPIC2012, BPIC2013, Receipt, BPIC2017, BPIC2020	Friedman test		x						x	Predictive business process monitoring	Event log		one-hot			x						x					x			x					x							Accuracy, Macro-precision, Macro-recall, Macro-Fscore, Macro-AUCROC, Macro AUCPR	split
Pasquadibisceglie, Appice, Castellano, Malerba, and Modugno (2020)	Real-world	4	Sepsis, BPIC2011, BPIC2012, Production	Nemenyi test		x						x	Predictive business process monitoring	Event log		min–max normalisation			x											x								x							AUC, F-Score	split
Pasquadibisceglie, Castellano, et al. (2021)	Real-world	4	Sepsis, BPIC2011, BPIC2012, Production			x						x	Predictive business process monitoring	Event log		Aggregation encoding			x											x								x							ROC AUC	
Pauwels and Calders (2020)	Real-world	4	BPIC2012, BPIC2015, Helpdesk, BPIC2018			x						x	Predictive business process monitoring	Event log									x											x						x					Accuracy of suffix prediction, Runtime	split
Pauwels and Calders (2021)	Real-world	4	Helpdesk, BPIC2011, BPIC2012, BPIC2015			x						x	Predictive business process monitoring	Event Log		one-hot			x				x				x			x								x		x					Accuracy, Runtime	
Pegoraro et al. (2021)	Real-world	2	BPIC2016, MIMIC									x	Predictive business process monitoring	Event log	text	one-hot encoding, min–max normalisation, exchangeable text encoding model (BoNG, Doc2Vec, LDA, BoW)			x											x								x							F1-Score, MAE	
Pfeiffer et al. (2021)	Real-world, Synthetic	7	Helpdesk, BPIC2012, WE, BPIC2013, CP, BPIC2017, RFP, MobIS			x						x	Predictive business process monitoring	Event log	re-source, time, cost, data	gramian angular fields			x			x						x		x					x			x							Accuracy, MAE	split
Pflug and Rinderle-Ma (2016)	Real-world	1	BPIC2011									x	Dynamic resource allocation	Event log	data								x											x			x								Cumulated throughput time, Average throughput time at activity fetching	
Pham et al. (2021)	Real-world	5	BPIC2012, BPIC2015, BPIC2017, Helpdesk									x	Predictive business process monitoring	Event log	re-source	one-hot			x											x								x							Accuracy	split
Philipp et al. (2020)	Real-world	3	BPIC2012, BPIC2013, German software company dataset									x	Predictive business process monitoring	Event log		positional encoding			x											x								x							Accuracy	split
Polančič et al. (2020)	Real-world	1	Questionnaire	T-test				x					Process model extraction	Image					x									x		x								x							Accuracy	split, cross
Polato et al. (2018)	Real-world	3	RTFM, Helpdesk, BPIC2012									x	Predictive business process monitoring	Event log	data	one-hot, set of activity (set, bag, list), transition-system-based features	x		x															x					x	x					Accuracy, MAPE, RMSPE, MAE	split, cross
Polato et al. (2014)	Real-world	1	RTFM			x						x	Predictive business process monitoring	Event log	data	last, one-hot			x															x	x					x					MAPE, RMSPE	cross

(continued on next page)

S. Weinzierl et al.

Table C.7 (*continued*).

Paper	Evaluation					BPM lifecycle phase							BPM task category	Data input		Feature engineering			Model building: ML paradigm						ML concept										ML algorithm group										Model assessment	
	Data type	Number of datasets	Name of datasets	Statistical method	Further	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Dataset type	Rel. context information	Feature extraction	Feature creation	Feature selection	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf.-learn.-based	Genetic-based	Other	Metrics	Validation strategy
Qafari and van der Aalst (2019)	modified Real-world	1	Hospital Billing			x			x				Performance analysis	Event log	data	situation features			x															x		x									Accuracy, Discrimination	split
Qian et al. (2020)	Real-world	2	Cooking Recipes, Maintenance Manuals	Two-tailed t-test		x		x					Process model extraction	Text		word2vec, bert			x			x			x					x								x							Accuracy, Behaviour similarity	cross
Qiao et al. (2011)	Real-world, Synthetic	5	SAP (industry), SAP (cross industry), SAP (reference), APQC, Real practice					x					Discovery of procedural process model	Set of business process models, process description documents and keywords		vector-space model, graph-based							x											x			x			x					Precision, Running time	split
Rizzi et al. (2020)	Real-world, Synthetic	8	BPIC2011, Claim Management			x						x	Predictive business process monitoring	Event log		frequency, simple, complex		x	x										x							x									Accuracy, AUC	split
Rogge-Solti and Kasneci (2014)	Real-world	1	Dutch surgery log			x			x				Pattern detection	Event log	time								x											x						x					ROC, Accuracy	cross
Rozinat and van der Aalst (2006)	NA	NA	Variety of domains (e.g. healthcare)			x		x					Decision mining	Event log					x															x		x										
Samiri et al. (2017)	Synthetic	1	Own	Auto-correlation, partial auto-correlation, mean percentage of residuals, augmented Dickey–Fuller for unit-root test								x	System adaption	Event log							x													x								x			Mean error, MAE, root-mean-square error, mean percentage error, mean absolute percentage error	
Saraeian et al. (2019)	Synthetic	1	Supply chain dataset									x	System adaption	Tabular data					x															x				x							MAE, relative root-mean square error (RRMSE), root RMSE	
Sarno, Ginardi, et al. (2013)	Synthetic	1	Synthetic process models						x				Trace clustering	Set of business process models		process matrix							x											x			x								Silhouette index, Similarity, Dissimilarity	
Sarno, Sari, et al. (2013)	Real-world	1	Special order case			x		x					Decision mining	Event log	re-source, data				x															x		x										
Sarno and Sungkono (2016a)	Synthetic	NA	Own					x					Discovery of procedural process model	Event log									x											x			x			x					Fitness, Validity	
Sarno and Sungkono (2016b)	Synthetic	NA	Own					x					Discovery of procedural process model	Event log									x											x			x			x					Fitness	
Savickas and Vasilecas (2018)	Real-world, Synthetic	3	MP, DFI, EIMSD						x				Performance analysis	Event log	time, re-source, data	sequence matrix	x						x											x						x					Rate of correct, partial, wrong, and missing activities	split
Seeliger et al. (2021)	Real-world, Synthetic	4	TC-DS18, TC-DS20, BPIC2015, BPIC2019			x	x						Process representation creation and learning	Event log	re-source, data	embedding, one-hot			x											x								x							F1-Bcubed, Fitness, Precision, Simplicity	
Seeliger et al. (2018)	Real-world, Synthetic	293	Own			x			x				Trace clustering	Event log	re-source, data	frequent itemsets, integer							x											x			x								Weighted fitness, Weighted precision, Weighted generalisation, Purity, Adjusted rand index, Cluster set entropy, Graph density, Cyclomatic number, Coefficient of connectivity, Coefficient of network complexity	

(continued on next page)

Table C.7 (*continued*).

Paper	Data type	Number of datasets	Name of datasets	Statistical method	Further	BPM task category	Dataset type	Rel. context information	Feature extraction	Metrics	Validation strategy
Senderovich et al. (2017)	Real-world	2	Israeli emergency department process, manufacturing process			Predictive business process monitoring	Event log	time, resource, data, inter-case	Window-based	MAE, RMSE	split, cross
Senderovich et al. (2015)	Real-world, Synthetic	3	Israeli bank's call centre, Israeli telecommunication company call centre, synthetic log			Predictive business process monitoring	Event log	time	transition-system-based	RASE (prediction error, seconds)	split
Senderovich et al. (2014)	Real-world	1	Call-centre log			Resource allocation planning	Service log			Misclassification rate	split
Shraga et al. (2020)	Real-world, Synthetic	5	Own, hospital, sepsis			Discovery of procedural process model	Event log	time, resource, data		Fitness, Precision, Generalisation	split
Silva et al. (2005)	Synthetic	1	Simulated report process			Discovery of procedural process model	Event log			X^2-test	
Sim et al. (2022)	Real-world	9	BPIC2012-BPIC2018, Steel processing, port logistics		Expert review	Event log creation	Event data		one-hot	Accuracy	split
Sindhgatta et al. (2020)	Real-world	4	BPIC2013, BPIC2012, BPIC2015, helpdesk			Predictive business process monitoring	Event log		one-hot	Accuracy, Attention	split
Sommers et al. (2021)	Real-world, Synthetic	11	PTAndLogGenerator, BPIC			Discovery of procedural process model	Event log		one-hot, trace graph	F-Score, Simplicity score	
Stierle, Weinzierl, et al. (2021)	Real-world	4	BPIC2017, BPIC2018, BPIC2020, sp2020		Case study	Performance analysis	Event log		one-hot	AUC ROC, Sensitivity, Specificity	cross
Tavares et al. (2019)	Real-world, Synthetic	2	Frozen, Healthcare			Descriptive business process monitoring	Event log	time	graph-based	Graph Distance, Variation	
Tax et al. (2016)	Real-world, Synthetic	2	Artificial Digital Photocopier, smart home environment			Process representation creation and learning	Event log	time, resource	n-gram	Levenshtein similarity	cross
Tax et al. (2017)	Real-world	3	Helpdesk, BPIC2012, Environmental permit			Predictive business process monitoring	Event Log	time	one-hot, index-based	MAE, Accuracy	split
Taymouri et al. (2020)	Real-world	3	Helpdesk, BPIC2012, BPIC2017			Predictive business process monitoring	Event log		one-hot	Weighted average accuracy, weighted average MAE	split
Taymouri, La Rosa, and Erfani (2021)	Real-world	3	Helpdesk, BPIC2012, BPIC2017	Paired t-test		Predictive business process monitoring	Event log	time	one-hot	MAE, Average SDL	split
Teinemaa et al. (2016)	Real-world	2	debt recovery process, the lead-to-contract			Predictive business process monitoring	Event log	resource, data, text	Index-based encoding + text encoding (Bag-of-n-grams model with and without Naive Bayes count ratios, Latent dichlet allocation topic modelling, doc2vec)	F-Score, Precision, Recall, Earliness, Efficiency	cross

Paper	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	Feature creation	Feature selection	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other
Senderovich et al. (2017)	x						x	x		x					
Senderovich et al. (2015)							x		x	x					
Senderovich et al. (2014)						x				x					
Shraga et al. (2020)	x		x					x						x	
Silva et al. (2005)			x											x	
Sim et al. (2022)		x								x					
Sindhgatta et al. (2020)	x						x			x					
Sommers et al. (2021)	x		x							x					
Stierle, Weinzierl, et al. (2021)	x			x						x					
Tavares et al. (2019)	x						x							x	
Tax et al. (2016)	x	x						x		x					
Tax et al. (2017)	x						x	x		x					
Taymouri et al. (2020)	x						x				x				
Taymouri, La Rosa, and Erfani (2021)	x						x	x			x				
Teinemaa et al. (2016)	x						x		x	x			x		

Paper	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf./learn.-based	Genetic-based	Other
Senderovich et al. (2017)					x							x			x					
Senderovich et al. (2015)										x		x			x					
Senderovich et al. (2014)					x							x			x					
Shraga et al. (2020)										x			x							
Silva et al. (2005)										x						x				
Sim et al. (2022)						x								x						
Sindhgatta et al. (2020)	x					x								x						
Sommers et al. (2021)						x								x						
Stierle, Weinzierl, et al. (2021)						x								x						
Tavares et al. (2019)			x										x							
Tax et al. (2016)										x						x				
Tax et al. (2017)						x								x						
Taymouri et al. (2020)	x					x								x						
Taymouri, La Rosa, and Erfani (2021)						x								x						
Teinemaa et al. (2016)					x							x		x	x	x				

(*continued on next page*)

Table C.7 (*continued*).

Paper	Data type	Number of datasets	Name of datasets	Statistical method	Further	BPM task category	Dataset type	Rel. context information	Feature extraction	Metrics	Validation strategy
Teinemaa et al. (2018)	Real-world	4	BPIC2017, RTFM, Unemployment			Prescriptive business process monitoring	Event log		aggregation encoding	Cost	cross
Tello et al. (2019)	Real-world	1	SAP			Event log creation	Event log	time, resource	event sequences	Accuracy, ROC, Testing time	split, cross
Theis and Darabi (2019)	Real-world	3	BPIC2012, BPIC2013, Helpdesk			Predictive business process monitoring	Event log		discretisation, normalisation	Accuracy, Precision, Recall, F-Score, Multiclass AUC	split
Theis and Darabi (2020)	Synthetic	1	Own	T-test, Wilcoxon signed-rank test		Performance analysis	Other			Generalisation score, Number of generated variants, TP rates, Fitness, Precision	split
Turner et al. (2008)	Synthetic	3	NA			Discovery of procedural process model	Event log		graph-based		
Valdés et al. (2022a)	Synthetic	1	Own	Gamma test		Pattern detection	Time series data		Trace-fitness-features	Precision, Recall, ROC	cross
Valdés et al. (2022b)	Synthetic	1	Own			Pattern detection	Time series data		DFG-based features	Precision, Recall	cross
van der Aalst et al. (2005)	Real-world, Synthetic	10	Own			Discovery of procedural process model	Event log			Ratio how often correct process model was found	
Varela-Vaca et al. (2019)	Real-world	1	ERP feature model			Trace clustering	Event log	data		Density, Cyclomatic number, Coefficient of connectivity, Control Flow complexity	
Vázquez-Barreiros et al. (2015)	Real-world, Synthetic	111	Own	Friedman test, Holm test		Discovery of procedural process model	Event log			Behavioural precision, Behavioural recall, Structural precision, Structural recall, Completeness, Precision and simplicity	
Vázquez-Barreiros et al. (2014)	NA	21	NA			Discovery of procedural process model	Event log			Behavioural precision, Behavioural recall, Structural precision, Structural recall, Proper completion, Alignment precision, Simplicity	
Verenich et al. (2016)	Real-world	2	Bondora, Environmental permit			Predictive business process monitoring	Event log	time, resource, data	index-based encoding	Accuracy, TP, FP, AUC, ROC	split, cross
Vergidis et al. (2007)	Real-world	5	Travel agent process			(Semi-)automatic process redesign	Business process designs (activities, connections, routing decisions)			Cost, Duration, Success ratio	
Wahid et al. (2019)	Real-world, Synthetic	2	Port logistics process, synthetic log			Predictive business process monitoring	Event log	Time	entity embedding, continuous variables	RMSE, MAE, Accuracy	split
Wang et al. (2012)	Real-world, Synthetic	3	Artificial, boiler, trans			Discovery of procedural process model	Event log, set of process models		principle components as features	Training time, Accuracy	split
Wang et al. (2019)	Real-world	6	BPIC2012, BPIC2017, Sepsis, Production, RTFM, Hospital Billing			Predictive business process monitoring	Event log		one-hot, normalisation	Accuracy, Execution time, Earliness, AUC	split
Weinzierl, Dunzer, et al. (2020)	Real-world	2	Helpdesk, BPIC2019			Prescriptive business process monitoring	Event log	time	one-hot	Damerau–Levenshtein Distance, % in Time	split

Paper	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	Feature creation	Feature selection	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other
Teinemaa et al. (2018)	x						x			x					
Tello et al. (2019)		x								x				x	
Theis and Darabi (2019)	x						x			x					
Theis and Darabi (2020)	x			x							x				
Turner et al. (2008)			x							x					
Valdés et al. (2022a)				x						x					
Valdés et al. (2022b)				x						x					
van der Aalst et al. (2005)			x							x					
Varela-Vaca et al. (2019)	x			x										x	
Vázquez-Barreiros et al. (2015)	x		x							x					
Vázquez-Barreiros et al. (2014)	x		x							x					
Verenich et al. (2016)	x						x			x					
Vergidis et al. (2007)					x					x					
Wahid et al. (2019)							x	x		x					
Wang et al. (2012)	x		x					x	x	x					
Wang et al. (2019)							x			x					
Weinzierl, Dunzer, et al. (2020)	x						x	x		x					

Paper	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf./learn.-based	Genetic-based	Other
Teinemaa et al. (2018)					x							x								
Tello et al. (2019)					x						x		x	x		x				
Theis and Darabi (2019)						x								x						
Theis and Darabi (2020)						x								x						
Turner et al. (2008)										x									x	
Valdés et al. (2022a)										x	x	x		x		x	x			
Valdés et al. (2022b)										x	x									
van der Aalst et al. (2005)										x									x	
Varela-Vaca et al. (2019)										x			x							
Vázquez-Barreiros et al. (2015)										x									x	
Vázquez-Barreiros et al. (2014)										x									x	
Verenich et al. (2016)										x	x				x					
Vergidis et al. (2007)										x									x	
Wahid et al. (2019)						x								x						
Wang et al. (2012)										x	x				x					
Wang et al. (2019)						x								x						
Weinzierl, Dunzer, et al. (2020)	x					x					x			x						

(*continued on next page*)

Table C.7 (*continued*).

Paper	Evaluation					BPM lifecycle phase							BPM task category	Data input		Feature engineering		
	Data type	Number of datasets	Name of datasets	Statistical method	Further	Implementation available	Process identification	Process discovery	Process analysis	Process redesign	Process implementation	Process monitoring	BPM task category	Dataset type	Rel. context information	Feature extraction	Feature creation	Feature selection
Weinzierl et al. (2022)	modified Real-world	4	BPIC2012, BPIC2013, BPIC2019, BPIC2020		Expert interviews	x			x				Pattern detection	Event log	re-source, data	binary, index-based		
Wetzstein et al. (2009)	NA	NA	Own		Simulation				x				Performance analysis	NA				
Weytjens and De Weerdt (2021)	Real-world	3	BPIC2017, BPIC2019, BPIC2020			x						x	Predictive business process monitoring	Event log	time, data	categorical (integer, embedding), numerical (standardised)	x	
Wibisono et al. (2015)	Real-world	1	Driver license application									x	Dynamic resource allocation	NA				
Wickramanayake et al. (2022)	Real-world	2	BPIC2012, BPIC2017		Expert review to check in-terpretability of the model(s)	x						x	Predictive business process monitoring	Event log		one-hot		
Xu et al. (2016)	Real-world	1	Process management system								x		Resource allocation planning	Event log	time, re-source			
Xu and Liu (2019)	Real-world	3	BPIC2011, BPIC2018, Hospital Billing		Case study		x						Data improvement	Event log		frequency-based		
Yang et al. (2018)	Real-world, Synthetic	2	WABO, BPIC2013					x					Organisational model discovery	Event log	re-source	performer by activity matrix		
Yang et al. (2017)	Real-world	3	Tracheal Intubation Data, Trauma Resuscitation Data, Emergency Department Data	Wald test					x				Trace clustering	Event log	data	activity occurrence per time unit		x
Yeshchenko et al. (2019)	Real-world, Synthetic	3	BPIC2011, Helpdesk, synthetic log			x			x				Pattern detection	Event log				
Yeshchenko et al. (2018)	Real-world	4	BPIC2012, BPIC2017, BPIC2013, Road traffic			x						x	Predictive business process monitoring	Event log, news	time, re-source, senti-ment	index-based, last-state, day/before/window for sentiments	x	
Zhao et al. (2016)	Real-world	1	E-healthcare process in a Chinese hospital									x	Dynamic resource allocation	Event log	re-source			
Zhao et al. (2020)	Real-world	1	Resource management system									x	Dynamic resource allocation	Event log, resource data	time, re-source, data	key characteristic and team faultiness features	x	x

Paper	Model building: ML paradigm						ML concept										ML algorithm group										Model assessment	
	Supervised learning	Semi-supervised learning	Reinforcement learning	Self-supervised learning	Unsupervised learning	Other	Multi-task learning	Active learning	Online/increm. learning	Transfer learning	Ensemble learning	Deep learning	Meta learning	Federated learning	Multi-view learning	Other	Instance-based	Decision tree-based	Clustering-based	Artificial-NN-based	Regression-based	Bayesian-based	Rule-based	Reinf.-learn.-based	Genetic-based	Other	Metrics	Validation strategy
Weinzierl et al. (2022)	x											x								x							Accuracy, Precision, Recall, F1-Score	split
Wetzstein et al. (2009)	x															x							x				Accuracy	
Weytjens and De Weerdt (2021)	x											x								x		x					MAE, Normalised MAE, Dropout inclusion, Byes inclusion	split
Wibisono et al. (2015)	x															x						x					Average completion time (mean, sd, single values)	
Wickramanayake et al. (2022)	x											x								x							Accuracy, Precision, Recall, F1-Score	split
Xu et al. (2016)	x															x									x		No. of instances scheduled, Ratio of scheduled instances by adjustment, Processing time	
Xu and Liu (2019)					x											x	x		x	x							Success rate	
Yang et al. (2018)					x											x			x			x					Bcubed precision, Bcubed recall, Bcubed f-measure, Average group size, Average number of membership	
Yang et al. (2017)	x				x											x			x		x						(Novel trace) similarity metric, F measure, G-mean	cross
Yeshchenko et al. (2019)					x				x										x								F-Score, Erratic measure	
Yeshchenko et al. (2018)	x										x							x									MAE	split
Zhao et al. (2016)					x						x								x								MAE, Recommendation Accuracy, Clustering time, Prediction time (compare these metrics with kmeans)	split, cross
Zhao et al. (2020)	x				x						x								x	x							MSE, Information value	split

References

Afflerbach, P., Hohendorf, M., & Manderscheid, J. (2017). Design it like Darwin – A value-based application of evolutionary algorithms for proper and unambiguous business process redesign. *Information Systems Frontiers*, *19*(5), 1101–1121.

Alpaydin, E. (2014). *Introduction to machine learning* (3rd ed.). MIT Press.

Alves, R. M., Barbieri, L., Stroeh, K., Peres, S. M., & Madeira, E. R. M. (2022). Context-aware completion time prediction for business process monitoring. In *Proceedings of the 2022 World Conference on Information Systems and Technologies* (pp. 355–365). Springer.

Appice, A., Di Mauro, N., & Malerba, D. (2019). Leveraging shallow machine learning to predict business process behavior. In *Proceedings of the 2019 IEEE International Conference on Services Computing* (pp. 184–188). IEEE.

Appice, A., & Malerba, D. (2015). A co-training strategy for multiple view clustering in process mining. *IEEE Transactions on Services Computing*, *9*(6), 832–845.

Bae, H., Lee, S., & Moon, I. (2014). Planning of business process execution in business process management environments. *Information Sciences*, *268*, 357–369.

Baier, L., Reimold, J., & Kühl, N. (2020). Handling concept drift for predictions in business process mining. In *Proceedings of the 22nd IEEE Conference on Business Informatics* (pp. 76–83). IEEE.

Bauer, K., Hinz, O., van der Aalst, W., & Weinhardt, C. (2021). Expl(AI)n it to me – Explainable AI and information systems research. *Business & Information Systems Engineering*, *63*(2), 79–82.

Bazhenova, E., Buelow, S., & Weske, M. (2016). Discovering decision models from event logs. In *Proceedings of the 19th International Conference on Business information systems* (pp. 237–251). Springer.

Beese, J., Haki, M. K., Aier, S., & Winter, R. (2019). Simulation-based research in information systems. *Business & Information Systems Engineering*, *61*(4), 503–521.

Berkenstadt, G., Gal, A., Senderovich, A., Shraga, R., & Weidlich, M. (2020). Queueing inference for process performance analysis with missing life-cycle data. In *Proceedings of the 2nd International Conference on Process Mining* (pp. 57–64). IEEE, http://dx.doi.org/10.1109/ICPM49681.2020.00019.

Bernard, G., & Andritsos, P. (2019). Accurate and transparent path prediction using process mining. In *Proceedings of the 23rd European Conference on Advances in Databases and Information Systems* (pp. 235–250). Springer.

Bevacqua, A., Carnuccio, M., Folino, F., Guarascio, M., & Pontieri, L. (2013). A data-driven prediction framework for analyzing and monitoring business process performances. In *Proceedings of the 15th International Conference on Enterprise Information Systems* (pp. 100–117). Springer.

Beverungen, D., Buijs, J., Becker, J., Di Ciccio, C., van der Aalst, W. M. P., Bartelheimer, C., et al. (2021). Seven paradoxes of business process management in a hyper-connected world. *Business & Information Systems Engineering*, *63*(2), 145–156. http://dx.doi.org/10.1007/s12599-020-00646-z.

Bezerra, F., Wainer, J., & van der Aalst, W. M. P. (2009). Anomaly detection using process mining. In *Proceedings of the 14th International Conference on Enterprise, Business Process and Information Systems Modeling* (pp. 149–161). Springer.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Blum, A. (1998). On-line algorithms in machine learning. In *Online algorithms – State of the art* (pp. 306–325). Springer.

Boltenhagen, M., Chatain, T., & Carmona, J. (2019). Generalized alignment-based trace clustering of process behavior. In *Proceedings of the 40th International Conference on Application and Theory of Petri Nets and Concurrency* (pp. 237–257). Springer.

Borkowski, M., Fdhila, W., Nardelli, M., Rinderle-Ma, S., & Schulte, S. (2019). Event-based failure prediction in distributed business processes. *Information Systems*, *81*, 220–235.

Bose, R. J. C., Mans, R. S., & van der Aalst, W. M. P. (2013). Wanna improve process mining results? In *Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining* (pp. 127–134). IEEE.

Bose, R. J. C., & van der Aalst, W. M. P. (2009). Context aware trace clustering: Towards improving process mining results. In *Proceedings of the 2009 SIAM International Conference on Data Mining* (pp. 401–412). SIAM.

Bozorgi, Z. D., Teinemaa, I., Dumas, M., La Rosa, M., & Polyvyanyy, A. (2020). Process mining meets causal machine learning: Discovering causal rules from event logs. In *Proceedings of the 2nd International Conference on Process Mining* (pp. 129–136). IEEE.

Bozorgi, Z. D., Teinemaa, I., Dumas, M., La Rosa, M., & Polyvyanyy, A. (2021). Prescriptive process monitoring for cost-aware cycle time reduction. In *Proceedings of the 3rd International Conference on Process Mining* (pp. 96–103). IEEE.

Breiman, L., Friedman, J. H., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth.

Breuker, D., Matzner, M., Delfmann, P., & Becker, J. (2016). Comprehensible predictive models for business processes. *MIS Quarterly*, *40*(4), 1009–1034. http://dx.doi.org/10.25300/MISQ/2016/40.4.10.

Brunk, J., Stierle, M., Papke, L., Revoredo, K., Matzner, M., & Becker, J. (2021). Cause vs. effect in context-sensitive prediction of business process instances. *Information Systems*, *95*, Article 101635.

Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence. URL https://hbr.org/2017/07/the-business-of-artificial-intelligence.

Buijs, J., van Dongen, B. F., & van der Aalst, W. M. P. (2012). A genetic algorithm for discovering process trees. In *Proceedings of the 2012 IEEE Congress on Evolutionary Computation* (pp. 1–8). IEEE.

Buijs, J., van Dongen, B. F., & van der Aalst, W. M. P. (2013). Mining configurable process models from collections of event logs. In *Proceedings of the 11th International Conference on Business Process Management* (pp. 33–48). Springer.

Camargo, M., Dumas, M., & González-Rojas, O. (2019). Learning accurate LSTM models of business processes. In *Proceedings of the 17th International Conference on Business Process management* (pp. 286–302). Springer.

Camargo, M., Dumas, M., & González-Rojas, O. (2022). Learning accurate business process simulation models from event logs via automated process discovery and deep learning. In *Proceedings of the 34th International Conference on Advanced Information Systems Engineering* (pp. 55–71). Springer.

Cao, F., Estert, M., Qian, W., & Zhou, A. (2006). Density-based clustering over an evolving data stream with noise. In *Proceedings of the 2006 SIAM International Conference on Data Mining* (pp. 328–339). SIAM.

Ceravolo, P., Damiani, E., Schepis, E. F., & Tavares, G. M. (2022). Real-time probing of control-flow and data-flow in event logs. *Procedia Computer Science*, *197*, 751–758.

Chesani, F., Lamma, E., Mello, P., Montali, M., Riguzzi, F., & Storari, S. (2009). Exploiting inductive logic programming techniques for declarative process mining. *Transactions on Petri Nets and Other Models of Concurrency II: Special Issue on Concurrency in Process-Aware Information Systems*, 278–295.

Ciasullo, M. V., Fenza, G., Loia, V., Orciuoli, F., Troisi, O., & Herrera-Viedma, E. (2018). Business process outsourcing enhanced by fuzzy linguistic consensus model. *Applied Soft Computing*, *64*, 436–444.

Cohen, W. W. (1995). Fast effective rule induction. In *Proceedings of the 12th International Conference on Machine Learning* (pp. 115–123). Elsevier, http://dx.doi.org/10.1016/B978-1-55860-377-6.50023-2.

Conforti, R., de Leoni, M., La Rosa, M., van der Aalst, W. M. P., & ter Hofstede, A. H. (2015). A recommendation system for predicting risks across multiple business process instances. *Decision Support Systems*, *69*, 1–19.

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, *1*(1), 104–126.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. http://dx.doi.org/10.1007/BF00994018.

Cuzzocrea, A., Folino, F., Guarascio, M., & Pontieri, L. (2015). A multi-view learning approach to the discovery of deviant process instances. In *Proceedings of the 2015 OTM Confederated International Conferences "On The Move to Meaningful Internet Systems"* (pp. 146–165). Springer.

Cuzzocrea, A., Folino, F., Guarascio, M., & Pontieri, L. (2016a). A multi-view multi-dimensional ensemble learning approach to mining business process deviances. In *Proceedings of the 2016 International Joint Conference on Neural Networks* (pp. 3809–3816). IEEE.

Cuzzocrea, A., Folino, F., Guarascio, M., & Pontieri, L. (2016b). A robust and versatile multi-view learning framework for the detection of deviant business process instances. *International Journal of Cooperative Information Systems*, *25*(4), Article 1740003.

Cuzzocrea, A., Folino, F., Guarascio, M., & Pontieri, L. (2019). Predictive monitoring of temporally-aggregated performance indicators of business processes against low-level streaming events. *Information Systems*, *81*, 236–266.

De Jong, K. A., Spears, W. M., & Gordon, D. F. (1993). Using genetic algorithms for concept learning. *Machine Learning*, *13*(2), 161–188.

De Koninck, P., & De Weerdt, J. (2019). Scalable mixed-paradigm trace clustering using super-instances. In *Proceedings of the 1st International Conference on Process Mining* (pp. 17–24). IEEE.

De Koninck, P., De Weerdt, J., & vanden Broucke, S. (2017). Explaining clusterings of process instances. *Data Mining and Knowledge Discovery*, *31*(3), 774–808.

De Koninck, P., Nelissen, K., Baesens, B., Snoeck, M., & De Weerdt, J. (2017). An approach for incorporating expert knowledge in trace clustering. In *Proceedings of the 29th International Conference on Advanced Information Systems Engineering* (pp. 561–576). Springer.

De Koninck, P., Nelissen, K., Baesens, B., Snoeck, M., & De Weerdt, J. (2021). Expert-driven trace clustering with instance-level constraints. *Knowledge and Information Systems*, *63*(5), 1197–1220.

De Koninck, P., vanden Broucke, S., & De Weerdt, J. (2018). Act2vec, trace2vec, log2vec, and model2vec: Representation learning for business processes. In *Proceedings of the 16th International Conference on Business Process Management* (pp. 305–321). Springer.

de Leoni, M., Dees, M., & Reulink, L. (2020). Design and evaluation of a process-aware recommender system based on prescriptive analytics. In *Proceedings of the 2nd International Conference on Process Mining* (pp. 9–16). IEEE.

de Leoni, M., Dumas, M., & García-Bañuelos, L. (2013). Discovering branching conditions from business process execution logs. In *Proceedings of the 16th International Conference on Fundamental Approaches to Software Engineering* (pp. 114–129). Springer.

de Leoni, M., van der Aalst, W. M. P., & Dees, M. (2014). A general framework for correlating business process characteristics. In *Proceedings of the 12th International Conference on Business Process Management* (pp. 250–266). Springer.

de Leoni, M., van der Aalst, W. M. P., & Dees, M. (2016). A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. *Information Systems*, *56*, 235–257.

De Maio, C., Fenza, G., Loia, V., Orciuoli, F., & Herrera-Viedma, E. (2016). A framework for context-aware heterogeneous group decision making in business processes. *Knowledge-Based Systems*, *102*, 39–50.

De Morais, R. M., & Kazan, S. (2014). An analysis of BPM lifecycles: From a literature review to a framework proposal. *Business Process Management Journal*, *20*, 1463–7154. http://dx.doi.org/10.1108/BPMJ-03-2013-0035.

De Weerdt, J., vanden Broucke, S., Vanthienen, J., & Baesens, B. (2012). Leveraging process discovery with trace clustering and text mining for intelligent analysis of incident management processes. In *Proceedings of the 2012 IEEE Congress on Evolutionary Computation* (pp. 1–8). IEEE.

De Weerdt, J., vanden Broucke, S., Vanthienen, J., & Baesens, B. (2013). Active trace clustering for improved process discovery. *IEEE Transactions on Knowledge and Data Engineering*, *25*(12), 2708–2720.

Delcoucq, L., Dupiereux-Fettweis, T., Lecron, F., & Fortemps, P. (2022). Resource and activity clustering based on a hierarchical cell formation algorithm. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, 1–10.

Delias, P., Doumpos, M., Grigoroudis, E., & Matsatsinis, N. (2019). A non-compensatory approach for trace clustering. *International Transactions in Operational Research*, *26*(5), 1828–1846.

Di Francescomarino, C., Dumas, M., Maggi, F. M., & Teinemaa, I. (2016). Clustering-based predictive process monitoring. *IEEE Transactions on Services Computing*, *12*(6), 896–909.

Di Francescomarino, C., Ghidini, C., Maggi, F. M., & Milani, F. (2018). Predictive process monitoring methods: Which one suits me best? In *Proceedings of 16th International Conference on Business Process Management* (pp. 462–479). Springer.

Di Francescomarino, C., Ghidini, C., Maggi, F. M., Petrucci, G., & Yeshchenko, A. (2017). An eye into the future: Leveraging a-priori knowledge in predictive business process monitoring. In *Proceedings of the 15th International Conference on Business Process Management* (pp. 252–268). Springer.

Di Mauro, N., Appice, A., & Basile, T. (2019). Activity prediction of business process instances with inception cnn models. In *Proceedings of the 18th International Conference of the Italian Association for Artificial Intelligence* (pp. 348–361). Springer.

Diamantini, C., Genga, L., & Potena, D. (2016). Behavioral process mining for unstructured processes. *Journal of Intelligent Information Systems*, *47*(1), 5–32.

Dietterich, T. G. (2002). Ensemble learning. In *The handbook of brain theory and neural networks* (2nd ed.). (pp. 110–125). MIT Press.

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Fundamentals of business process management, (2nd ed.). Springer, http://dx.doi.org/10.1007/978-3-642-33143-5.

Effendi, Y. A., & Sarno, R. (2017). Discovering process model from event logs by considering overlapping rules. In *Proceedings of the 4th International Conference on Electrical Engineering, Computer Science and Informatics* (pp. 1–6). IEEE.

Es-Soufi, W., Yahia, E., & Roucoules, L. (2016). On the use of process mining and machine learning to support decision making in systems design. In *Proceedings of the 13th IFIP International Conference on Product Lifecycle Management* (pp. 56–66). Springer, http://dx.doi.org/10.1007/978-3-319-54660-5_6.

Evermann, J., Rehse, J.-R., & Fettke, P. (2017). Predicting process behaviour using deep learning. *Decision Support Systems*, *100*, 129–140. http://dx.doi.org/10.1016/j.dss.2017.04.003.

Ferreira, D. R., & Gillblad, D. (2009). Discovering process models from unlabelled event logs. In *Proceedings of the 7th International Conference on Business process Management* (pp. 143–158). Springer.

Ferreira, D. R., Szimanski, F., & Ralha, C. G. (2013). Mining the low-level behaviour of agents in high-level business processes. *International Journal of Business Process Integration and Management*, *6*(2), 146–166.

Ferreira, D. R., & Vasilyev, E. (2015). Using logical decision trees to discover the cause of process delays from event logs. *Computers in Industry*, *70*, 194–207.

Ferreira, D. R., Zacarias, M., Malheiros, M., & Ferreira, P. (2007). Approaching process mining with sequence clustering: Experiments and findings. In *Proceedings of the 5th International Conference on Business Process Management* (pp. 360–374). Springer.

Ferri, C., Hernández-Orallo, J., & Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, *30*(1), 27–38.

Feuerriegel, S., Dolata, M., & Schwabe, G. (2020). Fair AI. *Business & Information Systems Engineering*, *62*(4), 379–384.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2023). Generative AI. *Business & Information Systems Engineering*, in press.

Firouzian, I., Zahedi, M., & Hassanpour, H. (2019a). Cycle time optimization of processes using an entropy-based learning for task allocation. *International Journal of Engineering*, *32*(8), 1090–1100.

Firouzian, I., Zahedi, M., & Hassanpour, H. (2019b). Investigation of the effect of concept drift on data-aware remaining time prediction of business processes. *International Journal of Nonlinear Analysis and Applications*, *10*(2), 153–166.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*(3), 613–619.

Folino, F., Folino, G., Guarascio, M., & Pontieri, L. (2022). Semi-supervised discovery of DNN-based outcome predictors from scarcely-labeled process logs. *Business & Information Systems Engineering*, 1–21.

Folino, F., Greco, G., Guzzo, A., & Pontieri, L. (2011). Mining usage scenarios in business processes: Outlier-aware discovery and run-time prediction. *Data & Knowledge Engineering*, *70*(12), 1005–1029.

Folino, F., Guarascio, M., & Pontieri, L. (2012). Discovering context-aware models for predicting business process performances. In *Proceedings of the 2012 OTM Confederated International Conferences "On The Move to Meaningful Internet Systems"* (pp. 287–304). Springer.

Folino, F., Guarascio, M., & Pontieri, L. (2015). Mining multi-variant process models from low-level logs. In *Proceedings of the 18th International Conference on Business Information Systems* (pp. 165–177). Springer.

Folino, F., & Pontieri, L. (2021). AI-empowered process mining for complex application scenarios: Survey and discussion. *Journal on Data Semantics*, *10*, 77–106. http://dx.doi.org/10.1007/s13740-021-00121-2.

Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics*, *21*(3), 768–769.

Fürnkranz, J., Gamberger, D., & Lavrač, N. (2012). *Foundations of rule learning*. Springer.

Galanti, R., Coma-Puig, B., de Leoni, M., Carmona, J., & Navarin, N. (2020). Explainable predictive process monitoring. In *Proceedings of the 2nd International Conference on Process Mining* (pp. 1–8). IEEE.

Garcez, A. d., Bader, S., Bowman, H., Lamb, L. C., de Penning, L., Illuminoo, B., et al. (2022). Neural-symbolic learning and reasoning: A survey and interpretation. In *Neuro-symbolic artificial intelligence: The state of the art* (pp. 1–51). IOS Press.

Garcez, A. d., Besold, T. R., De Raedt, L., Földiak, P., Hitzler, P., Icard, T., et al. (2015). Neural-symbolic learning and reasoning: Contributions and challenges. In *Proceedings of the 2015 AAAI Spring Symposium – Knowledge representation and reasoning: Integrating symbolic and neural approaches* (pp. 18–21). AI Access Foundation.

García-Bañuelos, L., Dumas, M., La Rosa, M., De Weerdt, J., & Ekanayake, C. C. (2014). Controlled automated discovery of collections of business process models. *Information Systems*, *46*, 85–101.

Gerlach, Y., Seeliger, A., Nolle, T., & Mühlhäuser, M. (2022). Inferring a multi-perspective likelihood graph from black-box next event predictors. In *Proceedings of the 34th International Conference on Advanced Information Systems Engineering* (pp. 19–35). Springer.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Greco, G., Guzzo, A., Pontieri, L., & Sacca, D. (2004). Mining expressive process models by clustering workflow traces. In *Proceedings of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 52–62). Springer.

Greco, G., Guzzo, A., Pontieri, L., & Sacca, D. (2006). Discovering expressive process models by clustering log traces. *IEEE Transactions on Knowledge and Data Engineering*, *18*(8), 1010–1027.

Grigori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M., & Shan, M.-C. (2004). Business process intelligence. *Computers in Industry*, *53*(3), 321–343. http://dx.doi.org/10.1016/j.compind.2003.10.007.

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., et al. (2023). How close is chatgpt to human experts? comparison corpus, evaluation, and detection. arXiv preprint arXiv:2301.07597, arXiv:2301.07597.

Guzzo, A., Joaristi, M., Rullo, A., & Serra, E. (2021). A multi-perspective approach for the analysis of complex business processes behavior. *Expert Systems with Applications*, *177*, Article 114934.

Ha, Q.-T., Bui, H.-N., & Nguyen, T.-T. (2016). A trace clustering solution based on using the distance graph model. In *Proceedings of the 8th International Conference on Computational Collective Intelligence* (pp. 313–322). Springer.

Hanga, K. M., Kovalchuk, Y., & Gaber, M. M. (2020). A graph-based approach to interpreting recurrent neural networks in process mining. *IEEE Access*, *8*, 172923–172938.

Harl, M., Weinzierl, S., Stierle, M., & Matzner, M. (2020). Explainable predictive business process monitoring using gated graph neural networks. *Journal of Decision Systems*, *29*(sup1), 312–327.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction: vol. 2*, Springer.

Hayes-Roth, F. (1985). Rule-based systems. *Communications of the ACM*, *28*(9), 921–932.

Heinrich, K., Zschech, P., Janiesch, C., & Bonin, M. (2021). Process data properties matter: Introducing gated convolutional neural networks (GCNN) and key-value-predict attention networks (KVP) for next event prediction with deep learning. *Decision Support Systems*, *143*, Article 113494. http://dx.doi.org/10.1016/j.dss.2021.113494.

Herbst, J. (2000). A machine learning approach to workflow management. In *Proceedings of the 11th European Conference on Machine Learning* (pp. 183–194). Springer.

Herbst, J., & Karagiannis, D. (2000). Integrating machine learning and workflow management to support acquisition and adaptation of workflow models. *Intelligent Systems in Accounting, Finance & Management*, *9*(2), 67–92.

Herbst, J., & Karagiannis, D. (2004). Workflow mining with InWoLvE. *Computers in Industry*, *53*(3), 245–264.

Herm, L.-V., Janiesch, C., Reijers, H. A., & Seubert, F. (2021). From symbolic RPA to intelligent RPA: Challenges for developing and operating intelligent software robots. In *Proceedings of the 19th International Conference on Business Process Management* (pp. 289–305). Springer.

Houy, C., Fettke, P., & Loos, P. (2010). Empirical research in business process management — Analysis of an emerging field of research. *Business Process Management Journal*, *16*, 619–661. http://dx.doi.org/10.1108/14637151011065946.

Hruschka, D. J., Schwartz, D., St. John, D. C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*, *16*(3), 307–331.

Hsieh, C., Moreira, C., & Ouyang, C. (2021). DICE4EL: Interpreting process predictions using a milestone-aware counterfactual approach. In *Proceedings of the 3rd International Conference on Process Mining* (pp. 88–95). IEEE.
Huang, Z., van der Aalst, W. M. P., Lu, X., & Duan, H. (2010). An adaptive work distribution mechanism based on reinforcement learning. *Expert Systems with Applications, 37*(12), 7533–7541.
Huang, Z., van der Aalst, W. M. P., Lu, X., & Duan, H. (2011). Reinforcement learning based resource allocation in business process management. *Data & Knowledge Engineering, 70*(1), 127–145.
Huo, S., Völzer, H., Reddy, P., Agarwal, P., Isahagian, V., & Muthusamy, V. (2021). Graph autoencoders for business process anomaly detection. In *Proceedings of the 19th International Conference on Business Process Management* (pp. 417–433). Springer.
Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., & Makedon, F. (2021). A survey on contrastive self-supervised learning. *Technologies, 9*(2), 1–22. http://dx.doi.org/10.3390/technologies9010002.
Jalayer, A., Kahani, M., Beheshti, A., Pourmasoumi, A., & Motahari-Nezhad, H. R. (2020). Attention mechanism in predictive business process monitoring. In *Proceedings of the 24th IEEE international Enterprise Distributed Object Computing Conference* (pp. 181–186). IEEE.
Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets, 31*, 685–695. http://dx.doi.org/10.1007/s12525-021-00475-2.
Jing, L., & Tian, Y. (2021). Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(11), 4037–4058. http://dx.doi.org/10.1109/TPAMI.2020.2992393.
Jlailaty, D., Grigori, D., & Belhajjame, K. (2017). Business process instances discovery from email logs. In *Proceeding of the 14th International Conference on Services Computing* (pp. 19–26). IEEE.
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.
Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260.
Jung, J.-Y., Bae, J., & Liu, L. (2008). Hierarchical business process clustering. In *Proceedings of the 2008 IEEE International Conference on Services Computing* (pp. 613–616). IEEE.
Jung, J.-Y., Bae, J., & Liu, L. (2009). Hierarchical clustering of business process models. *International Journal of Innovative Computing, Information and Control, 5*(12), 1349–4198.
Junior, S. B., Ceravolo, P., Damiani, E., Omori, N. J., & Tavares, G. M. (2020). Anomaly detection on event logs with a scarcity of labels. In *Proceedings of the 2nd International Conference on Process Mining* (pp. 161–168). IEEE.
Junior, S. B., Tavares, G. M., da Costa, V. G. T., Ceravolo, P., & Damiani, E. (2018). A framework for human-in-the-loop monitoring of concept-drift detection in event log stream. In *Proceedings of the Web Conference 2018* (pp. 319–326).
Kaddour, J., Lynch, A., Liu, Q., Kusner, M. J., & Silva, R. (2022). Causal machine learning: A survey and open problems. (pp. 1–188). arXiv preprint arXiv:2206.15475.
Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research, 4*(19), 237–285. http://dx.doi.org/10.1613/jair.301.
Kang, B., Kim, D., & Kang, S.-H. (2012). Periodic performance prediction for real-time business process monitoring. *Industrial Management & Data Systems, 112*(1), 4–23.
Kazakov, O. D., Novikov, S. P., Kulagina, N. A., & Shlapakova, S. N. (2018). Development of the concept of management of economic systems processes through construction and calling of machine learning models. In *Proceedings of the 2018 IEEE International Conference "Quality Management, Transport and Information Security, Information Technologies"* (pp. 316–321). IEEE.
Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns, 2*(9), Article 100314.
Khan, A., Ghose, A., & Dam, H. (2021). Decision support for knowledge intensive processes using RL based recommendations. In *Proceedings of the 19th International Conference of Business Process Management (Forum)* (pp. 246–262). Springer.
Khodyrev, I., & Popova, S. (2014). Discrete modeling and simulation of business processes using event logs. *Procedia Computer Science, 29*, 322–331.
Kim, J., Suh, W., & Lee, H. (2002). Document-based workflow modeling: A case-based reasoning approach. *Expert Systems with Applications, 23*(2), 77–93.
Klijn, E. L., & Fahland, D. (2020). Identifying and reducing errors in remaining time prediction due to inter-case dynamics. In *Proceedings of the 2nd International Conference on Process Mining* (pp. 25–32). IEEE.
Ko, J., & Comuzzi, M. (2023). A systematic review of anomaly detection for business process event logs. *Business & Information Systems Engineering*.
Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review, 26*(3), 159–190. http://dx.doi.org/10.1007/s10462-007-9052-3.
Krajsic, P., & Franczyk, B. (2020). Lambda architecture for anomaly detection in online process mining using autoencoders. In *Proceeding of the 12th International Conference on Computational Collective Intelligence* (pp. 579–589). Springer.
Krajsic, P., & Franczyk, B. (2021). Variational autoencoder for anomaly detection in event data in online process mining. In *Proceedings of the 23rd International Conferences on Enterprise Information Systems* (pp. 567–574). SciTePress.
Kratsch, W., König, F., & Röglinger, M. (2022). Shedding light on blind spots – Developing a reference architecture to leverage video data for process mining. *Decision Support Systems, 158*, Article 113794.
Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage Publications.
Lakshmanan, G. T., Shamsi, D., Doganata, Y. N., Unuvar, M., & Khalaf, R. (2015). A markov prediction model for data-driven semi-structured business processes. *Knowledge and Information Systems, 42*(1), 97–126.
Lamma, E., Mello, P., Montali, M., Riguzzi, F., & Storari, S. (2007). Inducing declarative logic-based models from labeled traces. In *Proceedings of the 5th International Conference on Business Process Management* (pp. 344–359). Springer.
Lamma, E., Mello, P., Riguzzi, F., & Storari, S. (2007). Applying inductive logic programming to process mining. In *Proceedings of the 17th International Conference on Inductive Logic Programming* (pp. 132–146). Springer.
Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363–374.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. http://dx.doi.org/10.1038/nature14539.
Lee, W. L. J., Burattin, A., Munoz-Gama, J., & Sepulveda, M. (2021). Orientation and conformance: A HMM-based approach to online conformance checking. *Information Systems, 102*, Article 101674. http://dx.doi.org/10.1016/j.is.2020.101674.
Lee, S., Lu, X., & Reijers, H. A. (2022). The analysis of online event streams: Predicting the next activity for anomaly detection. In *Proceeding of the 16th International Conference on Research Challenges in Information Science* (pp. 248–264). Springer.
Lee, W. L. J., Parra, D., Munoz-Gama, J., & Sepulveda, M. (2018). Predicting process behavior meets factorization machines. *Expert Systems with Applications, 112*, 87–98.
Leno, V., Dumas, M., & Maggi, F. M. (2018). Correlating activation and target conditions in data-aware declarative process discovery. In *Proceedings of the 16th International Conference on Business Process Management* (pp. 176–193). Springer.
Leno, V., Dumas, M., Maggi, F. M., La Rosa, M., & Polyvyanyy, A. (2020). Automated discovery of declarative process models with correlated data conditions. *Information Systems, 89*, Article 101482.
Leontjeva, A., Conforti, R., Di Francescomarino, C., Dumas, M., & Maggi, F. M. (2016). Complex symbolic sequence encodings for predictive monitoring of business processes. In *Proceedings of the 14th International Conference on Business Process Management* (pp. 297–313). Springer.
Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In *Proceedings of the 10th European Conference on Machine Learning* (pp. 4–15). Springer.
Li, G., de Murillas, E. G. L., de Carvalho, R. M., & van der Aalst, W. M. P. (2018). Extracting object-centric event logs to support process mining on databases. In *Proceedings of the 30th International Conference on Advanced Information Systems Engineering (Forum)* (pp. 182–199). Springer.
Li, C., Reichert, M., & Wombacher, A. (2010). The MinAdept clustering approach for discovering reference process models out of process variants. *International Journal of Cooperative Information Systems, 19*(03n04), 159–203.
Liu, T., Cheng, Y., & Ni, Z. (2012). Mining event logs to support workflow resource allocation. *Knowledge-Based Systems, 35*, 320–331.
Liu, H., & Motoda, H. (1998). *Feature extraction, construction and selection: A Data mining perspective: vol. 453*, Springer.
Liu, Y., Wang, J., Yang, Y., & Sun, J. (2008). A semi-automatic approach for workflow staff assignment. *Computers in Industry, 59*(5), 463–476.
Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2006). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587–604. http://dx.doi.org/10.1111/j.1468-2958.2002.tb00826.x, arXiv:https://academic.oup.com/hcr/article-pdf/28/4/587/22337849/jhumcom0587.pdf.
López, H. A., Strømsted, R., Niyodusenga, J.-M., & Marquard, M. (2021). Declarative process discovery: Linking process and textual views. In *Proceedings of the 33rd International Conference on Advanced Information Systems Engineering (Forum)* (pp. 109–117). Springer.
Lu, F., Zeng, Q., & Duan, H. (2016). Synchronization-core-based discovery of processes with decomposable cyclic dependencies. *ACM Transactions on Knowledge Discovery from Data, 10*(3), 1–29.
Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st Conference on Neural Information Processing Systems* (pp. 4765–4774). ACM.
Maggi, F. M., Bose, R. J., & van der Aalst, W. M. P. (2012). Efficient discovery of understandable declarative process models from event logs. In *Proceedings of the 24th International Conference on Advanced Information Systems Engineering* (pp. 270–285). Springer.
Maggi, F. M., Di Ciccio, C., Di Francescomarino, C., & Kala, T. (2018). Parallel algorithms for the automated discovery of declarative process models. *Information Systems, 74*, 136–152.
Maggi, F. M., Di Francescomarino, C., Dumas, M., & Ghidini, C. (2014). Predictive monitoring of business processes. In *Proceedings of the 26th International Conference on Advanced Information Systems Engineering* (pp. 457–472). Springer.
Maier, A. K., Syben, C., Stimpel, B., Würfl, T., Hoffmann, M., Schebesch, F., et al. (2019). Learning with known operators reduces maximum error bounds. *Nature Machine Intelligence, 1*(8), 373–380.

Maita, A. R. C., Martins, L. C., Paz, C. R. L., Peres, S. M., & Fantinato, M. (2015). Process mining through artificial neural networks and support vector machines: A systematic literature review. *Business Process Management Journal*, *21*(6), 1391–1415.

Mannhardt, F., de Leoni, M., Reijers, H. A., & van der Aalst, W. M. P. (2016). Decision mining revisited – discovering overlapping rules. In *Proceedings of the 28th International Conference on Advanced Information Systems Engineering* (pp. 377–392). Springer.

Márquez-Chamorro, A. E., Resinas, M., & Ruiz-Cortés, A. (2017). Predictive monitoring of business processes: A survey. *IEEE Transactions on Services Computing*, *11*(6), 962–977. http://dx.doi.org/10.1109/TSC.2017.2772256.

Márquez-Chamorro, A. E., Resinas, M., Ruiz-Cortés, A., & Toro, M. (2017). Run-time prediction of business process indicators using evolutionary decision rules. *Expert Systems with Applications*, *87*, 1–14.

Martens, A., Fettke, P., & Loos, P. (2014). A genetic algorithm for the inductive derivation of reference models using minimal graph-edit distance applied to real-world business process data. In *Proceedings of the 2014 Multikonferenz Wirtschaftsinformatik* (pp. 1613–1626). Universität Paderborn.

Măruşter, L., Weijters, A. J., van der Aalst, W. M. P., & Van Den Bosch, A. (2002). Process mining: Discovering direct successors in process logs. In *Proceedings of the 5th International Conference on Discovery Science* (pp. 364–373). Springer.

Măruşter, L., Weijters, A. J., van der Aalst, W. M. P., & Van Den Bosch, A. (2006). A rule-based approach for process discovery: Dealing with noise and imbalance in process logs. *Data Mining and Knowledge Discovery*, *13*(1), 67–87.

McCoy, D. W. (2002). Business activity monitoring: Calm before the storm. URL https://www.gartner.com/resources/105500/105562/105562.pdf.

Mehdiyev, N., Evermann, J., & Fettke, P. (2020). A novel business process prediction model using a deep learning method. *Business & Information Systems Engineering*, *62*(2), 143–157. http://dx.doi.org/10.1007/s12599-018-0551-3.

Mendling, J., Decker, G., Hull, R., Reijers, H. A., & Weber, I. (2018). How do machine learning, robotic process automation, and blockchains affect the human factor in business process management? *Communications of the Association for Information Systems*, *43*(1), 297–320. http://dx.doi.org/10.17705/1CAIS.04319.

Metzger, A., Kley, T., & Palm, A. (2020). Triggering proactive business process adaptations via online reinforcement learning. In *Proceedings of the 18th International Conference on Business Process Management* (pp. 273–290). Springer.

Metzger, A., Neubauer, A., Bohn, P., & Pohl, K. (2019). Proactive process adaptation using deep learning ensembles. In *Proceedings of the 31st International Conference on Advanced Information Systems Engineering* (pp. 547–562). Springer.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill Education.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT Press.

Montani, S., & Leonardi, G. (2012). Retrieval and clustering for business process monitoring: Results and improvements. In *Proceedings of the 20th International Conference on Case-based Reasoning* (pp. 269–283). Springer.

Mothilal, R. K., Sharma, A., & Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 607–617). ACM.

zur Muehlen, M., & Ho, D. T.-Y. (2005). Risk management in the BPM lifecycle. In *Proceedings of the 3rd International Conference on Business Process Management (Workshops)* (pp. 454–466). Springer, http://dx.doi.org/10.1007/11678564_42.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Mustansir, A., Shahzad, K., & Malik, M. K. (2022). Towards automatic business process redesign: An NLP based approach to extract redesign suggestions. *Automated Software Engineering*, *29*(12), 1–24. http://dx.doi.org/10.1007/s10515-021-00316-8.

Neu, D. A., Lahann, J., & Fettke, P. (2021). A systematic literature review on state-of-the-art deep learning methods for process prediction. *Artificial Intelligence Review*, *55*, 801–827. http://dx.doi.org/10.1007/s10462-021-09960-8.

Neuendorf, K. A. (2017). *The content analysis guidebook*. Sage Publications.

Nguyen, H. T. C., Lee, S., Kim, J., Ko, J., & Comuzzi, M. (2019). Autoencoders for improving quality of process event logs. *Expert Systems with Applications*, *131*, 132–147. http://dx.doi.org/10.1016/j.eswa.2019.04.052.

Nguyen, P., Slominski, A., Muthusamy, V., Ishakian, V., & Nahrstedt, K. (2016). Process trace clustering: A heterogeneous information network approach. In *Proceedings of the 2016 SIAM International Conference on Data Mining* (pp. 279–287). SIAM.

Nichol, A., Achiam, J., & Schulman, J. (2018). On first-order meta-learning algorithms. (pp. 1–15). arXiv preprint arXiv:1803.02999, arXiv:1803.02999.

Nigam, A., & Caswell, N. S. (2003). Business artifacts: An approach to operational specification. *IBM Systems Journal*, *42*(3), 428–445.

Nolle, T., Luettgen, S., Seeliger, A., & Mühlhäuser, M. (2018). Analyzing business process anomalies using autoencoders. *Machine Learning*, *107*(11), 1875–1893. http://dx.doi.org/10.1007/s10994-018-5702-8.

Nolle, T., Luettgen, S., Seeliger, A., & Mühlhäuser, M. (2019). BINet: Multi-perspective business process anomaly classification. *Information Systems*, *103*, Article 101458. http://dx.doi.org/10.1016/j.is.2019.101458.

Nolle, T., Seeliger, A., & Mühlhäuser, M. (2018). BINet: Multivariate business process anomaly detection using deep learning. In *Proceedings of the 16th International Conference on Business Process Management* (pp. 271–287). Springer.

Nolle, T., Seeliger, A., Thoma, N., & Mühlhäuser, M. (2020). DeepAlign: Alignment-based process anomaly correction using recurrent neural networks. In *Proceedings of the 32nd International Conference on Advanced Information Systems Engineering* (pp. 319–333). Springer.

Oberdorf, F., Schaschek, M., Weinzierl, S., Stein, N., Matzner, M., & Flath, C. M. (2023). Predictive end-to-end enterprise process network monitoring. *Business & Information Systems Engineering*, *49*–64.

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, *37*(1), 879–910.

OpenAI (2023). GPT-4 technical report. (pp. 1–100). arXiv preprint arXiv:2303.08774v3, arXiv:2303.08774v3.

Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(2), 604–624.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. http://dx.doi.org/10.1109/TKDE.2009.191.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183–199. http://dx.doi.org/10.1016/j.im.2014.08.008.

Park, G., & Song, M. (2019). Prediction-based resource allocation using LSTM and minimum cost and maximum flow algorithm. In *Proceedings of the 1st International Conference on Process Mining* (pp. 121–128). IEEE.

Park, G., & Song, M. (2020). Predicting performances in business processes using deep neural networks. *Decision Support Systems*, *129*, Article 113191.

Pasquadibisceglie, V., Appice, A., Castellano, G., & Malerba, D. (2019). Using convolutional neural networks for predictive process analytics. In *Proceedings of the 1st International Conference on Process Mining* (pp. 129–136). IEEE.

Pasquadibisceglie, V., Appice, A., Castellano, G., & Malerba, D. (2020). Predictive process mining meets computer vision. In *Proceedings of the 18th International Conference of Business Process Management (Forum)* (pp. 176–192). Springer.

Pasquadibisceglie, V., Appice, A., Castellano, G., & Malerba, D. (2021). A multi-view deep learning approach for predictive business process monitoring. *IEEE Transactions on Services Computing*, *15*(4), 2382–2395.

Pasquadibisceglie, V., Appice, A., Castellano, G., Malerba, D., & Modugno, G. (2020). ORANGE: Outcome-oriented predictive process monitoring based on image encoding and CNNs. *IEEE Access*, *8*, 184073–184086.

Pasquadibisceglie, V., Castellano, G., Appice, A., & Malerba, D. (2021). FOX: A neuro-fuzzy model for process outcome prediction and explanation. In *Proceedings of the 3rd International Conference on Process Mining* (pp. 112–119). IEEE.

Pauwels, S., & Calders, T. (2020). Bayesian network based predictions of business processes. In *Proceedings of the 18th International Conference of Business Process Management (Forum)* (pp. 159–175). Springer.

Pauwels, S., & Calders, T. (2021). Incremental predictive process monitoring: The next activity case. In *Proceedings of the 19th International Conference on Business Process Management* (pp. 123–140). Springer.

Pegoraro, M., Uysal, M. S., Georgi, D. B., & van der Aalst, W. M. P. (2021). Text-aware predictive monitoring of business processes. In *Proceedings of 24th International Conference on Business Information Systems* (pp. 120–129). Springer.

Pery, A., Rafiei, M., Simon, M., & van der Aalst, W. (2022). Trustworthy artificial intelligence and process mining: Challenges and opportunities. In *Proceedings of 3rd International Conference on Process Mining (Workshops)* (pp. 395–407). IEEE.

Pfeiffer, P., Lahann, J., & Fettke, P. (2021). Multivariate business process representation learning utilizing gramian angular fields and convolutional neural networks. In *Proceedings of the 19th International Conference on Business Process Management* (pp. 327–344). Springer.

Pflug, J., & Rinderle-Ma, S. (2016). Application of dynamic instance queuing to activity sequences in cooperative business process scenarios. *International Journal of Cooperative Information Systems*, *25*(01), Article 1650002.

Pham, D.-L., Ahn, H., Kim, K.-S., & Kim, K. P. (2021). Process-aware enterprise social network prediction and experiment using LSTM neural network models. *IEEE Access*, *9*, 57922–57940.

Philipp, P., Jacob, R., Robert, S., & Beyerer, J. (2020). Predictive analysis of business processes using neural networks with attention mechanism. In *Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication* (pp. 225–230). IEEE.

Polančič, G., Jagečič, S., & Kous, K. (2020). An empirical investigation of the effectiveness of optical recognition of hand-drawn business process elements by applying machine learning. *IEEE Access*, *8*, 206118–206131. http://dx.doi.org/10.1109/ACCESS.2020.3034603.

Polato, M., Sperduti, A., Burattin, A., & de Leoni, M. (2014). Data-aware remaining time prediction of business process instances. In *Proceedings of the 2014 International Joint Conference on Neural Networks* (pp. 816–823). IEEE.

Polato, M., Sperduti, A., Burattin, A., & de Leoni, M. (2018). Time and activity sequence prediction of business process instances. *Computing*, *100*(9), 1005–1031.

Qafari, M. S., & van der Aalst, W. M. P. (2019). Fairness-aware process mining. In *Proceedings of the 2019 OTM Confederated International Conferences "On The Move to Meaningful Internet Systems"* (pp. 182–192). Springer.

Qian, C., Wen, L., Kumar, A., Lin, L., Lin, L., Zong, Z., et al. (2020). An approach for process model extraction by multi-grained text classification. In *Proceedings of the 32nd International Conference on Advanced Information Systems Engineering* (pp. 268–282). Springer.

Qiao, M., Akkiraju, R., & Rembert, A. J. (2011). Towards efficient business process clustering and retrieval: combining language modeling and structure matching. In *Proceedings of the 9th International Conference on Business Process Management* (pp. 199–214). Springer.

Rama-Maneiro, E., Vidal, J. C., & Lama, M. (2021). Deep learning for predictive business process monitoring: Review and benchmark. *IEEE Transactions on Services Computing*, 1–18. http://dx.doi.org/10.1109/TSC.2021.3139807.

Recker, J., & Mendling, J. (2016). The state of the art of business process management research as published in the BPM conference: Recommendations for progressing the field. *Business & Information Systems Engineering, 58*(1), 55–72. http://dx.doi.org/10.1007/s12599-015-0411-3.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). ACM.

Rizzi, W., Di Francescomarino, C., & Maggi, F. M. (2020). Explainability in predictive process monitoring: When understanding helps improving. In *Proceedings of the 18th International Conference on Business Process Management (Forum)* (pp. 141–158). Springer.

Rogge-Solti, A., & Kasneci, G. (2014). Temporal anomaly detection in business processes. In *Proceedings of the 12th International Conference on Business Process Management* (pp. 234–249). Springer.

Romano, J. D., Le, T. T., La Cava, W., Gregg, J. T., Goldberg, D. J., Chakraborty, P., et al. (2021). PMLB v1.0: an open source dataset collection for benchmarking machine learning methods. arXiv preprint arXiv:2012.00058v2.

Rozinat, A., & van der Aalst, W. M. P. (2006). Decision mining in ProM. In *Proceedings of the 4th International Conference on Business Process Management* (pp. 420–425). Springer.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). Learning internal representations by error propagation. USA: California University.

Russel, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall.

Samiri, M. Y., Najib, M., El Fazziki, A., & Boukachour, J. (2017). Toward a self-adaptive workflow management system through learning and prediction models. *Arabian Journal for Science and Engineering, 42*(2), 897–912.

Saraeian, S., Shirazi, B., & Motameni, H. (2019). Optimal autonomous architecture for uncertain processes management. *Information Sciences, 501*, 84–99.

Sarno, R., Ginardi, H., Pamungkas, E. W., & Sunaryono, D. (2013). Clustering of ERP business process fragments. In *Proceedings of the 2013 International Conference on Computer, Control, Informatics and its Applications* (pp. 319–324). IEEE.

Sarno, R., Sari, P. L. I., Ginardi, H., Sunaryono, D., & Mukhlash, I. (2013). Decision mining for multi choice workflow patterns. In *Proceedings of the 2013 International Conference on Computer, Control, Informatics and its Applications* (pp. 337–342). IEEE.

Sarno, R., & Sungkono, K. R. (2016a). Hidden markov model for process mining of parallel business processes. *International Review on Computers and Software, 11*(4), 290–300.

Sarno, R., & Sungkono, K. R. (2016b). Coupled hidden markov model for process mining of invisible prime tasks. *International Review on Computers and Software, 11*(6), 539–547.

Savickas, T., & Vasilecas, O. (2018). Belief network discovery from event logs for business process analysis. *Computers in Industry, 100*, 258–266.

Seeliger, A., Luettgen, S., Nolle, T., & Mühlhäuser, M. (2021). Learning of process representations using recurrent neural networks. In *Proceedings of the 33rd International Conference on Advanced Information Systems Engineering* (pp. 109–124). Springer.

Seeliger, A., Nolle, T., & Mühlhäuser, M. (2018). Finding structure in the unstructured: Hybrid feature set clustering for process discovery. In *Proceedings of the 16th International Conference on Business Process Management* (pp. 288–304). Springer.

Senderovich, A., Di Francescomarino, C., Ghidini, C., Jorbina, K., & Maggi, F. M. (2017). Intra and inter-case features in predictive process monitoring: A tale of two dimensions. In *Proceedings of the 15th International Conference on Business Process Management* (pp. 306–323). Springer.

Senderovich, A., Weidlich, M., Gal, A., & Mandelbaum, A. (2014). Mining resource scheduling protocols. In *Proceedings of the 12th International Conference on Business Process Management* (pp. 200–216). Springer.

Senderovich, A., Weidlich, M., Gal, A., & Mandelbaum, A. (2015). Queue mining for delay prediction in multi-class service processes. *Information Systems, 53*, 278–295.

Settles, B. (2009). *Active learning literature survey: Technical Report*, USA: University of Wisconsin-Madison.

Shraga, R., Gal, A., Schumacher, D., Senderovich, A., & Weidlich, M. (2020). Process discovery with context-aware process trees. *Information Systems, 106*, Article 101533.

Silva, R., Zhang, J., & Shanahan, J. G. (2005). Probabilistic workflow mining. In *Proceedings of the 11th International Conference on Knowledge Discovery in Data Mining* (pp. 275–284). ACM.

Sim, S., Sutrisnowati, R. A., Won, S., Lee, S., & Bae, H. (2022). Automatic conversion of event data to event logs using CNN and event density embedding. *IEEE Access, 10*, 15994–16009.

Sindhgatta, R., Moreira, C., Ouyang, C., & Barros, A. (2020). Exploring interpretable predictive models for business processes. In *Proceedings of the 18th International Conference on Business Process Management* (pp. 257–272). Springer.

Sommers, D., Menkovski, V., & Fahland, D. (2021). Process discovery using graph neural networks. In *Proceedings of the 3rd International Conference on Process Mining* (pp. 40–47). IEEE.

Stierle, M., Brunk, J., Weinzierl, S., Zilker, S., Matzner, M., & Becker, J. (2021). Bringing light into the darkness – A systematic literature review on explainable predictive business process monitoring techniques. In *Proceedings of the 29th European Conference on Information Systems* (pp. 1–12). AIS.

Stierle, M., Weinzierl, S., Harl, M., & Matzner, M. (2021). A technique for determining relevance scores of process activities using graph-based neural networks. *Decision Support Systems, 144*, Article 113511. http://dx.doi.org/10.1016/j.dss.2021.113511.

Tavares, G. M., Ceravolo, P., da Costa, V. G. T., Damiani, E., & Junior, S. B. (2019). Overlapping analytic stages in online process mining. In *Proceedings of the 2019 IEEE International Conference on Services Computing* (pp. 167–175). IEEE.

Tax, N., Sidorova, N., Haakma, R., & van der Aalst, W. M. P. (2016). Event abstraction for process mining using supervised learning techniques. In *Proceedings of SAI Intelligent Systems Conference* (pp. 251–269). Springer.

Tax, N., Verenich, I., La Rosa, M., & Dumas, M. (2017). Predictive business process monitoring with LSTM neural networks. In *Proceedings of the 29th International Conference on Advanced Information Systems Engineering* (pp. 477–492). Springer.

Taymouri, F., La Rosa, M., Dumas, M., & Maggi, F. M. (2021). Business process variant analysis: Survey and classification. *Knowledge-Based Systems, 211*, Article 106557. http://dx.doi.org/10.1016/j.knosys.2020.106557.

Taymouri, F., La Rosa, M., & Erfani, S. (2021). A deep adversarial model for suffix and remaining time prediction of event sequences. In *Proceedings of the 2021 SIAM International Conference on Data Mining* (pp. 522–530). SIAM.

Taymouri, F., La Rosa, M., Erfani, S., Bozorgi, Z. D., & Verenich, I. (2020). Predictive business process monitoring via generative adversarial nets: The case of next event prediction. In *Proceedings of the 18th International Conference on Business Process Management* (pp. 237–256). Springer.

Teinemaa, I., Dumas, M., La Rosa, M., & Maggi, F. M. (2019). Outcome-oriented predictive process monitoring: Review and benchmark. *ACM Transactions on Knowledge Discovery from Data, 13*(2), 1–57. http://dx.doi.org/10.1145/3301300.

Teinemaa, I., Dumas, M., Maggi, F. M., & Di Francescomarino, C. (2016). Predictive business process monitoring with structured and unstructured data. In *Proceedings of the 14th International Conference on Business Process Management* (pp. 401–417). Springer.

Teinemaa, I., Tax, N., de Leoni, M., Dumas, M., & Maggi, F. M. (2018). Alarm-based prescriptive process monitoring. In *Proceedings of the 16th International Conference on Business Process Management (Forum)* (pp. 91–107). Springer.

Tello, G., Gianini, G., Mizouni, R., & Damiani, E. (2019). Machine learning-based framework for log-lifting in business process mining applications. In *Proceedings of the 17th International Conference on Business Process Management* (pp. 232–249). Springer.

Theis, J., & Darabi, H. (2019). Decay replay mining to predict next process events. *IEEE Access, 7*, 119787–119803.

Theis, J., & Darabi, H. (2020). Adversarial system variant approximation to quantify process model generalization. *IEEE Access, 8*, 194410–194427.

Tiwari, A., Turner, C. J., & Majeed, B. (2008). A review of business process mining: State-of-the-art and future trends. *Business Process Management Journal, 14*(1), 5–22.

Trkman, P. (2010). The critical success factors of business process management. *International Journal of Information Management, 30*, 125–134. http://dx.doi.org/10.1016/j.ijinfomgt.2009.07.003.

Turner, C. J., Tiwari, A., & Mehnen, J. (2008). A genetic programming approach to business process mining. In *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation* (pp. 1307–1314).

Valdés, J. J., Céspedes-González, Y., & Pou, A. (2022a). Process mining as a time series analysis tool via conformance checking. In *Proceedings of the 2021 Future Technologies Conference* (pp. 636–649). Springer, http://dx.doi.org/10.1007/978-3-030-89906-6_42.

Valdés, J. J., Céspedes-González, Y., & Pou, A. (2022b). Process mining as a time series analysis tool via graph kernels. In *Proceedings of the 2021 Future Technologies Conference* (pp. 832–847). Springer, http://dx.doi.org/10.1007/978-3-030-89906-6_53.

van der Aa, H., Carmona Vargas, J., Leopold, H., Mendling, J., & Padró, L. (2018). Challenges and opportunities of applying natural language processing in business process management. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 2791–2801). ACL.

van der Aalst, W. M. P. (2016). Process mining. In *Data science in action* (2nd ed.). Springer.

van der Aalst, W. (2019). Object-centric process mining: Dealing with divergence and convergence in event data. In *Proceedings of the 17th International Conference on Software Engineering and Formal Methods* (pp. 3–25). Springer.

van der Aalst, W. M. P., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., et al. (2011). Process mining manifesto. In *Proceedings of the 9th International Conference of Business Process Management (Workshops)* (pp. 169–194). Springer.

van der Aalst, W. M. P., Bichler, M., & Heinzl, A. (2017). Responsible data science. *Business & Information Systems Engineering*, *59*(5), 311–313.
van der Aalst, W. M. P., De Medeiros, A. K. A., & Weijters, A. J. (2005). Genetic process mining. In *Proceedings of the 26th International Conference on Application and Theory of Petri Nets* (pp. 48–69). Springer.
van Dongen, B. F., Jansen-Vullers, M. H., Verbeek, H., & van der Aalst, W. M. (2007). Verification of the SAP reference models using EPC reduction, state-space analysis, and invariants. *Computers in Industry*, *58*(6), 578–601.
Varela-Vaca, Á. J., Galindo, J. A., Ramos-Gutiérrez, B., Gómez-López, M. T., & Benavides, D. (2019). Process mining to unleash variability management: Discovering configuration workflows using logs. In *Proceedings of the 23rd International Systems and Software Product Line Conference* (pp. 265–276). ACM.
Vázquez-Barreiros, B., Mucientes, M., & Lama, M. (2014). A genetic algorithm for process discovery guided by completeness, precision and simplicity. In *Proceedings of the 12th International Conference on Business Process Management* (pp. 118–133). Springer.
Vázquez-Barreiros, B., Mucientes, M., & Lama, M. (2015). ProDiGen: Mining complete, precise and minimal structure process models with a genetic algorithm. *Information Sciences*, *294*, 315–333.
Verenich, I., Dumas, M., La Rosa, M., Maggi, F. M., & Di Francescomarino, C. (2016). Minimizing overprocessing waste in business processes via predictive activity ordering. In *Proceedings of the 28th International Conference on Advanced Information Systems Engineering* (pp. 186–202). Springer.
Verenich, I., Dumas, M., La Rosa, M., Maggi, F. M., & Teinemaa, I. (2019). Survey and cross-benchmark comparison of remaining time prediction methods in business process monitoring. *ACM Transactions on Intelligent Systems and Technology*, *10*(4), 1–34. http://dx.doi.org/10.1145/3331449.
Vergidis, K., Tiwari, A., Majeed, B., & Roy, R. (2007). Optimisation of business process designs: An algorithmic approach with multiple objectives. *International Journal of Production Economics*, *109*(1–2), 105–121.
Vidgof, M., Bachhofner, S., & Mendling, J. (2023). Large language models for business process management: Opportunities and challenges. (pp. 1–16). arXiv preprint arXiv:2304.04309, arXiv:2304.04309.
vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems*, *37*, 205–224.
Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, *2018*.
Wahid, N. A., Adi, T. N., Bae, H., & Choi, Y. (2019). Predictive business process monitoring – Remaining time prediction using deep neural network with entity embedding. *Procedia Computer Science*, *161*, 1080–1088.
Wang, J., Wong, R. K., Ding, J., Guo, Q., & Wen, L. (2012). Efficient selection of process mining algorithms. *IEEE Transactions on Services Computing*, *6*(4), 484–496.
Wang, J., Yu, D., Liu, C., & Sun, X. (2019). Outcome-oriented predictive process monitoring with attention-based bidirectional LSTM neural networks. In *Proceedings of the 2019 IEEE International Conference on Web Services* (pp. 360–367). IEEE.
Wanner, J., Wissuchek, C., & Janiesch, C. (2020). Machine learning and complex event processing: A review of real-time data analytics for the industrial internet of things. *Enterprise Modelling and Information Systems Architectures*, *15*, 1.
Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. King's College.
Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, xiii–xxiii.
Weinzierl, S., Dunzer, S., Zilker, S., & Matzner, M. (2020). Prescriptive business process monitoring for recommending next best actions. In *Proceedings of the 18th International Conference of Business Process Management (Forum)* (pp. 193–209). Springer.
Weinzierl, S., Wolf, V., Pauli, T., Beverungen, D., & Matzner, M. (2022). Detecting temporal workarounds in business processes – A deep-learning-based method for analysing event log data. *Journal of Business Analytics*, *5*(1), 76–100.
Weinzierl, S., Zilker, S., Stierle, M., Matzner, M., & Park, G. (2020). From predictive to prescriptive process monitoring: Recommending the next best actions instead of calculating the next most likely events. In *Proceedings of the 15th International Conference on Wirtschaftsinformatik* (pp. 364–368).
Wetzstein, B., Leitner, P., Rosenberg, F., Brandic, I., Dustdar, S., & Leymann, F. (2009). Monitoring and analyzing influential factors of business process performance. In *Proceedings of the 2009 IEEE International Enterprise Distributed Object Computing Conference* (pp. 141–150). IEEE.
Weytjens, H., & De Weerdt, J. (2021). Learning uncertainty with artificial neural networks for improved remaining time prediction of business processes. In *Proceedings of the 19th International Conference on Business Process Management* (pp. 141–157). Springer.
Weytjens, H., & Weerdt, J. D. (2021). Creating unbiased public benchmark datasets with data leakage prevention for predictive process monitoring. In *Proceedings of the 19th International Conference on Business Process Management (Workshops)* (pp. 18–29). Springer.
Wibisono, A., Nisafani, A. S., Bae, H., & Park, Y.-J. (2015). On-the-fly performance-aware human resource allocation in the business process management systems environment using naïve Bayes. In *Proceedings of the 3rd Asia-Pacific Conference on Business Process Management* (pp. 70–80). Springer.
Wickramanayake, B., He, Z., Ouyang, C., Moreira, C., Xu, Y., & Sindhgatta, R. (2022). Building interpretable models for business process prediction using shared and specialised attention mechanisms. *Knowledge-Based Systems*, *248*, Article 108773.
Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, *135*, 364–381.
Xu, J., & Liu, J. (2019). A profile clustering based event logs repairing approach for process mining. *IEEE Access*, *7*, 17872–17881.
Xu, J., Liu, C., Zhao, X., Yongchareon, S., & Ding, Z. (2016). Resource management for business process scheduling in the presence of availability constraints. *ACM Transactions on Management Information Systems*, *7*(3), 1–26.
Xu, C., Tao, D., & Xu, C. (2013). A survey on multi-view learning. (pp. 1–59). arXiv preprint arXiv:1304.5634, arXiv:1304.5634.
Yang, S., Dong, X., Sun, L., Zhou, Y., Farneth, R. A., Xiong, H., et al. (2017). A data-driven process recommender framework. In *Proceedings of the 23rd International Conference on Knowledge Discovery and Data Mining* (pp. 2111–2120). ACM.
Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, *10*(2), 1–19. http://dx.doi.org/10.1145/3298981.
Yang, J., Ouyang, C., Pan, M., Yu, Y., & ter Hofstede, A. H. (2018). Finding the "liberos": Discover organizational models with overlaps. In *Proceedings of the 16th International Conference on Business Process Management* (pp. 339–355). Springer.
Yeshchenko, A., Di Ciccio, C., Mendling, J., & Polyvyanyy, A. (2019). Comprehensive process drift detection with visual analytics. In *Proceedings of the 38th International Conference on Conceptual Modeling* (pp. 119–135). Springer.
Yeshchenko, A., Durier, F., Revoredo, K., Mendling, J., & Santoro, F. (2018). Context-aware predictive process monitoring: The impact of news sentiment. In *Proceedings of the 2018 OTM Confederated International Conferences "On The Move to Meaningful Internet Systems"* (pp. 586–603). Springer.
Zairi, M. (1997). Business process management: A boundaryless approach to modern competitiveness. *Business Process Management Journal*, *3*, 64–80. http://dx.doi.org/10.1108/14637159710161585.
Zanzotto, F. M. (2019). Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research*, *64*, 243–252.
Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 1–20. http://dx.doi.org/10.1109/TKDE.2021.3070203.
Zhao, W., Liu, H., Dai, W., & Ma, J. (2016). An entropy-based clustering ensemble method to support resource allocation in business process management. *Knowledge and Information Systems*, *48*(2), 305–330. http://dx.doi.org/10.1007/s10115-015-0879-7.
Zhao, W., Pu, S., & Jiang, D. (2020). A human resource allocation method for business processes using team faultlines. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, *50*(9), 2887–2900.
Zhu, X. (2005). *Semi-supervised learning literature survey: Technical Report*, USA: University of Wisconsin-Madison.

ANEXO 16: GUZZO, JOARISTI, & ANTONIO RULLO, 2021 - A MULTI-PERSPECTIVE APPROACH FOR THE ANALYSIS OF COMPLEX BUSINESS PROCESSES BEHAVIOR

Expert Systems With Applications 177 (2021) 114934

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A multi-perspective approach for the analysis of complex business processes behavior

Antonella Guzzo [a,*,1], Mikel Joaristi [b,4], Antonino Rullo [a,2], Edoardo Serra [b,3]

[a] University of Calabria, via Pietro Bucci, 87036 Rende, Italy
[b] Boise State University, 777 W Main St, 83702 Boise, ID, USA

ARTICLE INFO

Keywords:
Process mining
Trace embedding
Multi-perspective analysis
Deep learning

ABSTRACT

Business processes are often monitored by transactional information systems that produce massive dataset called event logs. Such logs contain the process execution traces, typically characterized by heterogeneous and high-dimensional data. Process mining techniques offer a great opportunity to gain valuable knowledge hidden in the data to be used for analysing the multiple characteristics of processes (i.e. perspectives in process mining, like structural aspects, activities, resources, data and time). Therefore, raw data must be encoded into a suitable format that can be more conveniently provided to the mining algorithms. However, most of the existing process encoding techniques focus on the control-flow perspective, i.e. only encode the sequence of activities that characterize a trace, leaving out other process perspectives that are fundamental for describing the process behavior in all its aspects. In this paper we address the problem of computing a concise and informative representation of execution traces that considers the multiple perspectives of the process behavior. We propose a holistic approach that computes trace embedding able to capture patterns of dependencies between the perspectives that are lost in a one-dimensional analysis and, at the same time, it is unsupervised, meaning that no a priori knowledge is needed. The experiments conducted on two real life logs demonstrate that our proposed embedding is appropriate to concisely describe the multiple and various characteristics of the processes, and that the proposed method outperforms existing trace encoding techniques. Furthermore, the embedding includes the elapsed time between events as an additional feature to make us capable to use it as a further dimension of analysis.

1. Introduction

Process mining is a research field aiming at developing methodologies and techniques to discover and improve business processes on the basis of the knowledge that can be mined from the data collected in the logs of transaction information systems (van der Aalst, 2016; dos Santos Garcia et al., 2019). In real-life applications, such logs are given as a set of process executions, called *traces* or event log, that contains high-dimensional and heterogeneous information as for example activities performed, activity executors (e.g. employers, systems, departments, etc.), timestamp and data in general. These information can be analysed along with three process behavior perspectives:

- the *control-flow* perspective, which focuses on the activities that are executed and the relationships of precedence among them (in terms of preconditions and postconditions);
- the *organizational* perspective, which focuses on the actors that are involved, on their roles, and on how they are mapped/assigned to the activities;
- the *data* perspective, which focuses on the data used and generated during the process execution, including the temporal data, the

* Corresponding author.
E-mail addresses: antonella.guzzo@unical.it (A. Guzzo), mikeljoaristi@boisestate.edu (M. Joaristi), n.rullo@dimes.unical.it (A. Rullo), edoardoserra@boisestate.edu (E. Serra).
1 ORCID ID: https://orcid.org/0000-0003-3159-0536
2 ORCID ID: https://orcid.org/0000-0002-6030-0027
3 ORCID ID: https://orcid.org/0000-0003-0689-5063
4 ORCID ID: https://orcid.org/0000-0001-9878-8366

https://doi.org/10.1016/j.eswa.2021.114934
Received 1 December 2020; Received in revised form 24 February 2021; Accepted 20 March 2021
Available online 29 March 2021

attributes that characterize the given enactment, and the attributes that characterize the activities and the triggered events.

Effective process mining methods must be capable to simultaneously look at all the process' perspectives, along with their associated information, and at their cross-correlations in a holistic way. However, most of the process mining techniques focus only on the control-flow perspective (Bezerra, Wainer, & van der Aalst, 2009; De Koninck, vanden Broucke, & De Weerdt, 2018; Yang & Hwang, 2006; Jans, Van Der Werf, Lybaert, & Vanhoof, 2011) (i.e. the sequence of the activities) or look into different perspectives to each one. Recently, multivariate-based techniques (Böhmer & Rinderle-Ma, 2016; Nolle, Seeliger, & Mühlhäuser, 2018; Rullo, Guzzo, Serra, & Tirrito, 2020; van Eck, Sidorova, & van der Aalst, 2016; De Weerdt, Schupp, Vanderloock, & Baesens, 2013) have been proposed. In these works more than one perspective is taken into account for the process discovery task, and in some cases the *time* perspective is considered as well, as a further data dimension, but separately from the others. However, these techniques *(i)* do not consider all the perspectives at once, leaving out important aspects of the behavior of the process under analysis; *(ii)* only consider the time instant in which events occur, leaving out the time elapsed between events which is a useful feature for describing the process behavior on the basis of the rate at which events occur.

Another important aspect when dealing with learning tasks of multidimensional data is the format the data is represented with. In order for mining algorithms to work with heterogeneous data, it is crucial that the process logs provided as input are represented in a way that is *(i)* appropriate to concisely and accurately describe the heterogeneous characteristics of the process behavior, and *(ii)* useful to build classifiers and other predictive/descriptive models (Bengio, Courville, & Vincent, 2013).

Representation learning plays therefore a key role in process mining—much more than it happens in other application domains for mining/learning methods.

Actually, most of the representation learning methods proposed for business processes come in the form of *feature selection* algorithms, which either look at a single perspective at a time (Böhmer & Rinderle-Ma, 2017; Folino, Guarascio, & Pontieri, 2012), or compute "aggregate" features by just concatenating the features extracted over the different perspectives (Bose & van der Aalst, 2009; Luca, Pimentel, Watkinson, & Clifton, 2018; Polato et al., 2018; Rullo et al., 2020). However, it has been noticed that these approaches are not the best choice when dealing with high-dimensional and multivariate data (Mhaskar & Poggio, 2016).

De Koninck et al. (2018) proposed a method that overcomes these drawbacks. Their approach does not focus on the extraction of specific features, but rather computes an *embedding* of the input data into a novel vectorial space— hereinafter we refer to this kind of approach as a *trace embedding* method. Even though this method deals with the control-flow perspective only, it has been proven to be quite effective when used for pre-processing data in inference tasks such as trace clustering, process comparison, predictive process monitoring and process anomaly detection. However, the authors left open *(i)* to develop embedding methods that simultaneously take into account the various process perspectives and *(ii)* to investigate more elaborated architectures in order to improve the performances of the inference tasks.

1.1. Contributions

To address the above research issues, we proposed a holistic method for analysing process behavior aiming at capturing patterns of dependencies between perspectives that are lost in one-dimensional analysis. Moreover, our method is tailored to time analysis by expliciting elapsed time between events as further dimension of analysis. We compute a *multi-perspective representation* of the traces by means a neural network architecture, enriched with an Long Short Time Memory (LSTM) attention mechanism and a time feature, thus avoiding the expensive feature engineering phase task.

LSTMs have proven to be useful for learning sequences containing long-term patterns of unknown length due to their ability to maintain long term memory, which gives them an advantage over alternative approaches like Recurrent Neural Networks and Hidden Markov Models. Our contributions can be summarized as follows:

- we propose a deep learning based framework to find a concise and informative representation of the process behavior by automatically integrating the multidimensional aspects of the traces.
- our approach is based on a LSTM-autoencoder model, Deep-Trace2-Vec, which is able to learn the representation directly from raw temporal data in an unsupervised way, that is, without the need of any a priori knowledge and thus avoiding the expensive feature engineering phase task;
- we include in the input log the elapsed time between events so as to embed useful information about the "speed" at which traces are executed; this allows to explain the process behavior also on the basis of the rate at which events occur;
- we empirically validate the proposed approach over two datasets consisting of real data. In particular, we use the datasets adopted in the BPI Challenge 2015 (Van Dongen & Boudewijn, 2015) and in the BPI Challenge 2019 (Van Dongen & Boudewijn, 2019) competitions. Our results show that the performances of the most relevant process mining methods (clustering and classification) drastically improve w.r.t. the performances that can be obtained with embedding methods earlier proposed in literature.

The rest of the paper is organized as follows: a critical evaluation of literature is provided in Section 2; some preliminary notions are illustrated in Section 3; the multi-perspective embedding framework is elaborated in Section 4; results of our experimental activity are illustrated in Section 5; finally, concluding remarks with avenues for further research are discussed in Section 6.

2. Related work

Representation learning is the task of finding a condensed representation of a given amount of data which guarantees that characteristics relevant for further analysis are preserved after the dimensionality of data has been reduced. In the case of processes, this is a challenging task since real-life event logs contain a large number of cases (traces), potentially representing a highly varied set of distinct event sequences, and usually, also containing additional information on resources and a diverse set of other event or case-related attributes.

There exists a considerable amount of encoding techniques that can be applied to event logs (Tavares, Barbon Junior, Ceravolo, & Damiani, 2020), here we evaluate the most used in the process mining domain.

Graph embedding methods (Goyal & Ferrara, 2018) can encode information structured as graphs, thus, an event-log can be appropriately represented as a graph using vertices as activities/attributes and edges as relations between them. Van Vanongen and Van der Aalst (2004) proposed to represent control-flow instances as graphs with the aim of performing mining tasks on graph instances instead of constructing a process model. Böhmer and Rinderle-Ma (2016) proposed a multi-perspective approach with a probabilistic graph-based model. The graph contains three types of vertices, activities, resources and weekdays, and edges are labeled with the probability that, given the previous triple activity-resource-weekday, a certain activity happens next. The graph is then used to score ongoing traces with the purpose of finding anomalous process executions.

Text-based encodings, grounded in information retrieval and natural language processing, have been used in the context of process mining in various forms due to the analogy between activities and words, and between traces and sentences. The *integer encoding* maps all possible attribute values for each attribute to a unique positive integer. Nolle

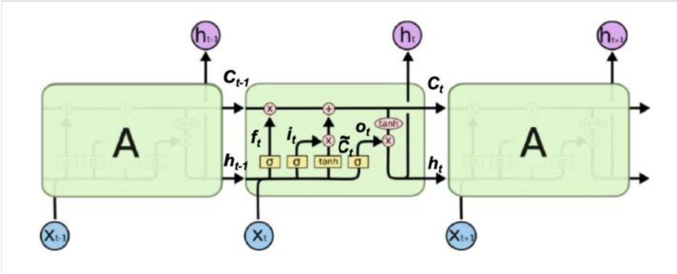


Fig. 1. An LSTM cell C, source (Staudemeyer & Morris, 2019).

et al. (2018) adopted this encoding technique to preprocess data for anomaly detection purposes. The *count vectorizer* encode words/activities by accounting their frequencies in a corpus/trace. The *one-hot encoding* transforms a categorical attribute which shows m different values into m binary features, each observation in the categorical feature is thus converted into a vector of size m with only one of the values as 1. The *n-gram encoding* has the ability to represent relationships between neighboring elements of a sequence by representing a trace as all contiguous subsequences of length n. Mehdiyev, Evermann, and Fettke (2018) applied deep feed-forward multilayer neural networks to predict the next business process event, and encoded traces adopting the n-gram encoding. Rullo et al. (2020) adopted a hybrid approach for the detection of anomalous traces, in which the trace embedding is the concatenation of the vector obtained by one-hot encoding the categorical trace attributes, and the vector obtained by count-vectorizing the 2-grams of the activities' sequence. Still from the field of natural language processing, *word embeddings* are the result of a two-layer neural network built to reconstruct the context of a word in a corpus, so as words occurring in similar contexts generate similar encodings. De Koninck et al. (2018) applied this technique in the field of process mining, and published a landmark paper where general purpose representations for activities (called *act2vec*), traces (called *trace2Vec*), event logs (called *log2vec*) and process models (called *model2vec*) have been proposed. The paper has the merit to present a novel neural network-based embedding technique specified for process concepts and to show how it can be efficiently used for automated inference tasks such as trace clustering, process comparison, predictive process monitoring, and anomaly detection. However, the authors left open to develop embedding methods that simultaneously take into account the various process perspectives, and to investigate more elaborated architectures in order to improve the performances of the inference tasks.

The major drawbacks of the aforementioned techniques are mainly two: *(i)* they do not consider all the process perspective at once, and *(ii)* temporal data is considered only as time points, leaving out time intervals (i.e. time spans between events) that are fundamental for explaining the temporal behavior of processes. On the contrary, our paper computes a trace embedding aiming at preserving the multi-perspectives of processes by exploiting LSTM architecture and autoencoder. In fact, LSTMs have proven to be useful for learning sequences containing long-term patterns of unknown length due to their ability to maintain long term memory. As process' executions are temporal sequences, we think the LSTM be the best choice to build embeddings that maintain information about the temporal relationships between events. The LSTM is the architecture used as prediction of process model in (Navarin, Vincenzi, Polato, & Sperduti, 2017; Evermann, Rehse, & Fettke, 2017), however, differently for us, these works don't compute an embedding and only consider the sequence of the activities and their timestamps.

For completeness it is noteworthy to mention that autoencoders have been adopted in process mining literature mainly as learning tool fed with text-based embeddings with the aim of finding anomalous process executions (Nolle, Luettgen, Seeliger, & Mühlhäuser, 2018). Rather, in this paper we adopt the autoencoder for the purpose of learning highly informative trace embeddings, leaving the data mining task (e.g. classification, clustering, anomaly detection, etc.) to well established machine learning algorithms.

3. Background knowledge

3.1. Long Short-Term Memory

An LSTM is a neural network architecture tailored to deal with sequences of data (Staudemeyer & Morris, 2019). The core idea is to allow the network to "accumulate" information over some period of time, and to create paths through time having derivatives that neither vanish nor explode. An LSTM network contains LSTM units instead of, or in addition to, other standard network units. An LSTM unit remembers values for either long or short time periods, and several units are often implemented in "blocks". LSTM blocks contain three or four "gates" that control the information flow. These gates are implemented using the logistic function to compute a value between 0 and 1. Multiplication is applied to this value to partially allow or deny information to flow into or out of the memory. An LSTM block uses weights to direct gates' operations, applying them to the values that feed into the block. Thus, the LSTM block determines how to maintain its memory as a function of those values, and training its weights causes the block to learn the function that minimizes loss. LSTMs also have a form of a chain of repeating module of a neural network (chain-like structure), where each module has a different structure and different interacting layers. Fig. 1 describes the architecture of a single LSTM cell that outputs the next state h_t by taking in input the previous state h_{t-1} and the next symbol x_t. The operations done by the single LSTM cell C are described by the following equations:

$$\begin{aligned}
a_t &= \rho(W_a \cdot [h_{t-1}; x_t]) \\
b_t &= \rho(W_b \cdot [h_{t-1}; x_t]) \\
y_t &= \tanh(W_y \cdot [h_{t-1}; x_t]) \\
g_t &= \rho(W_g \cdot [h_{t-1}; x_t]) \\
c_t &= c_{t-1} \cdot a_t + b_t \cdot y_t \\
h_t &= \tanh(c_t) \cdot g_t
\end{aligned}$$

where the W_a, W_b, W_y and W_g are the weights representing the LSTM cell C and the entire LSTM neural network.

If a sequence is passed to an LSTM, the network will output, for each element of the sequence, a vectorial representation h_i representing the sub-sequence from the beginning till the element i.

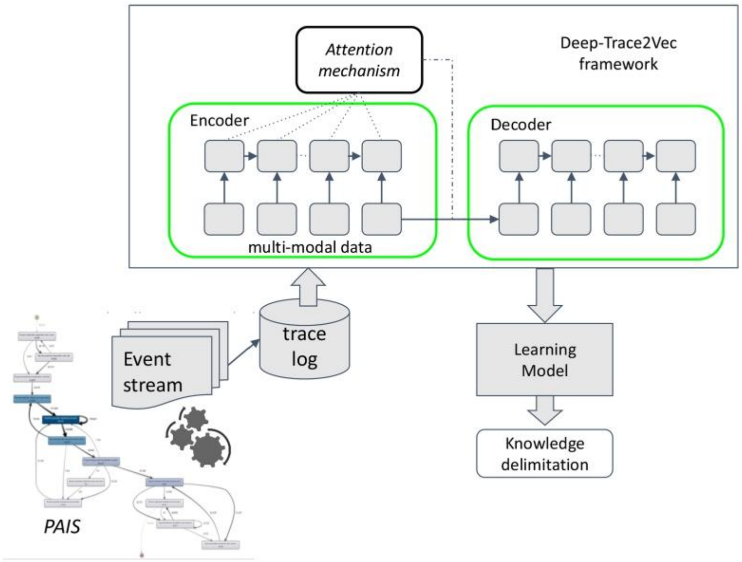


Fig. 2. Overview of the framework.

3.2. Autoencoder

An autoencoder (Meng, Catchpoole, Skillicom, & Kennedy, 2017) is a neural network that is capable to produce an encoding of the input data in an unsupervised manner, in particular by reducing its dimensionality while minimizing reconstruction loss. An autoencoder has two architectural parts: (i) the encoder, aiming at learning an encoded representation (embedding) of the training in different feature spaces by efficiently reducing the dimensionality of the original data space; (ii) the decoder, that tries to reconstruct the original data by taking the embeddings (compressed feature vectors) as input. The output of an autoencoder has the same number of computational units as the input (original feature dimensions). The intuition behind an autoencoder to be used as representation algorithm comes from its ability to effectively reconstruct the examples having similar statistical properties in the original feature space, thus obtaining the smaller reconstruction errors for the known type of objects.

Specifically, the encoder is a function f that maps an input X to a hidden representation Y as formulated by the following equation:

$$Y = f(X) = s_f(WX + b)$$

where s_f is a nonlinear activation function. The encoder is parameterized by a weight matrix W and a bias vector b. The decoder function g maps the hidden representation Y back to a reconstruction $X^{'}$:

$$X^{'} = g(Y) = s_g(W^{'}Y + b^{'})$$

where s_g is the decoder's activation function. The decoder's parameters are a bias vector $b^{'}$ and matrix $W^{'}$. Training an autoencoder involves finding parameters $\pi = (W, W^{'}, b, b^{'})$ that minimize the reconstruction loss on the given dataset X. The objective function is given as

$$\Theta = \min_{\pi} L(X, X^{'}) = \min_{\pi} L(X, g(f(X)))$$

For linear reconstruction, the reconstruction loss (L1) is generally from the squared error:

$$L_1(\pi) = \sum_{i=1}^{n} ||x_i - x_i^{'}|| = \sum_{i=1}^{n} ||x_i - g(f(x_i))||^2$$

For nonlinear reconstruction, the reconstruction loss (L2) is generally from cross-entropy:

$$L_2(\pi) = -\sum_{i=1}^{n} [x_i log(y_i) + (1 - x_i) log(1 - y_i)]$$

where $x_i \in X, x_i^{'} \in X^{'}$, and $y_i \in Y$.

4. Multi-perspective embedding framework

A high-level view of the multi-perspective trace embedding framework we are going to illustrate is depicted in Fig. 2. Without loss of generality, we assume that process executions are caught in a stream of event log and then represented as a bag of traces. In our framework, these traces are passed trough an autoencoder architecture that embeds the process behavior into an informative vectorial representation. The autoencoder uses some LSMT components, plus an *attention mechanism* (Chorowski, Bahdanau, Serdyuk, Cho, & Bengio, 2015) which is a technical component conceived to improve the quality of the results (and discussed in details later in this section). The embedding (that is, the representation that has been learned via the proposed framework) can be then used as input by a number of process mining methods. Our focus will be mainly to assess the gain (in terms of improvement of performances over the respective inference tasks) that our trace embeddings can provide for process clustering and classification algorithms.

The trace embedding method we have just sketched will be called Deep-Trace2Vec. Differently from earlier approaches, Deep-Trace2Vec focuses on process execution traces by integrating multidimensional data into a unique representation. Moreover, it is able to handle all types of attributes: text, categorical, ordinal, and numerical. It is a totally unsupervised algorithm, and as such, it does not need of expert knowledge. In addition, it provides a transformation function able to map a log

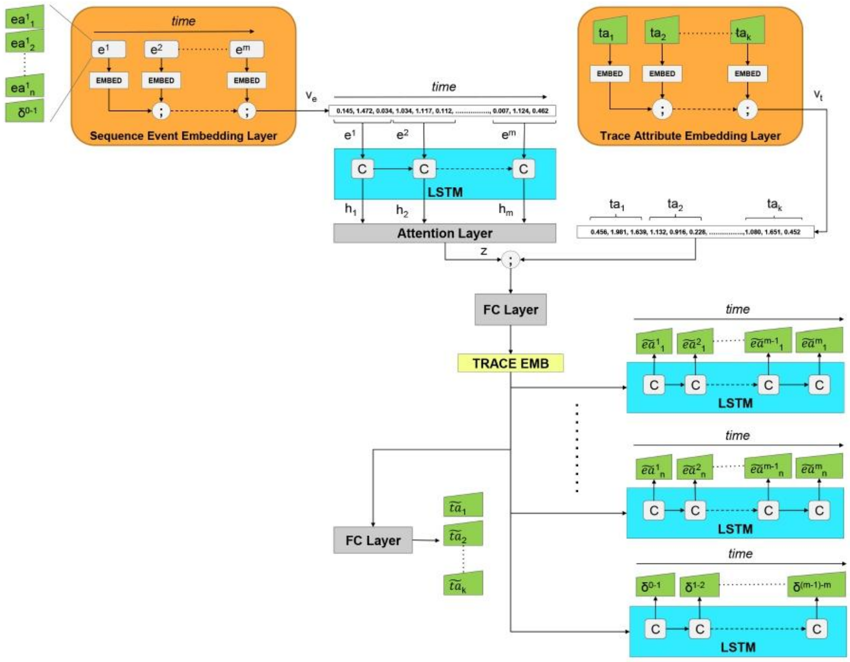


Fig. 3. Deep-Trace2Vec architecture. The arrows labeled with *time* denote the temporal relationship between events, which is preserved in the vector v_e, and in the output $< \widetilde{ea}_1^1 ... \widetilde{ea}_1^m > ... < \widetilde{ea}_n^1 ... \widetilde{ea}_n^m >, < \delta^{0-1} ... \delta^{(m-1)-m} >$.

trace into a vector without the need to re-train the model.

Each architectural component will be illustrated in the rest of this section.

4.1. Encoder

By abstracting from the specificity of the various systems, we can view a log as a set of execution traces of a specific process. An execution trace can be represented as a pair (A, S) where $A = \{ta_1, ..., ta_k\}$ is a set of trace attributes describing the global properties of the trace (e.g. caseID/identifier, who originate execution and system, beginning time), and $S = < e^1, ..., e^m >$ is a temporal ordered sequence of events mapping the sequence of activities performed in the execution's trace with their attributes. Specifically, each event $e^i \in S$ is characterized by a timestamp t^i and a set of event attributes $\{ea_1^i, ..., ea_n^i\}$ describing the intrinsic properties of the event (e.g. activity's name, performing resource, state of activity, input data). Given two events e^i and e^j with $i > j$, we have that $t^i \geqslant t^j$. In particular if $t^i = t^j$ then e^i and e^j are occurring in parallel. Both the trace attributes and the event attributes can be nominal (e.g. strings) or numeric.

Deep-Trace2Vec is trained by using an autoencoder structure operating a multi-perspective integration. The architecture is sketched in Fig. 3. This architecture is inspired to the autoencoder-LSTM model that has been successfully developed for natural language processing problems, and for image captioning. These models are capable of learning the complex dynamics within the temporal ordering of input sequences, and of using an internal memory to remember information across long input sequences.

Specifically, our encoder comprises two separate embeddings: a *Trace Attribute Embedding* layer, which embeds the trace's attributes ta_1, $..., ta_k$ into a numeric vector (denoted as v_t in Fig. 3); and a *Sequence Event Embedding* layer, which embeds the sequence of event attributes.

$$< ea_1^1, ..., ea_n^1, \delta^{0-1}, ..., ea_1^m, ..., ea_n^m, \delta^{(m-1)-m} >$$

into a numeric vector (denoted as v_e in Fig. 3). Note that the elapsed time $\delta^{(i-1)-i}$ between events e^{i-1} and e^i is introduced as part of the events' attribute set, in order to take into account the temporal relation between events. Since trace attributes (resp. event attributes) can be either nominal or numerical, for the nominal attributes our technique creates an embedded representation for each value of the nominal domain, i.e. a vectorial numeric representation for each nominal value. Once the nominal attributes are converted, the trace attributes (resp. event attributes for each event of the sequence) are concatenated (concatenation is denoted as ";" in Fig. 3).

The vectorial representations of event's attributes is consequently passed through an LSTM recurrent neural network, which preserves the temporal correlation of events. For each event e^i of the sequence, the LSTM outputs a vectorial representation h_i summarizing all the events from e^1 to e^i. The sequence $< h_1, ..., h_m >$ is then passed trough an attention layer producing a summarized temporal representation z. In turn, z is concatenated with the vectorial representation of the trace attributes v_t. The concatenation of v_t and z results in a unique vectorial numeric representation which encodes both the trace's and events' attributes and the events temporal ordered sequence. This allows us to capture the multi-perspective aspect of a process, as information of different nature from the traces are included in the embeddings. The resulting vector is then passed to a Fully Connected (FC) layer (with a hyperbolic tangent activation function) that creates the trace embedding representation (TRACE EMB). The key idea is to reduce and remove the sparsity from the concatenation of the two embeddings v_t and z.

4.1.1. Attention mechanism

In the architecture depicted in Fig. 3 a key role is played by the attention layer. We recall here that *attention* (Chorowski et al., 2015) is a mechanism for discovering the parts of the input sequence that are the most relevant for describing the trace behavior. The ides is that these parts should be weighted more than the other parts, in the computation of the embedding. To have an intuition on how the attention mechanism is implemented, consider how human brain processes input information. Notably, the human brain chooses to focus on certain inputs, while ignoring or diminishing the importance of other ones. Inspired by this analogy, in (Bahdanau, Cho, & Bengio, 2014) the authors added an attention layer to the encoder-decoder framework allowing to automatically search for the parts of a source sentence that are relevant for predicting a target word, without having to form these parts as a hard segment explicitly. In our architecture, we shift this idea from the analysis of sentences to the analysis of process traces.

In our context, the attention layer takes as input the vectorial representation sequence $< h_1, \ldots, h_m >$ from the LSTM and returns the output $z = tanh(W_c[co; h_n])$ where.

- $W_c \in \mathscr{R}^{2|h_n| \times |h_n|}$ is a set of weights to learn, and
- co is the attention context vector obtained by the weighed mean of all h_i vectors, i.e., $co = \sum_{i=1}^{n} c_i^s \cdot h_i$.

The weights $c_1^s, \ldots, c_n^s$ are obtained by using (1) a unique fully connected layer that, applied singularly to each h_i, produces a single value q_i, and (2) a softmax layer taking in input all elements in $\{q_1, \ldots, q_n\}$ and producing $c_1^s, \ldots, c_n^s$.

4.2. Decoder

Given the comprehensive trace embedding (TRACE EMB), the final part of our network works on the reconstruction of the attributes taken in the input. In the case of trace's attributes, the reconstruction is performed as in the decoder of an autoencoder using one fully connected layer. In the case of event's attributes, instead, they are reconstructed by using a set of LSTMs, one for each attribute. For instance, consider a toy example where the event's attributes are $\{task, resource, state\}$. According with the notation introduced in the paper, a trace of length m is represented by a sequence of event's attributes: $< ea^1_{task}, ea^1_{resource}, ea^1_{state}, \ldots, ea^m_{task}, ea^m_{resource}, ea^m_{state} >$ with its instantiating values, for example $e^1 = \{ sendEmail, Bob, finished\}$.

In this case, the decoder uses three LSTMs, one for *task*, one for *resource* and one for *state*, in order to reconstruct the original sequence of their values. The idea is to capture the different temporal relations among each event and its attributes. The outputs of the network are the predicted values $\{\widetilde{ta}_1, \ldots, \widetilde{ta}_k\}$ of the trace's attributes, and the predicted values.

$$\{\widetilde{ea}_1^1, \ldots, \widetilde{ea}_n^1, \delta^{0-1}, \ldots, \widetilde{ea}_1^m, \ldots, \widetilde{ea}_n^m, \delta^{(m-1)-m}\}$$

of the events' attributes (included the elapsed time between events δ). To train the assembled neural network, we use a cumulative loss function that:

- for each numerical attribute sums the mean squared error between the reconstructed attribute $\widetilde{ta}_i$ (or the reconstructed sequence of attribute values $< \widetilde{ea}_i^1, \widetilde{ea}_i^2, \ldots, \widetilde{ea}_i^{m-1}, \widetilde{ea}_i^m >$) with the original attribute value ta_i (or the original attribute values of the sequences $< ea_i^1, ea_i^2, \ldots, ea_i^{m-1}, ea_i^m >$), and
- for each nominal attribute uses the cross-entropy and the softmax activation function to create a distribution among all the nominal values. In particular, softmax function outputs a vector that represents the probability distribution of the list of nominal attributes (with all probabilities sum to one), and the cross-entropy is used as a loss function to minimize in order to reduce the gap between the predicted value and the ground truth.

5. Experimental evaluation

In this section we illustrate the experimental settings we used to evaluate the effectiveness of our approach, and then we discuss the results of our experimental activity. Our goal is to provide answers to the following research questions:

RQ1 How does Deep-Trace2Vec perform compared with state-of-the-art techniques in process representation?
RQ2 How does the performance of Deep-Trace2Vec vary with respect to the multi-perspective specification?

The rest of the section is organized as follows: in Section 5.1 we describe the datasets used to test our approach; in Section 5.2 we provide details on the trace classification and trace clustering tasks used to evaluate and compare our embedding approach with the baselines; in Section 5.3 and Section 5.4 we describe the baselines we compare with and the implementation details of each embedding technique, respectively; in Section 5.5 we describe the experiment results for both the process mining tasks considered; finally, in Section 5.6 we summarize the experiment results and answer to the above research questions.

5.1. Datasets

To evaluate the usefulness of the proposed embedding technique, we apply it to two event logs for which a ground truth is known. In particular, we perform the evaluation over two of the well known real-life datasets provided by the BPI Challenge, an annual process mining competition. This datasets are widely used as benchmarks in the evaluation of process mining techniques. Below it follows a brief description of the datasets.

BPI Challenge 2015. The dataset (Van Dongen & Boudewijn, 2015), namely *BPIC_15* in short, consists of data generated from the building permit application process in five municipalities located in The Netherlands. The data contains 5.647 cases, and 27 types of activities for a total of 92.426 events over a period of four years. Each record describes a single step taken by an employee in the process, and consists of the case ID, the activity performed (given by its ID as well as its description in both English and Dutch), the time the action was completed, involved resources, and many other fields. Each trace is characterized by 6 nominal attributes and 1 numeric attribute, while each event is characterized by 6 nominal attributes. We consider the five municipalities as the class labels (values 1, 2, 3, 4, 5). *BPIC_15* can be considered a balanced dataset as the five classes are distributed as following: 1 −17%, 2 −17%, 3 −19%, 4 −22%, 5 −25%.

BPI Challenge 2019. The dataset (Van Dongen & Boudewijn, 2019), namely *BPIC_19* in short, collects data from a large multinational company operating with 60 subsidiaries from The Netherlands in the area of coatings and paints. Specifically, the data is recorded for the execution of Purchase-To-Pay processes (possibly of an ERP system) of purchase orders submitted in 2018, where each purchase order (or purchase document) contains one or more line items. For each line item, there are roughly two types of flows in the data, designed by the *Item Category* attribute which we use as class label (values 1, 2). The data consists of 76.349 traces, and 42 types of activities performed by 627 users, for a total of 1.595.923 events recorded over a period of three years. Each trace is characterized by 6 nominal attributes and 1 numeric attribute, while each event is characterized by 2 nominal attributes. *BPIC_19* is an unbalanced dataset as the two classes are distributed as following: 1 −77%, 2 −23%.

5.2. Experimental setting

To validate our embedding technique we used the trace clustering and trace classification tasks, that are actually the most used in the process mining settings.

5.2.1. Trace classification

Trace classification is a process mining task that allows to determine whether traces belong or not to a predefined class. Trace classification has been proved to be useful in the case of conformance checking and predictive monitoring. In fact, the former can be reformulated as a binary classification task where the problem is to determine if traces fit (or not) the process model, i.e. traces can be replayed by the underline model (Okoye Dr, Naeem Dr, Islam Dr, Tawil Dr, & Lamine Dr, 2017). Similarly, the latter can be reformulated as a multi-classification task where the problem is to classify traces based on their processing time, or the associated outcome (e.g. success or failure). We executed three classification algorithms, namely Decision Tree, Random Forest and Logistic Regression. We applied a 10-fold cross-validation, so as 90% of the traces was used to train the model and the remaining 10% was used to test the performance of the learning technique. To measure the effectiveness of the model we adopted the F1-score, Precision, and Recall measures, obtained as the average value computed over the results of each folder.

5.2.2. Trace clustering

Trace clustering, that is, splitting an event log horizontally by grouping traces based on a specific criterion of similarity, has received great attention from the process mining community and, as a matter of fact, a plethora of trace clustering approaches have been proposed in the last decade. However, as pointed out in (De Koninck et al., 2018), whether a clustering technique is using the right types of features with respect to the class label is typically not known by the user, and the similarity of the obtained clusters to the ground truth mainly depends on the clustering algorithm used. In fact, different clustering algorithms group data points according to different criteria, and they do it in an unsupervised way, which means that they do not leverage any relation between the features and the class label. Thus, we believe that the ability of a condensed representation of a process of representing its execution traces cannot be properly evaluated with a clustering task, especially if a multi-perspective condensed representation such as Deep-Trace2Vec is considered. Such kind of condensed representations contains a large amount of information, and it is difficult for clustering algorithms to extract only one particular aspect arbitrarily selected, unless a feature selection task is carried preliminarily. Feature selection is the process of selecting the subset of the most relevant features. By performing feature selection we can carry out clustering only considering those features that are more representative with respect to the target attribute, so as to obtain more significative results for our analysis.

The experiments consisted in performing a hierarchical clustering algorithm (based on Ward's minimum variance method), a k-means algorithm, and the Gaussian Mixture Models (GMM) on the vectorial representations of the execution traces. Then, we measured the extent to which the resulting cluster solution captures the same information of the ground truth using four metrics, namely the Adjusted Rand Index (ARI) (Rand, 1971), the Adjusted Mutual Information score (AMI) (Vinh, Epps, & Bailey, 2010), the Homogeneity score (HS), and the Completeness score (CS) (Hubert & Arabie, 1985). ARI and AMI are very similar indexes, they are both used to compare clusterings of the same dataset. However, according to Romano et al. (Romano, Vinh, Bailey, & Verspoor, 2016) ARI should be used when the reference clustering has large equal sized clusters, while AMI should be used when the reference clustering is unbalanced. For this reason we use ARI in the evaluation of the *BPIC*_15 dataset, and AMI in the evaluation of the *BPIC*_19 dataset.

5.3. Baseline

For both the classification and the clustering tasks we compared the results obtained with our method with those obtained with three existing embedding techniques. Such techniques were chosen as the most suitable for the purpose of representing process execution traces and the ability of representing multidimensional data. Below we provide a brief description of such techniques.

Trace2Vec (De Koninck et al., 2018): is a representation of traces based on the trace identifier, which allows for joint learning of representations of activities and traces. The embedding is based on the Distributed Memory Model of Paragraph Vectors (PVDM) approach, which can be regarded as an extension of CBOW (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), a model architecture used in the natural language processing domain. Here, exploiting the analogy between activities and words, traces can be regarded as sentences as well. However, this method is not suitable for the representation of multidimensional data as it does take into account traces and event identifiers only, disregarding traces and events attributes. Nevertheless, as far as we know, Trace2Vec is the only unsupervised learning method for representing process executions, and as such, we have chosen to adopt it as a baseline for the comparison.

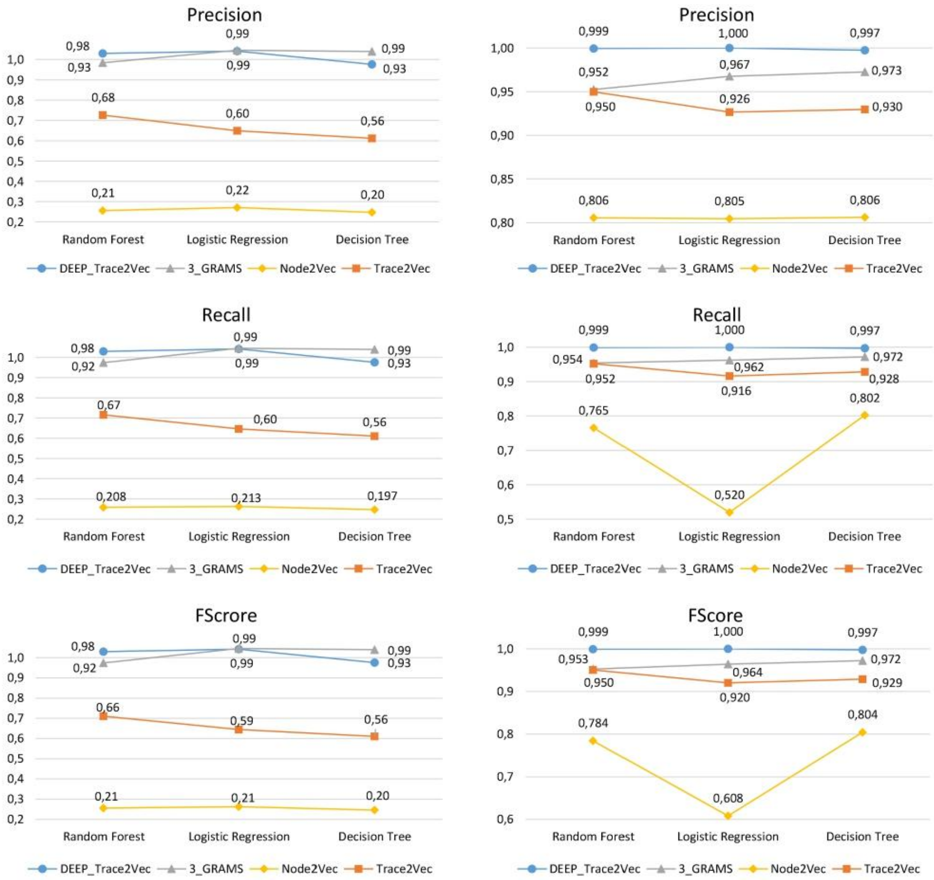
node2Vec (Grover & Leskovec, 2016): is a feature representation method which learns a mapping of nodes to a $n-$*dimensional* space of features and maximizes the likelihood of preserving network neighborhoods of nodes. An event-log can be appropriately represented as a graph using vertices as features and edges as significant relation between them. In particular, we represented an event log as a directed graph with three types of vertices, namely trace vertices v_t, event vertices v_e, and attribute vertices v_a, and with a directed edge from v_t to v_e if e is the first event of trace t, from v_{e1} to v_{e2} if event $e1$ precedes event $e2$ (and both events pertain to the same trace), from v_t to v_a if a is an attribute value of trace t, and from v_e to v_a if a is an attribute value of event e. According to this mapping, traces (resp. events) with same attribute values are mapped as vertices in the graph linked with common attribute vertices. This kind of representation enables the representation of multidimensional data as both trace and event attributes are taken into account. Graph representation has been already used in the field of text mining (Sonawane & Kulkarni, 2014), and for activity detection and prediction in events logs (Augustyniak & Ślusarczyk, 2018), due to its ability of preserving relationships and structural information effectively.

n-grams: are widely used in computational linguistics (e.g. statistical natural language processing (Manning & Schütze, 1999)), computational biology (e.g. DNA sequence analysis (Kim & Shawe-Taylor, 1994)), data compression (Witten, Bray, Mahoui, & Teahan, 1999), information retrieval (Baeza-Yates, 1992), and business process management (Horita, Hirayama, Hayase, Tahara, & Ohsuga, 2016; Wombacher, 2006). The authors of (Horita et al., 2016) proposed the usage of n-grams for a trace classification task. In particular, traces are converted in 2-grams vectors, so as 2-grams are the features representing event orders in each trace, and used as input to a decision tree algorithm for trace classification. We took inspiration from this work, and, besides the sequences of events, we applied the same principle to the sequences of trace's and events' attributes, so as to represent multidimensionality.

5.4. Implementation details

All the experiments have been carried out in a computer with an Intel Xeon E5-2620 CPU. This CPU has 6 cores running at 2.00 GHz and 12 threads. The computer also has 102 GB of RAM memory and a Nvidia GeForce GTX 1080 Ti GPU with 11 GB of memory.

Deep-Taace2Vec was implemented using Python and the neural network API Keras (Chollet et al., 2015) in conjunction with Tensorflow (Abadi et al., 2016). For the training regimen we used 100 epochs, the batch sizes where selected by taking into account the sizes of the utilized



(a) (b)

Fig. 4. Evaluation metrics values of the classification task for the *BPIC*_15 (a) and the *BPIC*_19 (b) datasets.

datasets, with the objective to utilize the maximum amount of available memory and, at the same time, minimize computation times. Thus the selected batch size depends on the dataset and was set to 400 and 1000 for *BPIC*_15 and *BPIC*_19 respectively. Each individual feature was trained using a embedding size of 4. The embeddings obtained as the result of the auto-encoder are numeric vectors of size 100. The code is publicly available.[5].

For Trace2Vec and node2Vec we used the Python implementations and default parameters provided by the authors of (De Koninck et al., 2018 & Grover & Leskovec, 2016), respectively.

For n-grams we adopted the `CountVectorizer` class of the `sklearn` Python library in order to compute the so called *document-term matrix*, where documents are the trace identifiers, and terms are the n-grams computed over the sequences of events and sequences of event attributes.

5.5. Results

The results of the classification task are represented in Fig. 4. First of all, we notice that Deep-Trace2Vec provides very high values of precision, recall, and f-score, outperforming or equalizing the other approaches for every classification algorithm and for both datasets, except for n-grams when the classification is performed by the decision tree algorithm.[6] Furthermore, the gap between Deep-Trace2Vec and Trace2-Vec, which to date is the only unsupervised learning method for representing process executions, is significantly high for *BPIC*_15 (on average 35 percentage point higher), and 2 percentage point circa higher for *BPIC*_19.

The results of clustering methods are illustrated in Fig. 5. Each point visualizes the similarity of the resulting clustering to the ground truth.

[5] https://github.com/edoardoserra/DeepTrace2Vec

[6] For n-grams we only show results with $n = 3$ as we empirically found that for different values of n performances do not change significantly.

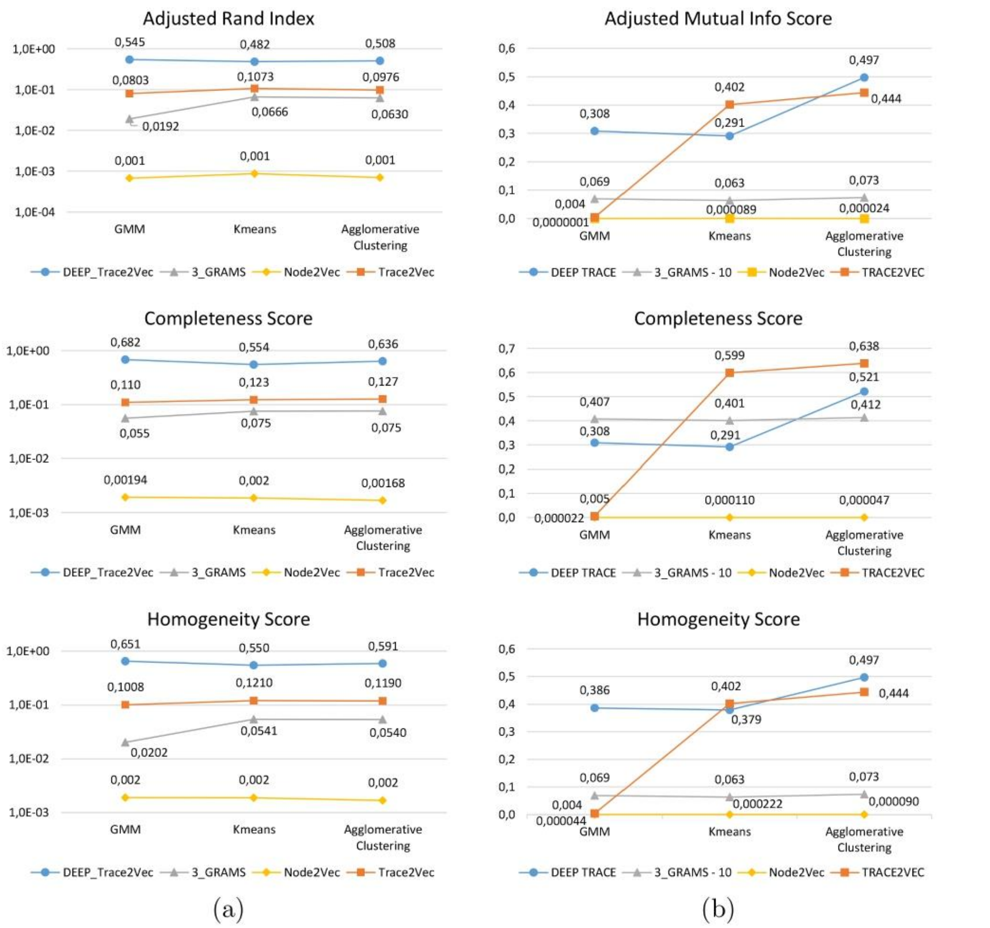


Fig. 5. Evaluation metrics values of the clustering task for the *BPIC*_15 (a) and the *BPIC*_19 (b) datasets.

Before computing the clusters we applied a feature selection task to both datasets, selecting 20 features for *BPIC*_15, and 10 features for *BPIC*_19. We can notice that for *BPIC*_15 Deep-Trace2Vec performs much better than other techniques as it provides metrics values one or two order of magnitude higher, while for *BPIC*_19, although not always Deep-Trace2Vec is the best choice, it still provides best values of AMI and HS with 2 out of 3 clustering algorithms.

These results are further supported by the visualization of the embedding spaces depicted in Figs. 6–8. To visualize the spaces we took advantage of the Principal Component Analysis (PCA) (Jolliffe, 2011), a well known dimensionality reduction technique which identifies the subspace in which the data approximately lies, allowing us to properly visualize the original 100 dimensional embedding space to a more convenient 2 dimensional space. Figs. 6 and 8 show the PCA plots obtained with the Deep-Trace2Vec technique before and after the feature selection task. We can observe how the feature selection task has noticeably improved the representation capabilities of the embeddings for both datasets. In fact, the plots on the left hand side of Figs. 6 and 8 show a set of points forming more defined clusters with respect to those obtained without feature selection. Figs. 7 and 9 show the PCA plots obtained after the feature selection task with the embeddings computed by the three approaches we compare to. From these plots it is clear that these embeddings form more overlapping clusters with respect to those obtained with Deep-Trace2Vec. This, (plus Figs. 4 and 5) confirms that Deep-Trace2Vec generates embeddings with higher representational capabilities than those provided by other techniques.

To conclude, we conducted an ablation study of Deep-Trace2Vec in order to answer to **RQ2**. In particular, we compared Deep-Trace2Vec with a modified version that does not include multiple data dimensions. In other words, it takes into account the control flow perspective only, that is, for each trace, neither trace nor event attributes are fed into the framework but the sequence of events only. It can be noticed that Deep-Trace2Vec drastically improves the performances of both the classification and the clustering tasks for both datasets. This result (see Figs. 10 and 11) confirms that taking into account the multiple process perspectives is fundamental when representation learning is applied in the

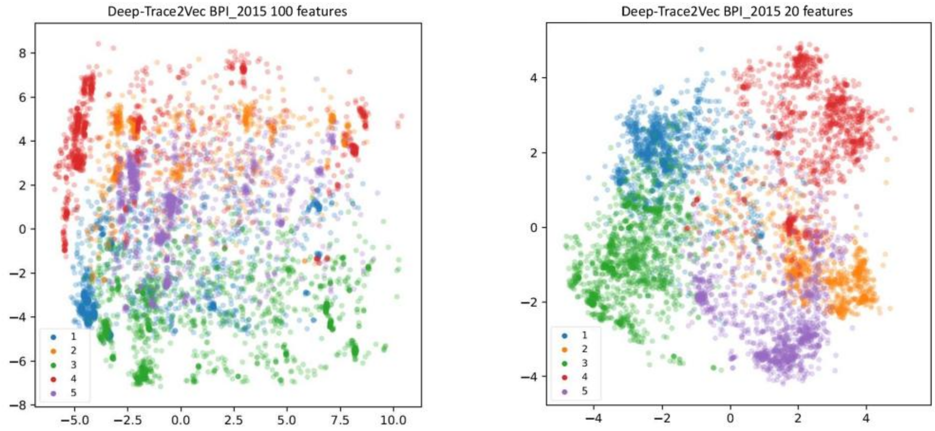


Fig. 6. PCA plots of the Deep-Trace2Vec embeddings before (left) and after (right) the feature selection task for the *BPIC*_15 dataset.

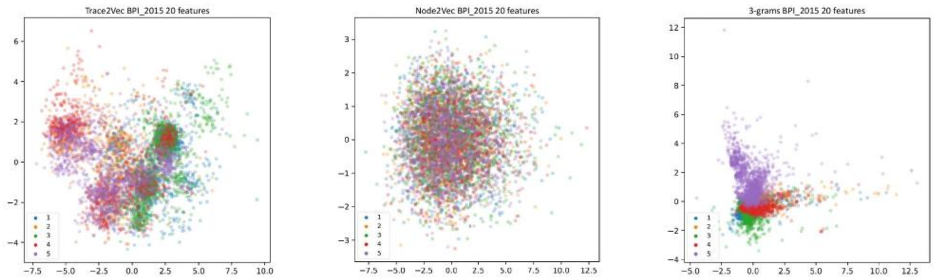


Fig. 7. PCA plots of the Trace2Vec (left), Node2Vec (middle) and 3-grams (right) embeddings after the feature selection task for the *BPIC*_15 dataset.

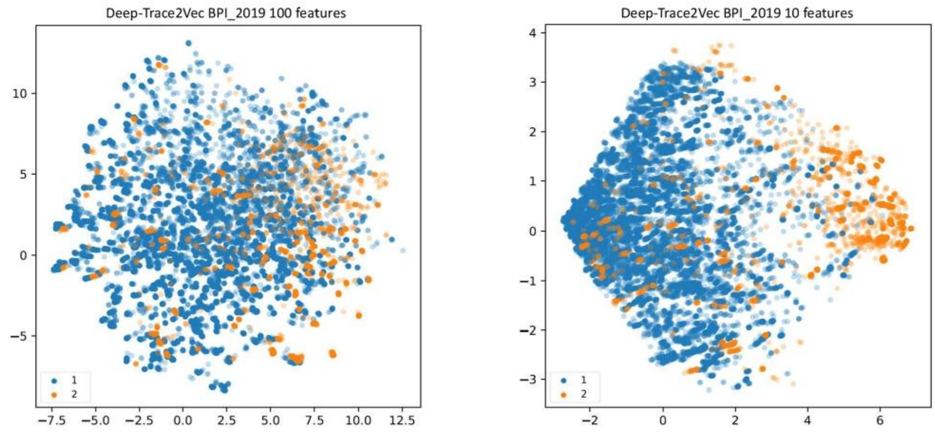


Fig. 8. PCA plots of the Deep-Trace2Vec embeddings before (left) and after (right) the feature selection task for the *BPIC*_19 dataset.

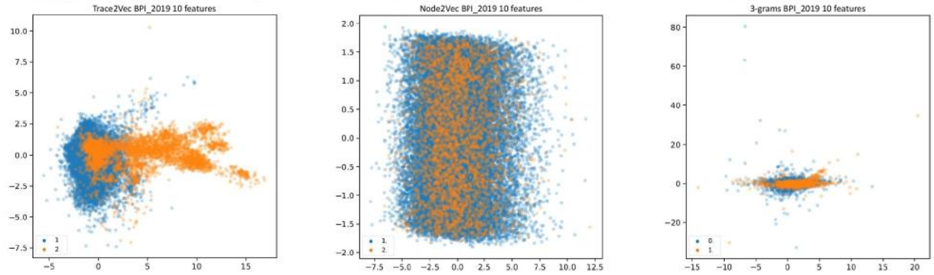


Fig. 9. PCA plots of the Trace2Vec (left), Node2Vec (middle) and 3-grams (right) embeddings after the feature selection task for the *BPIC*_19 dataset.

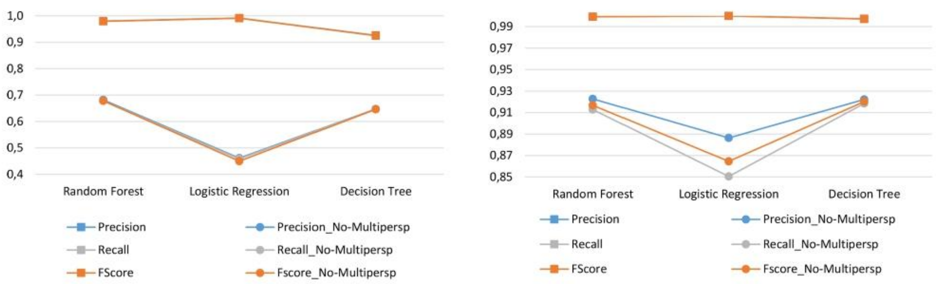


Fig. 10. Comparison of Deep-Trace2Vec performances with and w/o multi-perspective representation for the classification task applied at the *BPIC*_15 (left) and the *BPIC*_19 (right) datasets.

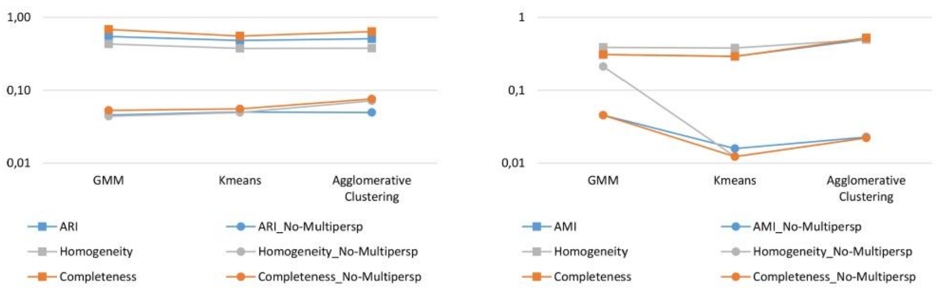


Fig. 11. Comparison of Deep-Trace2Vec performances with and w/o multi-perspective representation for the clustering task applied at the *BPIC*_15 (left) and the *BPIC*_19 (right) datasets.

field of process mining, as it allows to get a deep insight into the process behavior. In fact, for their very nature process executions are a heterogeneous collection of attributes, that confer the embeddings higher representational capabilities compared when such attributes are omitted from the computation.

5.6. Conclusive discussion

The main results of Section 5.5 are summarized below to provide a clear answer to the research questions defined at the beginning of this section.

RQ1 Our embedding in comparison with the state-of-art techniques usually performs better in both trace classification and trace clustering tasks. In particular, Deep-Trace2Vec significantly outperforms all the baselines in the classification task when tested with *BPIC*_19, and in the clustering task when tested with *BPIC*_15, while it results to be comparable with the n-gram-based approach in the classification task when tested with *BPIC*_15, and with Trace2Vec in the clustering task when tested with *BPIC*_19. In addition, from the qualitative visualization analysis shown in Figs. 6–8, we observe that Deep-Trace2Vec produces the most homogeneous and separable regions characterizing the classes provided in the ground truth, meaning that our embedding technique has more representative power than the other ones.

RQ2 We demonstrated that taking into account the multiple perspectives of the process increases the representative power of the trace embeddings. This is revealed by the ablation study conducted in Section 5.5 where Deep-Trace2Vec drastically outperforms the single-perspective approach for both the trace classification and trace clustering tasks. We can thus conclude that modeling the multi-dimensional nature of processes allows a better representation of the processes behavior.

6. Conclusions

Nowadays, there is an enormous amount of data storing process executions containing high-dimensional and heterogenous information. This data can be formatted as event log and analyzed with process mining techniques in order to extract useful knowledge for the analysis of the multiple and various characteristics of processes. However, current literature primarily focuses on the control-flow perspective, or in a better case considers additional perspectives separately. Moreover, there is very little prior process mining research on exploiting neural network-based embedding. To overcome these issues, in this paper we have addressed the problem of computing a concise and informative representation of execution traces that includes the multiple perspectives of the process behavior. We proposed Deep-TRace2Vec, a deep learning-based approach that is able to encode complex process behavior in the form of a vectorial representation of the traces. Deep-TRace2Vec is able to learn robust and compact representations automatically in an usupervised way, that is, it does not rely on any prior knowledge about the process. We used several deep recurrent multilayer neural networks in order to embed the multidimensional characteristic of traces along with their connections. We evaluated our embedding with two real world datasets taken from the BPI Challenge 2015 and the BPI Challenge 2019 competitions. Results have shown that Deep-TRace2Vec compared with the state of art techniques, provides a richer and more informative representation space that is better suited in trace classification and clustering tasks. Moreover, the performed ablation study showed how the multi-perspective feature that also includes the temporal dimension of a trace, drastically improves the performance in both the trace classification and clustering tasks.

As future work we aim at extending the proposed model in order to obtain good performances also when used for different data mining tasks like anomaly detection and novelty detection. Moreover, we will also seek to investigate how effectively combine dimensionality reduction of the embedding architecture with surrogate modelling (Alizadeh et al., 2019; Alizadeh, Allen, & Mistree, 2020; Jia et al., 2020), with the aim of improving the performances of simulation tasks for engineering design. We will also investigate extensions of our approach with transformer neural networks (Devlin, Chang, Lee, & Toutanova, 2018) that are another suitable processing unit for sequential and temporal data.

CRediT authorship contribution statement

Antonella Guzzo: Conceptualization, Methodology, Writing - review & editing. **Antonino Rullo:** Software, Methodology, Writing - review & editing. **Edoardo Serra:** Conceptualization, Methodology, Writing - review & editing. **Mikel Joaristi:** Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* (pp. 265–283).

Alizadeh, R., Allen, J. K., & Mistree, F. (2020). Managing computational complexity using surrogate models: a critical review. *Research in Engineering Design, 31*(3), 275–298.

Alizadeh, R., Jia, L., Nellippallil, A. B., Wang, G., Hao, J., Allen, J. K., & Mistree, F. (2019). Ensemble of surrogates and cross-validation for rapid and accurate predictions using small data sets. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AI EDAM, 33*(4), 484–501.

Augustyniak, P., & Ślusarczyk, G. (2018). Graph-based representation of behavior in detection and prediction of daily living activities. *Computers in Biology and Medicine, 95*, 261–270.

Baeza-Yates, R. A. (1992). Text-retrieval: Theory and practice. In *IFIP congress (1)*, Vol. 12, pp. 465–476.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. CoRR abs/1409.0473.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(8), 1798–1828.

Bezerra, F., Wainer, J., & van der Aalst, W. M. (2009). Anomaly detection using process mining. In *Enterprise, business-process and information systems modeling* (pp. 149–161). Springer.

Böhmer, K., & Rinderle-Ma, S. (2017). Multi instance anomaly detection in business process executions, in *International conference on business process management* (pp. 77–93). Springer.

Böhmer, K., & Rinderle-Ma, S. (2016). Multi-perspective anomaly detection in business process execution events. In *OTM confederated international conferences on the move to meaningful internet systems* (pp. 80–98). Springer.

Bose, R. P. J. C., & van der Aalst, W. M. P. (2009). Context aware trace clustering: Towards improving process mining results. In *Proc of the SIAM international conference on data mining (SDM 2009)* (pp. 401–412).

Chollet, F., et al. (2015). *Keras*. URL: https://github.com/fchollet/keras.

Chorowski, J. K., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y. (2015). Attention-based models for speech recognition. In *Advances in neural information processing systems* (pp. 577–585).

De Koninck, P., vanden Broucke, S., & De Weerdt, J. (2018). act2vec, trace2vec, log2vec, and model2vec: Representation learning for business processes. In International conference on business process management. (pp. 305–321). Springer.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv: 1810.04805.

De Weerdt, J., Schupp, A., Vanderloock, A., & Baesens, B. (2013). Process mining for the multi-faceted analysis of business processes—a case study in a financial services organization. *Computers in Industry, 64*(1), 57–67.

dos Santos Garcia, C., Meincheim, A., Faria Junior, E. R., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., Santos, E. A. P., & Scalabrin, E. E. (2019). Process mining techniques and applications – a systematic mapping study. *Expert Systems with Applications, 133*, 260–295. URL: https://www.sciencedirect.com/science/article/pii/S0957417419303161.

Evermann, J., Rehse, J., & Fettke, P. (2017). Predicting process behaviour using deep learning. *Decision Support Systems, 100*, 129–140.

Folino, F., Guarascio, M., & Pontieri, L. (2012). Context-aware predictions on business processes: An ensemble-based solution. *New frontiers in mining complex patterns – first international workshop, NFMCP 2012 revised selected papers*, 215–229.

Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems, 151*, 78–94.

Grover, A., & Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. In *Proceedings of the 22Nd ACM SIGKDD international conference on knowledge discovery and data mining, KDD '16* (pp. 855–864). ACM.

Horita, H., Hirayama, H., Hayase, T., Tahara, Y., & Ohsuga, A. (2016). Business process verification and restructuring ltl formula based on machine learning approach. In *Computer and information science* (pp. 89–102). Springer.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(1), 193–218.

Jans, M., Van Der Werf, J. M., Lybaert, N., & Vanhoof, K. (2011). A business process mining application for internal transaction fraud mitigation. *Expert Systems with Applications, 38*(10), 13351–13359.

Jia, L., Alizadeh, R., Hao, J., Wang, G., Allen, J. K., & Mistree, F. (2020). A rule-based method for automated surrogate model selection. *Advanced Engineering Informatics, 45*, Article 101123.

Jolliffe, I. (2011). *Principal component analysis*. Springer.

Kim, J. Y., & Shawe-Taylor, J. (1994). Fast string matching using an n-gram algorithm. *Software: Practice and Experience, 24*(1), 79–88.

Luca, S. E., Pimentel, M. A., Watkinson, P. J., & Clifton, D. A. (2018). Point process models for novelty detection on spatial point patterns and their extremes. *Computational Statistics and Data Analysis, 125*, 86–103.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA, USA: MIT Press.

Mehdiyev, N., Evermann, J., & Fettke, P. (2018). A novel business process prediction model using a deep learning method. *Business & Information Systems Engineering*, 1–15.

Meng, Q., Catchpoole, D., Skillicom, D., & Kennedy, P. J. (2017). Relational autoencoder for feature extraction. In *in '2017 International Joint Conference on Neural Networks (IJCNN)', IEEE* (pp. 364–371). IEEE.

Mhaskar, H., & Poggio, T. A. (2016). *Deep vs. shallow networks: An approximation theory perspective*. CoRR abs/1608.03287.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality*. CoRR abs/1310.4546.

Navarin, N., Vincenzi, B., Polato, M., & Sperduti, A. (2017). Lstm networks for data-aware remaining time prediction of business process instances. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1–7).

Nolle, T., Luettgen, S., Seeliger, A., & Mühlhäuser, M. (2018). Analyzing business process anomalies using autoencoders. *Machine Learning* (pp. 1–19).

Nolle, T., Seeliger, A., & Mühlhäuser, M. (2018). Binet: multivariate business process anomaly detection using deep learning. In *International conference on business process management* (pp. 271–287). Springer.

Okoye Dr, K., Naeem Dr, U., Islam Dr, S., Tawil Dr, A.-R. H., & Lamine Dr, E. (2017). Process models discovery and traces classification: A fuzzy-bpmn mining approach. *Journal of International Technology and Information Management, 26*(4), 2–50.

Polato M., Sperduti A., Burattin, A., & de Leoni, M. (2018). Time and activity sequence prediction of business process instances. *Computing 100*(9), 1005–1031.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association, 66*(336), 846–850.

Romano, S., Vinh, N. X., Bailey, J., & Verspoor, K. (2016). Adjusting for chance clustering comparison measures. *The Journal of Machine Learning Research, 17*(1), 4635–4666.

Rullo, A., Guzzo, A., Serra, E., & Tirrito, E. (2020). A framework for the multi-modal analysis of novel behavior in business processes. In *International conference on intelligent data engineering and automated learning* (pp. 51–63). Springer.

Sonawane, S. S., & Kulkarni, P. A. (2014). Graph based representation and analysis of text document: A survey of techniques. *International Journal of Computer Applications, 96*(19).

Staudemeyer, R. C., & Morris, E. R. (2019). *Understanding lstm–a tutorial into long short-term memory recurrent neural networks*. arXiv preprint arXiv:1909.09586.

Tavares, G. M., Barbon Junior, S., Ceravolo, P., & Damiani, E. (2020). Evaluating trace encoding methods in process mining. In *9th International symposium DATAMOD 2020*. Springer.

van der Aalst, W. M. P. (2016). *Process mining – Data science in action* (2nd Ed.). Springer.

Van Dongen, B.F., & Boudewijn (2015). BPI Challenge 2015, Eindhoven University of Technology. URL: https://data.4tu.nl/repository/uuid:31a308ef-c844-48da-948c-305d167a0ec1 [DATASET].

Van Dongen, B.F. Boudewijn (2019). BPI Challenge 2019, 4TU.Centre for Research Data. URL: https://data.4tu.nl/repository/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1 [DATASET] .

Van Dongen, B. F., & Van der Aalst, W. M. (2004). Multi-phase process mining: Building instance graphs. In *International conference on conceptual modeling* (pp. 362–376). Springer.

van Eck, M. L., Sidorova, N., & van der Aalst, W. M. P. (2016). Discovering and exploring state-based models for multi-perspective processes. In M. L. Rosa, P. Loos, & O. Pastor (Eds.), *Business process management – 14th international conference, BPM 2016, Rio de Janeiro, Brazil, September 18–22, 2016. Proceedings', Vol. 9850 of Lecture notes in computer science* (pp. 142–157). Springer.

Vinh, N. X., Epps, J., & Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research, 11*(Oct), 2837–2854.

Witten, I. H., Bray, Z., Mahoui, M., & Teahan, W. J. (1999). Text mining: A new frontier for lossless compression. In *Data compression conference, DCC 1999, Snowbird, Utah, USA, March 29–31, 1999* (pp. 198–207). doi: 10.1109/DCC.1999.755669.

Wombacher, A. (2006). Evaluation of technical measures for workflow similarity based on a pilot study. In *OTM confederated international conferences on the move to meaningful internet systems* (pp. 255–272). Springer.

Yang, W.-S., & Hwang, S.-Y. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert Systems with Applications, 31*(1), 56–68.

ANEXO 17: GUZZO, JOARISTI, & ANTONIO RULLO, 2021 - A MULTI-PERSPECTIVE APPROACH FOR THE ANALYSIS OF COMPLEX BUSINESS PROCESSES BEHAVIOR

Expert Systems With Applications 177 (2021) 114934

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A multi-perspective approach for the analysis of complex business processes behavior

Antonella Guzzo [a,*,1], Mikel Joaristi [b,4], Antonino Rullo [a,2], Edoardo Serra [b,3]

[a] University of Calabria, via Pietro Bucci, 87036 Rende, Italy
[b] Boise State University, 777 W Main St, 83702 Boise, ID, USA

ARTICLE INFO

Keywords:
Process mining
Trace embedding
Multi-perspective analysis
Deep learning

ABSTRACT

Business processes are often monitored by transactional information systems that produce massive dataset called event logs. Such logs contain the process execution traces, typically characterized by heterogeneous and high-dimensional data. Process mining techniques offer a great opportunity to gain valuable knowledge hidden in the data to be used for analysing the multiple characteristics of processes (i.e. perspectives in process mining, like structural aspects, activities, resources, data and time). Therefore, raw data must be encoded into a suitable format that can be more conveniently provided to the mining algorithms. However, most of the existing process encoding techniques focus on the control-flow perspective, i.e. only encode the sequence of activities that characterize a trace, leaving out other process perspectives that are fundamental for describing the process behavior in all its aspects. In this paper we address the problem of computing a concise and informative representation of execution traces that considers the multiple perspectives of the process behavior. We propose a holistic approach that computes trace embedding able to capture patterns of dependencies between the perspectives that are lost in a one-dimensional analysis and, at the same time, it is unsupervised, meaning that no a priori knowledge is needed. The experiments conducted on two real life logs demonstrate that our proposed embedding is appropriate to concisely describe the multiple and various characteristics of the processes, and that the proposed method outperforms existing trace encoding techniques. Furthermore, the embedding includes the elapsed time between events as an additional feature to make us capable to use it as a further dimension of analysis.

1. Introduction

Process mining is a research field aiming at developing methodologies and techniques to discover and improve business processes on the basis of the knowledge that can be mined from the data collected in the logs of transaction information systems (van der Aalst, 2016; dos Santos Garcia et al., 2019). In real-life applications, such logs are given as a set of process executions, called *traces* or event log, that contains high-dimensional and heterogeneous information as for example activities performed, activity executors (e.g. employers, systems, departments, etc.), timestamp and data in general. These information can be analysed along with three process behavior perspectives:

- the *control-flow* perspective, which focuses on the activities that are executed and the relationships of precedence among them (in terms of preconditions and postconditions);
- the *organizational* perspective, which focuses on the actors that are involved, on their roles, and on how they are mapped/assigned to the activities;
- the *data* perspective, which focuses on the data used and generated during the process execution, including the temporal data, the

* Corresponding author.
E-mail addresses: antonella.guzzo@unical.it (A. Guzzo), mikeljoaristi@boisestate.edu (M. Joaristi), n.rullo@dimes.unical.it (A. Rullo), edoardoserra@boisestate.edu (E. Serra).
[1] ORCID ID: https://orcid.org/0000-0003-3159-0536
[2] ORCID ID: https://orcid.org/0000-0002-6030-0027
[3] ORCID ID: https://orcid.org/0000-0003-0689-5063
[4] ORCID ID: https://orcid.org/0000-0001-9878-8366

https://doi.org/10.1016/j.eswa.2021.114934
Received 1 December 2020; Received in revised form 24 February 2021; Accepted 20 March 2021
Available online 29 March 2021
0957-4174/

attributes that characterize the given enactment, and the attributes that characterize the activities and the triggered events.

Effective process mining methods must be capable to simultaneously look at all the process' perspectives, along with their associated information, and at their cross-correlations in a holistic way. However, most of the process mining techniques focus only on the control-flow perspective (Bezerra, Wainer, & van der Aalst, 2009; De Koninck, vanden Broucke, & De Weerdt, 2018; Yang & Hwang, 2006; Jans, Van Der Werf, Lybaert, & Vanhoof, 2011) (i.e. the sequence of the activities) or look into different perspectives to each one. Recently, multivariate-based techniques (Böhmer & Rinderle-Ma, 2016; Nolle, Seeliger, & Mühlhäuser, 2018; Rullo, Guzzo, Serra, & Tirrito, 2020; van Eck, Sidorova, & van der Aalst, 2016; De Weerdt, Schupp, Vanderloock, & Baesens, 2013) have been proposed. In these works more than one perspective is taken into account for the process discovery task, and in some cases the *time* perspective is considered as well, as a further data dimension, but separately from the others. However, these techniques *(i)* do not consider all the perspectives at once, leaving out important aspects of the behavior of the process under analysis; *(ii)* only consider the time instant in which events occur, leaving out the time elapsed between events which is a useful feature for describing the process behavior on the basis of the rate at which events occur.

Another important aspect when dealing with learning tasks of multidimensional data is the format the data is represented with. In order for mining algorithms to work with heterogeneous data, it is crucial that the process logs provided as input are represented in a way that is *(i)* appropriate to concisely and accurately describe the heterogeneous characteristics of the process behavior, and *(ii)* useful to build classifiers and other predictive/descriptive models (Bengio, Courville, & Vincent, 2013).

Representation learning plays therefore a key role in process mining—much more than it happens in other application domains for mining/learning methods.

Actually, most of the representation learning methods proposed for business processes come in the form of *feature selection* algorithms, which either look at a single perspective at a time (Böhmer & Rinderle-Ma, 2017; Folino, Guarascio, & Pontieri, 2012), or compute "aggregate" features by just concatenating the features extracted over the different perspectives (Bose & van der Aalst, 2009; Luca, Pimentel, Watkinson, & Clifton, 2018; Polato et al., 2018; Rullo et al., 2020). However, it has been noticed that these approaches are not the best choice when dealing with high-dimensional and multivariate data (Mhaskar & Poggio, 2016).

De Koninck et al. (2018) proposed a method that overcomes these drawbacks. Their approach does not focus on the extraction of specific features, but rather computes an *embedding* of the input data into a novel vectorial space— hereinafter we refer to this kind of approach as a *trace embedding* method. Even though this method deals with the control-flow perspective only, it has been proven to be quite effective when used for pre-processing data in inference tasks such as trace clustering, process comparison, predictive process monitoring and process anomaly detection. However, the authors left open *(i)* to develop embedding methods that simultaneously take into account the various process perspectives and *(ii)* to investigate more elaborated architectures in order to improve the performances of the inference tasks.

1.1. Contributions

To address the above research issues, we proposed a holistic method for analysing process behavior aiming at capturing patterns of dependencies between perspectives that are lost in one-dimensional analysis. Moreover, our method is tailored to time analysis by expliciting elapsed time between events as further dimension of analysis. We compute a *multi-perspective representation* of the traces by means a neural network architecture, enriched with an Long Short Time Memory (LSTM) attention mechanism and a time feature, thus avoiding the expensive feature engineering phase task.

LSTMs have proven to be useful for learning sequences containing long-term patterns of unknown length due to their ability to maintain long term memory, which gives them an advantage over alternative approaches like Recurrent Neural Networks and Hidden Markov Models. Our contributions can be summarized as follows:

- we propose a deep learning based framework to find a concise and informative representation of the process behavior by automatically integrating the multidimensional aspects of the traces.
- our approach is based on a LSTM-autoencoder model, Deep-Trace2-Vec, which is able to learn the representation directly from raw temporal data in an unsupervised way, that is, without the need of any a priori knowledge and thus avoiding the expensive feature engineering phase task;
- we include in the input log the elapsed time between events so as to embed useful information about the "speed" at which traces are executed; this allows to explain the process behavior also on the basis of the rate at which events occur;
- we empirically validate the proposed approach over two datasets consisting of real data. In particular, we use the datasets adopted in the BPI Challenge 2015 (Van Dongen & Boudewijn, 2015) and in the BPI Challenge 2019 (Van Dongen & Boudewijn, 2019) competitions. Our results show that the performances of the most relevant process mining methods (clustering and classification) drastically improve w.r.t. the performances that can be obtained with embedding methods earlier proposed in literature.

The rest of the paper is organized as follows: a critical evaluation of literature is provided in Section 2; some preliminary notions are illustrated in Section 3; the multi-perspective embedding framework is elaborated in Section 4; results of our experimental activity are illustrated in Section 5; finally, concluding remarks with avenues for further research are discussed in Section 6.

2. Related work

Representation learning is the task of finding a condensed representation of a given amount of data which guarantees that characteristics relevant for further analysis are preserved after the dimensionality of data has been reduced. In the case of processes, this is a challenging task since real-life event logs contain a large number of cases (traces), potentially representing a highly varied set of distinct event sequences, and usually, also containing additional information on resources and a diverse set of other event or case-related attributes.

There exists a considerable amount of encoding techniques that can be applied to event logs (Tavares, Barbon Junior, Ceravolo, & Damiani, 2020), here we evaluate the most used in the process mining domain.

Graph embedding methods (Goyal & Ferrara, 2018) can encode information structured as graphs, thus, an event-log can be appropriately represented as a graph using vertices as activities/attributes and edges as relations between them. Van Vanongen and Van der Aalst (2004) proposed to represent control-flow instances as graphs with the aim of performing mining tasks on graph instances instead of constructing a process model. Böhmer and Rinderle-Ma (2016) proposed a multi-perspective approach with a probabilistic graph-based model. The graph contains three types of vertices, activities, resources and weekdays, and edges are labeled with the probability that, given the previous triple activity-resource-weekday, a certain activity happens next. The graph is then used to score ongoing traces with the purpose of finding anomalous process executions.

Text-based encodings, grounded in information retrieval and natural language processing, have been used in the context of process mining in various forms due to the analogy between activities and words, and between traces and sentences. The *integer encoding* maps all possible attribute values for each attribute to a unique positive integer. Nolle

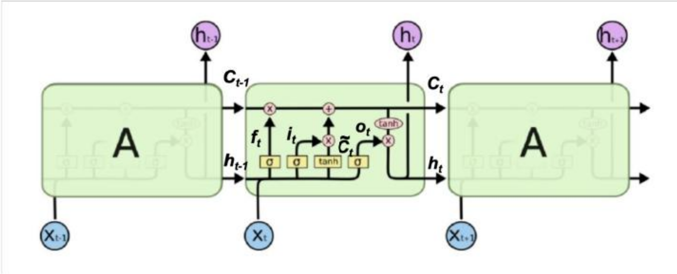


Fig. 1. An LSTM cell C, source (Staudemeyer & Morris, 2019).

et al. (2018) adopted this encoding technique to preprocess data for anomaly detection purposes. The *count vectorizer* encode words/activities by accounting their frequencies in a corpus/trace. The *one-hot encoding* transforms a categorical attribute which shows *m* different values into *m* binary features, each observation in the categorical feature is thus converted into a vector of size *m* with only one of the values as 1. The *n-gram encoding* has the ability to represent relationships between neighboring elements of a sequence by representing a trace as all contiguous subsequences of length *n*. Mehdiyev, Evermann, and Fettke (2018) applied deep feed-forward multilayer neural networks to predict the next business process event, and encoded traces adopting the n-gram encoding. Rullo et al. (2020) adopted a hybrid approach for the detection of anomalous traces, in which the trace embedding is the concatenation of the vector obtained by one-hot encoding the categorical trace attributes, and the vector obtained by count-vectorizing the 2-grams of the activities' sequence. Still from the field of natural language processing, *word embeddings* are the result of a two-layer neural network built to reconstruct the context of a word in a corpus, so as words occurring in similar contexts generate similar encodings. De Koninck et al. (2018) applied this technique in the field of process mining, and published a landmark paper where general purpose representations for activities (called *act2vec*), traces (called *trace2Vec*), event logs (called *log2vec*) and process models (called *model2vec*) have been proposed. The paper has the merit to present a novel neural network-based embedding technique specified for process concepts and to show how it can be efficiently used for automated inference tasks such as trace clustering, process comparison, predictive process monitoring, and anomaly detection. However, the authors left open to develop embedding methods that simultaneously take into account the various process perspectives, and to investigate more elaborated architectures in order to improve the performances of the inference tasks.

The major drawbacks of the aforementioned techniques are mainly two: *(i)* they do not consider all the process perspective at once, and *(ii)* temporal data is considered only as time points, leaving out time intervals (i.e. time spans between events) that are fundamental for explaining the temporal behavior of processes. On the contrary, our paper computes a trace embedding aiming at preserving the multi-perspectives of processes by exploiting LSTM architecture and autoencoder. In fact, LSTMs have proven to be useful for learning sequences containing long-term patterns of unknown length due to their ability to maintain long term memory. As process' executions are temporal sequences, we think the LSTM be the best choice to build embeddings that maintain information about the temporal relationships between events. The LSTM is the architecture used as prediction of process model in (Navarin, Vincenzi, Polato, & Sperduti, 2017; Evermann, Rehse, & Fettke, 2017), however, differently for us, these works don't compute an embedding and only consider the sequence of the activities and their timestamps.

For completeness it is noteworthy to mention that autoencoders have been adopted in process mining literature mainly as learning tool fed with text-based embeddings with the aim of finding anomalous process executions (Nolle, Luettgen, Seeliger, & Mühlhäuser, 2018). Rather, in this paper we adopt the autoencoder for the purpose of learning highly informative trace embeddings, leaving the data mining task (e.g. classification, clustering, anomaly detection, etc.) to well established machine learning algorithms.

3. Background knowledge

3.1. Long Short-Term Memory

An LSTM is a neural network architecture tailored to deal with sequences of data (Staudemeyer & Morris, 2019). The core idea is to allow the network to "accumulate" information over some period of time, and to create paths through time having derivatives that neither vanish nor explode. An LSTM network contains LSTM units instead of, or in addition to, other standard network units. An LSTM unit remembers values for either long or short time periods, and several units are often implemented in "blocks". LSTM blocks contain three or four "gates" that control the information flow. These gates are implemented using the logistic function to compute a value between 0 and 1. Multiplication is applied to this value to partially allow or deny information to flow into or out of the memory. An LSTM block uses weights to direct gates' operations, applying them to the values that feed into the block. Thus, the LSTM block determines how to maintain its memory as a function of those values, and training its weights causes the block to learn the function that minimizes loss. LSTMs also have a form of a chain of repeating module of a neural network (chain-like structure), where each module has a different structure and different interacting layers. Fig. 1 describes the architecture of a single LSTM cell that outputs the next state h_t by taking in input the previous state h_{t-1} and the next symbol x_t. The operations done by the single LSTM cell C are described by the following equations:

$$\begin{aligned}
a_t &= \rho(W_a \cdot [h_{t-1}; x_t]) \\
b_t &= \rho(W_b \cdot [h_{t-1}; x_t]) \\
y_t &= \tanh(W_y \cdot [h_{t-1}; x_t]) \\
g_t &= \rho(W_g \cdot [h_{t-1}; x_t]) \\
c_t &= c_{t-1} \cdot a_t + b_t \cdot y_t \\
h_t &= \tanh(c_t) \cdot g_t
\end{aligned}$$

where the W_a, W_b, W_y and W_g are the weights representing the LSTM cell C and the entire LSTM neural network.

If a sequence is passed to an LSTM, the network will output, for each element of the sequence, a vectorial representation h_i representing the sub-sequence from the beginning till the element i.

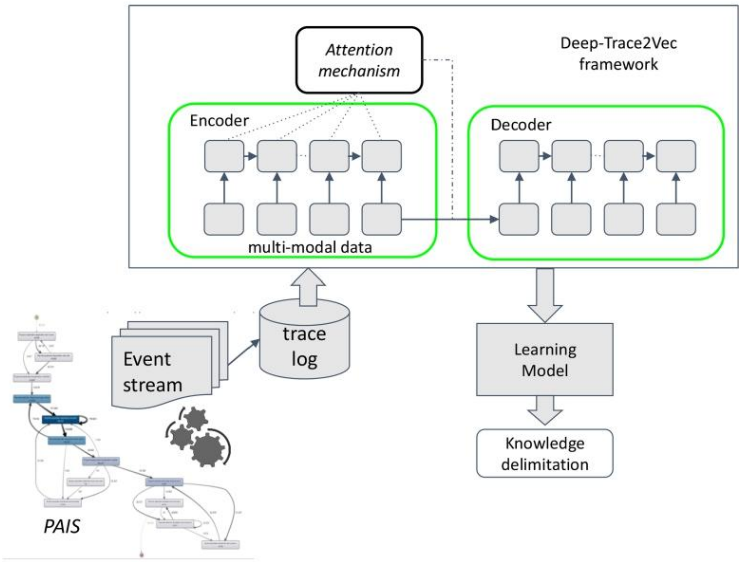


Fig. 2. Overview of the framework.

3.2. Autoencoder

An autoencoder (Meng, Catchpoole, Skillicom, & Kennedy, 2017) is a neural network that is capable to produce an encoding of the input data in an unsupervised manner, in particular by reducing its dimensionality while minimizing reconstruction loss. An autoencoder has two architectural parts: (i) the encoder, aiming at learning an encoded representation (embedding) of the training in different feature spaces by efficiently reducing the dimensionality of the original data space; (ii) the decoder, that tries to reconstruct the original data by taking the embeddings (compressed feature vectors) as input. The output of an autoencoder has the same number of computational units as the input (original feature dimensions). The intuition behind an autoencoder to be used as representation algorithm comes from its ability to effectively reconstruct the examples having similar statistical properties in the original feature space, thus obtaining the smaller reconstruction errors for the known type of objects.

Specifically, the encoder is a function f that maps an input X to a hidden representation Y as formulated by the following equation:

$$Y = f(X) = s_f(WX + b)$$

where s_f is a nonlinear activation function. The encoder is parameterized by a weight matrix W and a bias vector b. The decoder function g maps the hidden representation Y back to a reconstruction X':

$$X' = g(Y) = s_g(W'Y + b')$$

where s_g is the decoder's activation function. The decoder's parameters are a bias vector b' and matrix W'. Training an autoencoder involves finding parameters $\pi = (W, W', b, b')$ that minimize the reconstruction loss on the given dataset X. The objective function is given as

$$\Theta = \min_{\pi} L(X, X') = \min_{\pi} L(X, g(f(X)))$$

For linear reconstruction, the reconstruction loss (L1) is generally from the squared error:

$$L_1(\pi) = \sum_{i=1}^{n} ||x_i - x_i'|| = \sum_{i=1}^{n} ||x_i - g(f(x_i))||^2$$

For nonlinear reconstruction, the reconstruction loss (L2) is generally from cross-entropy:

$$L_2(\pi) = -\sum_{i=1}^{n} [x_i log(y_i) + (1 - x_i) log(1 - y_i)]$$

where $x_i \in X, x_i' \in X'$, and $y_i \in Y$.

4. Multi-perspective embedding framework

A high-level view of the multi-perspective trace embedding framework we are going to illustrate is depicted in Fig. 2. Without loss of generality, we assume that process executions are caught in a stream of event log and then represented as a bag of traces. In our framework, these traces are passed trough an autoencoder architecture that embeds the process behavior into an informative vectorial representation. The autoencoder uses some LSMT components, plus an *attention mechanism* (Chorowski, Bahdanau, Serdyuk, Cho, & Bengio, 2015) which is a technical component conceived to improve the quality of the results (and discussed in details later in this section). The embedding (that is, the representation that has been learned via the proposed framework) can be then used as input by a number of process mining methods. Our focus will be mainly to assess the gain (in terms of improvement of performances over the respective inference tasks) that our trace embeddings can provide for process clustering and classification algorithms.

The trace embedding method we have just sketched will be called Deep-Trace2Vec. Differently from earlier approaches, Deep-Trace2Vec focuses on process execution traces by integrating multidimensional data into a unique representation. Moreover, it is able to handle all types of attributes: text, categorical, ordinal, and numerical. It is a totally unsupervised algorithm, and as such, it does not need of expert knowledge. In addition, it provides a transformation function able to map a log

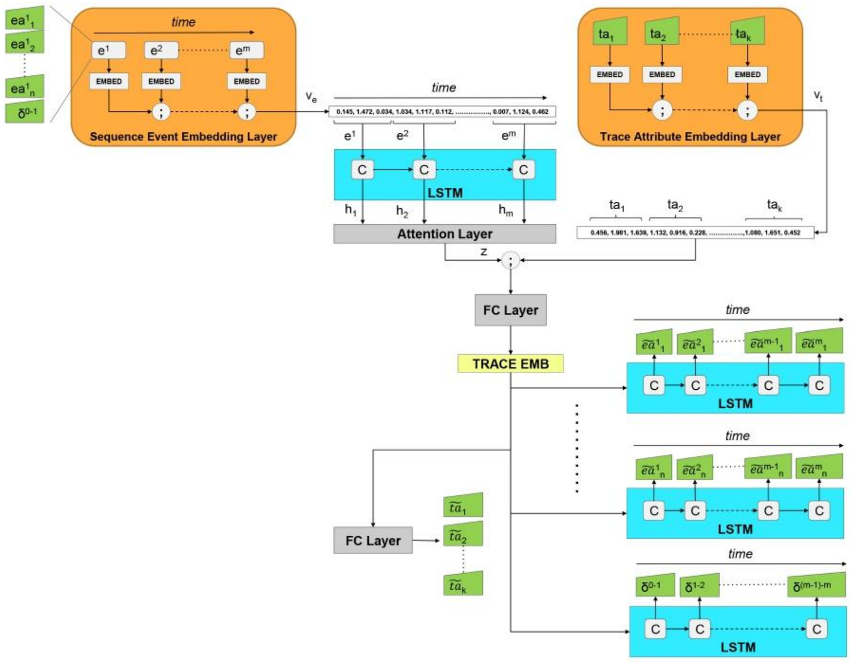


Fig. 3. Deep-Trace2Vec architecture. The arrows labeled with *time* denote the temporal relationship between events, which is preserved in the vector v_e, and in the output $< \widetilde{ea}_1^1 \ldots \widetilde{ea}_1^m > \ldots < \widetilde{ea}_n^1 \ldots \widetilde{ea}_n^m >, < \delta^{0-1} \ldots \delta^{(m-1)-m} >$.

trace into a vector without the need to re-train the model.

Each architectural component will be illustrated in the rest of this section.

4.1. Encoder

By abstracting from the specificity of the various systems, we can view a log as a set of execution traces of a specific process. An execution trace can be represented as a pair (A, S) where $A = \{ta_1, \ldots, ta_k\}$ is a set of trace attributes describing the global properties of the trace (e.g. caseID/identifier, who originate execution and system, beginning time), and $S =< e^1, \ldots, e^m >$ is a temporal ordered sequence of events mapping the sequence of activities performed in the execution's trace with their attributes. Specifically, each event $e^i \in S$ is characterized by a timestamp t^i and a set of event attributes $\{ea_1^i, \ldots, ea_n^i\}$ describing the intrinsic properties of the event (e.g. activity's name, performing resource, state of activity, input data). Given two events e^i and e^j with $i > j$, we have that $t^i \geqslant t^j$. In particular if $t^i = t^j$ then e^i and e^j are occurring in parallel. Both the trace attributes and the event attributes can be nominal (e.g. strings) or numeric.

Deep-Trace2Vec is trained by using an autoencoder structure operating a multi-perspective integration. The architecture is sketched in Fig. 3. This architecture is inspired to the autoencoder-LSTM model that has been successfully developed for natural language processing problems, and for image captioning. These models are capable of learning the complex dynamics within the temporal ordering of input sequences, and of using an internal memory to remember information across long input sequences.

Specifically, our encoder comprises two separate embeddings: a *Trace Attribute Embedding* layer, which embeds the trace's attributes $ta_1, \ldots, ta_k$ into a numeric vector (denoted as v_t in Fig. 3); and a *Sequence Event Embedding* layer, which embeds the sequence of event attributes.

$$< ea_1^1, \ldots, ea_n^1, \delta^{0-1}, \ldots, ea_1^m, \ldots, ea_n^m, \delta^{(m-1)-m} >$$

into a numeric vector (denoted as v_e in Fig. 3). Note that the elapsed time $\delta^{(i-1)-i}$ between events e^{i-1} and e^i is introduced as part of the events' attribute set, in order to take into account the temporal relation between events. Since trace attributes (resp. event attributes) can be either nominal or numerical, for the nominal attributes our technique creates an embedded representation for each value of the nominal domain, i.e. a vectorial numeric representation for each nominal value. Once the nominal attributes are converted, the trace attributes (resp. event attributes for each event of the sequence) are concatenated (concatenation is denoted as ";" in Fig. 3).

The vectorial representations of event's attributes is consequently passed through an LSTM recurrent neural network, which preserves the temporal correlation of events. For each event e^i of the sequence, the LSTM outputs a vectorial representation h_i summarizing all the events from e^1 to e^i. The sequence $< h_1, \ldots, h_m >$ is then passed trough an attention layer producing a summarized temporal representation z. In turn, z is concatenated with the vectorial representation of the trace attributes v_t. The concatenation of v_t and z results in a unique vectorial numeric representation which encodes both the trace's and events' attributes and the events temporal ordered sequence. This allows us to capture the multi-perspective aspect of a process, as information of different nature from the traces are included in the embeddings. The resulting vector is then passed to a Fully Connected (FC) layer (with a hyperbolic tangent activation function) that creates the trace embedding representation (TRACE EMB). The key idea is to reduce and remove the sparsity from the concatenation of the two embeddings v_t and z.

4.1.1. Attention mechanism

In the architecture depicted in Fig. 3 a key role is played by the attention layer. We recall here that *attention* (Chorowski et al., 2015) is a mechanism for discovering the parts of the input sequence that are the most relevant for describing the trace behavior. The ides is that these parts should be weighted more than the other parts, in the computation of the embedding. To have an intuition on how the attention mechanism is implemented, consider how human brain processes input information. Notably, the human brain chooses to focus on certain inputs, while ignoring or diminishing the importance of other ones. Inspired by this analogy, in (Bahdanau, Cho, & Bengio, 2014) the authors added an attention layer to the encoder-decoder framework allowing to automatically search for the parts of a source sentence that are relevant for predicting a target word, without having to form these parts as a hard segment explicitly. In our architecture, we shift this idea from the analysis of sentences to the analysis of process traces.

In our context, the attention layer takes as input the vectorial representation sequence $< h_1, \ldots, h_m >$ from the LSTM and returns the output $z = tanh(W_c[co; h_n])$ where.

- $W_c \in \mathscr{R}^{2|h_n| \times |h_n|}$ is a set of weights to learn, and
- co is the attention context vector obtained by the weighed mean of all h_i vectors, i.e., $co = \sum_{i=1}^{n} c_i^s \cdot h_i$.

The weights $c_1^s, \ldots, c_n^s$ are obtained by using (1) a unique fully connected layer that, applied singularly to each h_i, produces a single value q_i, and (2) a softmax layer taking in input all elements in $\{q_1, \ldots, q_n\}$ and producing $c_1^s, \ldots, c_n^s$.

4.2. Decoder

Given the comprehensive trace embedding (TRACE EMB), the final part of our network works on the reconstruction of the attributes taken in the input. In the case of trace's attributes, the reconstruction is performed as in the decoder of an autoencoder using one fully connected layer. In the case of event's attributes, instead, they are reconstructed by using a set of LSTMs, one for each attribute. For instance, consider a toy example where the event's attributes are $\{task, resource, state\}$. According with the notation introduced in the paper, a trace of length m is represented by a sequence of event's attributes: $< ea^1_{task}, ea^1_{resource}, ea^1_{state}, \ldots, ea^m_{task}, ea^m_{resource}, ea^m_{state} >$ with its instantiating values, for example $e^1 = \{$ *sendEmail*, *Bob*, *finished*$\}$.

In this case, the decoder uses three LSTMs, one for *task*, one for *resource* and one for *state*, in order to reconstruct the original sequence of their values. The idea is to capture the different temporal relations among each event and its attributes. The outputs of the network are the predicted values $\{\widetilde{ta}_1, \ldots, \widetilde{ta}_k\}$ of the trace's attributes, and the predicted values.

$$\{\widetilde{ea}_1^1, \ldots, \widetilde{ea}_n^1, \delta^{0-1}, \ldots, \widetilde{ea}_1^m, \ldots, \widetilde{ea}_n^m, \delta^{(m-1)-m}\}$$

of the events' attributes (included the elapsed time between events δ). To train the assembled neural network, we use a cumulative loss function that:

- for each numerical attribute sums the mean squared error between the reconstructed attribute $\widetilde{ta}_i$ (or the reconstructed sequence of attribute values $< \widetilde{ea}_i^1, \widetilde{ea}_i^2, \ldots, \widetilde{ea}_i^{m-1}, \widetilde{ea}_i^m >$) with the original attribute value ta_i (or the original attribute values of the sequences $< ea_i^1, ea_i^2, \ldots, ea_i^{m-1}, ea_i^m >$), and
- for each nominal attribute uses the cross-entropy and the softmax activation function to create a distribution among all the nominal values. In particular, softmax function outputs a vector that represents the probability distribution of the list of nominal attributes (with all probabilities sum to one), and the cross-entropy is used as a loss function to minimize in order to reduce the gap between the predicted value and the ground truth.

5. Experimental evaluation

In this section we illustrate the experimental settings we used to evaluate the effectiveness of our approach, and then we discuss the results of our experimental activity. Our goal is to provide answers to the following research questions:

RQ1 How does Deep-Trace2Vec perform compared with state-of-the-art techniques in process representation?
RQ2 How does the performance of Deep-Trace2Vec vary with respect to the multi-perspective specification?

The rest of the section is organized as follows: in Section 5.1 we describe the datasets used to test our approach; in Section 5.2 we provide details on the trace classification and trace clustering tasks used to evaluate and compare our embedding approach with the baselines; in Section 5.3 and Section 5.4 we describe the baselines we compare with and the implementation details of each embedding technique, respectively; in Section 5.5 we describe the experiment results for both the process mining tasks considered; finally, in Section 5.6 we summarize the experiment results and answer to the above research questions.

5.1. Datasets

To evaluate the usefulness of the proposed embedding technique, we apply it to two event logs for which a ground truth is known. In particular, we perform the evaluation over two of the well known real-life datasets provided by the BPI Challenge, an annual process mining competition. This datasets are widely used as benchmarks in the evaluation of process mining techniques. Below it follows a brief description of the datasets.

BPI Challenge 2015. The dataset (Van Dongen & Boudewijn, 2015), namely *BPIC_15* in short, consists of data generated from the building permit application process in five municipalities located in The Netherlands. The data contains 5.647 cases, and 27 types of activities for a total of 92.426 events over a period of four years. Each record describes a single step taken by an employee in the process, and consists of the case ID, the activity performed (given by its ID as well as its description in both English and Dutch), the time the action was completed, involved resources, and many other fields. Each trace is characterized by 6 nominal attributes and 1 numeric attribute, while each event is characterized by 6 nominal attributes. We consider the five municipalities as the class labels (values $1, 2, 3, 4, 5$). *BPIC_15* can be considered a balanced dataset as the five classes are distributed as following: $1-17\%, 2-17\%, 3-19\%, 4-22\%, 5-25\%$.

BPI Challenge 2019. The dataset (Van Dongen & Boudewijn, 2019), namely *BPIC_19* in short, collects data from a large multinational company operating with 60 subsidiaries from The Netherlands in the area of coatings and paints. Specifically, the data is recorded for the execution of Purchase-To-Pay processes (possibly of an ERP system) of purchase orders submitted in 2018, where each purchase order (or purchase document) contains one or more line items. For each line item, there are roughly two types of flows in the data, designed by the *Item Category* attribute which we use as class label (values $1, 2$). The data consists of 76.349 traces, and 42 types of activities performed by 627 users, for a total of 1.595.923 events recorded over a period of three years. Each trace is characterized by 6 nominal attributes and 1 numeric attribute, while each event is characterized by 2 nominal attributes. *BPIC_19* is an unbalanced dataset as the two classes are distributed as following: $1-77\%, 2-23\%$.

5.2. Experimental setting

To validate our embedding technique we used the trace clustering and trace classification tasks, that are actually the most used in the process mining settings.

5.2.1. Trace classification

Trace classification is a process mining task that allows to determine whether traces belong or not to a predefined class. Trace classification has been proved to be useful in the case of conformance checking and predictive monitoring. In fact, the former can be reformulated as a binary classification task where the problem is to determine if traces fit (or not) the process model, i.e. traces can be replayed by the underline model (Okoye Dr, Naeem Dr, Islam Dr, Tawil Dr, & Lamine Dr, 2017). Similarly, the latter can be reformulated as a multi-classification task where the problem is to classify traces based on their processing time, or the associated outcome (e.g. success or failure). We executed three classification algorithms, namely Decision Tree, Random Forest and Logistic Regression. We applied a 10-fold cross-validation, so as 90% of the traces was used to train the model and the remaining 10% was used to test the performance of the learning technique. To measure the effectiveness of the model we adopted the F1-score, Precision, and Recall measures, obtained as the average value computed over the results of each folder.

5.2.2. Trace clustering

Trace clustering, that is, splitting an event log horizontally by grouping traces based on a specific criterion of similarity, has received great attention from the process mining community and, as a matter of fact, a plethora of trace clustering approaches have been proposed in the last decade. However, as pointed out in (De Koninck et al., 2018), whether a clustering technique is using the right types of features with respect to the class label is typically not known by the user, and the similarity of the obtained clusters to the ground truth mainly depends on the clustering algorithm used. In fact, different clustering algorithms group data points according to different criteria, and they do it in an unsupervised way, which means that they do not leverage any relation between the features and the class label. Thus, we believe that the ability of a condensed representation of a process of representing its execution traces cannot be properly evaluated with a clustering task, especially if a multi-perspective condensed representation such as Deep-Trace2Vec is considered. Such kind of condensed representations contains a large amount of information, and it is difficult for clustering algorithms to extract only one particular aspect arbitrarily selected, unless a feature selection task is carried preliminarily. Feature selection is the process of selecting the subset of the most relevant features. By performing feature selection we can carry out clustering only considering those features that are more representative with respect to the target attribute, so as to obtain more significative results for our analysis.

The experiments consisted in performing a hierarchical clustering algorithm (based on Ward's minimum variance method), a k-means algorithm, and the Gaussian Mixture Models (GMM) on the vectorial representations of the execution traces. Then, we measured the extent to which the resulting cluster solution captures the same information of the ground truth using four metrics, namely the Adjusted Rand Index (ARI) (Rand, 1971), the Adjusted Mutual Information score (AMI) (Vinh, Epps, & Bailey, 2010), the Homogeneity score (HS), and the Completeness score (CS) (Hubert & Arabie, 1985). ARI and AMI are very similar indexes, they are both used to compare clusterings of the same dataset. However, according to Romano et al. (Romano, Vinh, Bailey, & Verspoor, 2016) ARI should be used when the reference clustering has large equal sized clusters, while AMI should be used when the reference clustering is unbalanced. For this reason we use ARI in the evaluation of the *BPIC*_15 dataset, and AMI in the evaluation of the *BPIC*_19 dataset.

5.3. Baseline

For both the classification and the clustering tasks we compared the results obtained with our method with those obtained with three existing embedding techniques. Such techniques were chosen as the most suitable for the purpose of representing process execution traces and the ability of representing multidimensional data. Below we provide a brief description of such techniques.

Trace2Vec (De Koninck et al., 2018): is a representation of traces based on the trace identifier, which allows for joint learning of representations of activities and traces. The embedding is based on the Distributed Memory Model of Paragraph Vectors (PVDM) approach, which can be regarded as an extension of CBOW (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), a model architecture used in the natural language processing domain. Here, exploiting the analogy between activities and words, traces can be regarded as sentences as well. However, this method is not suitable for the representation of multidimensional data as it does take into account traces and event identifiers only, disregarding traces and events attributes. Nevertheless, as far as we know, Trace2Vec is the only unsupervised learning method for representing process executions, and as such, we have chosen to adopt it as a baseline for the comparison.

node2Vec (Grover & Leskovec, 2016): is a feature representation method which learns a mapping of nodes to a $n-dimensional$ space of features and maximizes the likelihood of preserving network neighborhoods of nodes. An event-log can be appropriately represented as a graph using vertices as features and edges as significant relation between them. In particular, we represented an event log as a directed graph with three types of vertices, namely trace vertices v_t, event vertices v_e, and attribute vertices v_a, and with a directed edge from v_t to v_e if e is the first event of trace t, from v_{e1} to v_{e2} if event $e1$ precedes event $e2$ (and both events pertain to the same trace), from v_t to v_a if a is an attribute value of trace t, and from v_e to v_a if a is an attribute value of event e. According to this mapping, traces (resp. events) with same attribute values are mapped as vertices in the graph linked with common attribute vertices. This kind of representation enables the representation of multidimensional data as both trace and event attributes are taken into account. Graph representation has been already used in the field of text mining (Sonawane & Kulkarni, 2014), and for activity detection and prediction in events logs (Augustyniak & Ślusarczyk, 2018), due to its ability of preserving relationships and structural information effectively.

n-grams: are widely used in computational linguistics (e.g. statistical natural language processing (Manning & Schütze, 1999)), computational biology (e.g. DNA sequence analysis (Kim & Shawe-Taylor, 1994)), data compression (Witten, Bray, Mahoui, & Teahan, 1999), information retrieval (Baeza-Yates, 1992), and business process management (Horita, Hirayama, Hayase, Tahara, & Ohsuga, 2016; Wombacher, 2006). The authors of (Horita et al., 2016) proposed the usage of n-grams for a trace classification task. In particular, traces are converted in 2-grams vectors, so as 2-grams are the features representing event orders in each trace, and used as input to a decision tree algorithm for trace classification. We took inspiration from this work, and, besides the sequences of events, we applied the same principle to the sequences of trace's and events' attributes, so as to represent multidimensionality.

5.4. Implementation details

All the experiments have been carried out in a computer with an Intel Xeon E5-2620 CPU. This CPU has 6 cores running at 2.00 GHz and 12 threads. The computer also has 102 GB of RAM memory and a Nvidia GeForce GTX 1080 Ti GPU with 11 GB of memory.

Deep-Taace2Vec was implemented using Python and the neural network API Keras (Chollet et al., 2015) in conjunction with Tensorflow (Abadi et al., 2016). For the training regimen we used 100 epochs, the batch sizes where selected by taking into account the sizes of the utilized

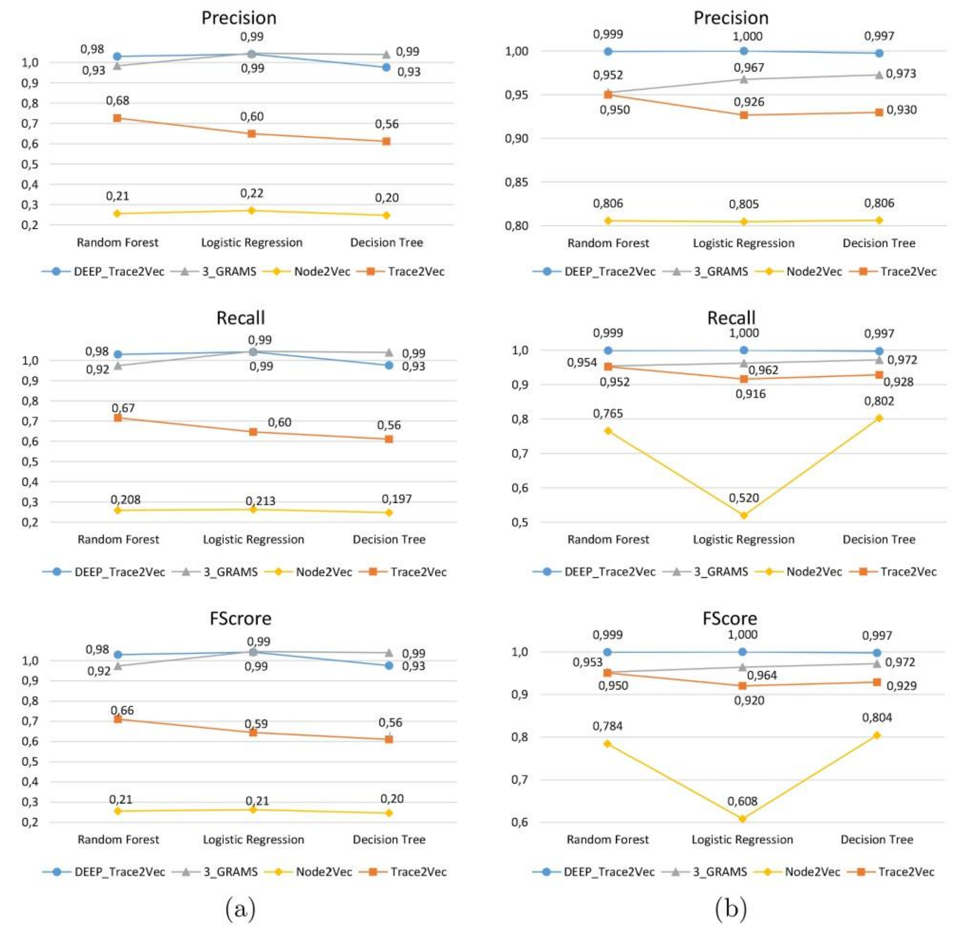


Fig. 4. Evaluation metrics values of the classification task for the *BPIC*_15 (a) and the *BPIC*_19 (b) datasets.

datasets, with the objective to utilize the maximum amount of available memory and, at the same time, minimize computation times. Thus the selected batch size depends on the dataset and was set to 400 and 1000 for *BPIC*_15 and *BPIC*_19 respectively. Each individual feature was trained using a embedding size of 4. The embeddings obtained as the result of the auto-encoder are numeric vectors of size 100. The code is publicly available.[5].

For Trace2Vec and node2Vec we used the Python implementations and default parameters provided by the authors of (De Koninck et al., 2018 & Grover & Leskovec, 2016), respectively.

For n-grams we adopted the `CountVectorizer` class of the `sklearn` Python library in order to compute the so called *document-term matrix*, where documents are the trace identifiers, and terms are the n-grams computed over the sequences of events and sequences of event attributes.

5.5. Results

The results of the classification task are represented in Fig. 4. First of all, we notice that Deep-Trace2Vec provides very high values of precision, recall, and f-score, outperforming or equalizing the other approaches for every classification algorithm and for both datasets, except for n-grams when the classification is performed by the decision tree algorithm.[6] Furthermore, the gap between Deep-Trace2Vec and Trace2-Vec, which to date is the only unsupervised learning method for representing process executions, is significantly high for *BPIC*_15 (on average 35 percentage point higher), and 2 percentage point circa higher for *BPIC*_19.

The results of clustering methods are illustrated in Fig. 5. Each point visualizes the similarity of the resulting clustering to the ground truth.

[5] https://github.com/edoardoserra/DeepTrace2Vec

[6] For n-grams we only show results with $n = 3$ as we empirically found that for different values of n performances do not change significantly.

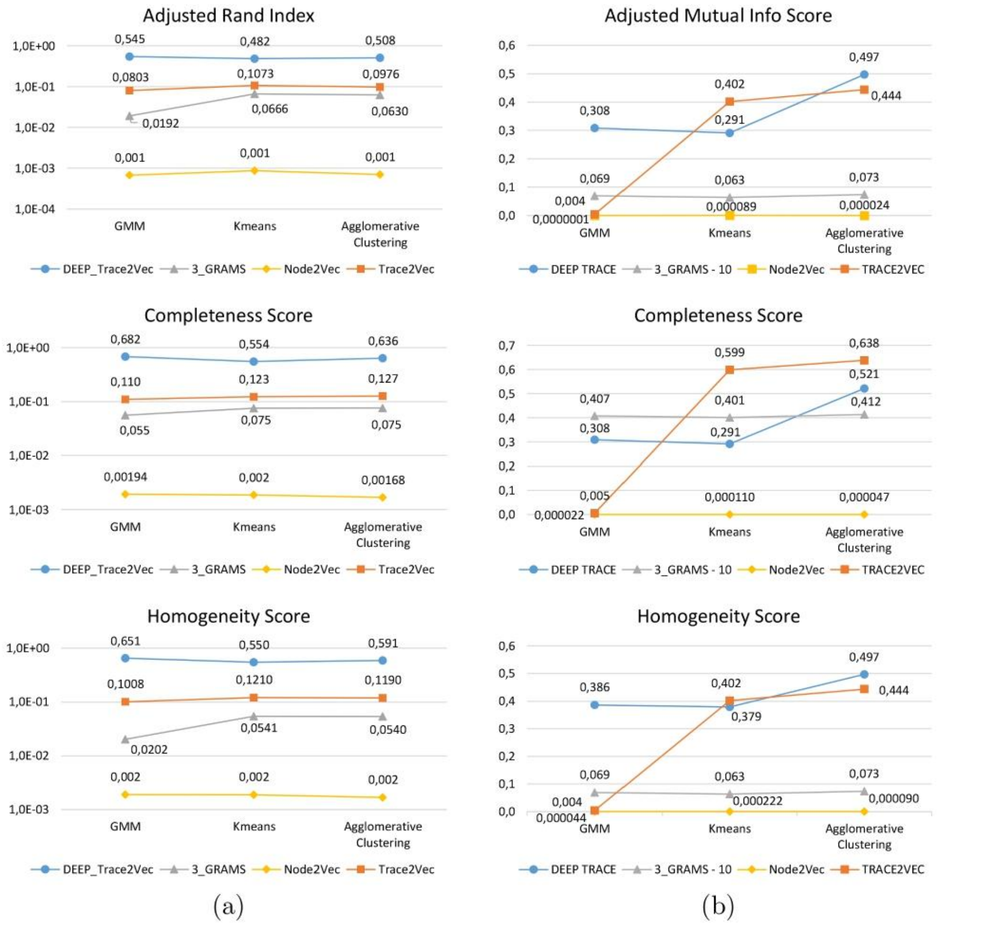


Fig. 5. Evaluation metrics values of the clustering task for the *BPIC*_15 (a) and the *BPIC*_19 (b) datasets.

Before computing the clusters we applied a feature selection task to both datasets, selecting 20 features for *BPIC*_15, and 10 features for *BPIC*_19. We can notice that for *BPIC*_15 Deep-Trace2Vec performs much better than other techniques as it provides metrics values one or two order of magnitude higher, while for *BPIC*_19, although not always Deep-Trace2Vec is the best choice, it still provides best values of AMI and HS with 2 out of 3 clustering algorithms.

These results are further supported by the visualization of the embedding spaces depicted in Figs. 6–8. To visualize the spaces we took advantage of the Principal Component Analysis (PCA) (Jolliffe, 2011), a well known dimensionality reduction technique which identifies the subspace in which the data approximately lies, allowing us to properly visualize the original 100 dimensional embedding space to a more convenient 2 dimensional space. Figs. 6 and 8 show the PCA plots obtained with the Deep-Trace2Vec technique before and after the feature selection task. We can observe how the feature selection task has noticeably improved the representation capabilities of the embeddings for both datasets. In fact, the plots on the left hand side of Figs. 6 and 8 show a set of points forming more defined clusters with respect to those obtained without feature selection. Figs. 7 and 9 show the PCA plots obtained after the feature selection task with the embeddings computed by the three approaches we compare to. From these plots it is clear that these embeddings form more overlapping clusters with respect to those obtained with Deep-Trace2Vec. This, (plus Figs. 4 and 5) confirms that Deep-Trace2Vec generates embeddings with higher representational capabilities than those provided by other techniques.

To conclude, we conducted an ablation study of Deep-Trace2Vec in order to answer to **RQ2**. In particular, we compared Deep-Trace2Vec with a modified version that does not include multiple data dimensions. In other words, it takes into account the control flow perspective only, that is, for each trace, neither trace nor event attributes are fed into the framework but the sequence of events only. It can be noticed that Deep-Trace2Vec drastically improves the performances of both the classification and the clustering tasks for both datasets. This result (see Figs. 10 and 11) confirms that taking into account the multiple process perspectives is fundamental when representation learning is applied in the

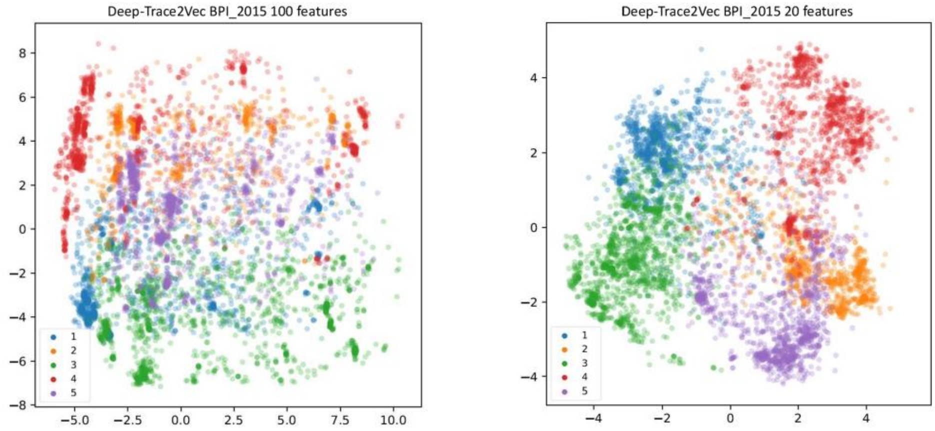


Fig. 6. PCA plots of the Deep-Trace2Vec embeddings before (left) and after (right) the feature selection task for the *BPIC*_15 dataset.

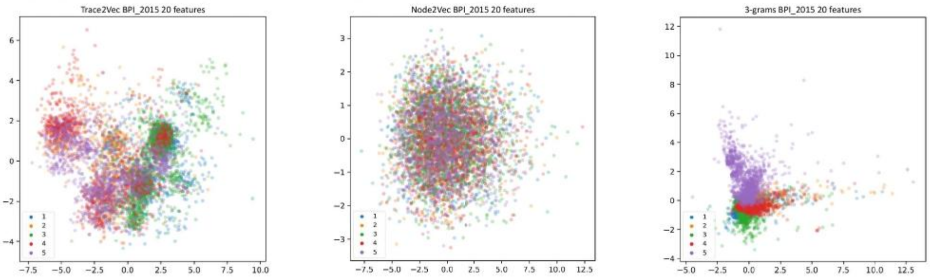


Fig. 7. PCA plots of the Trace2Vec (left), Node2Vec (middle) and 3-grams (right) embeddings after the feature selection task for the *BPIC*_15 dataset.

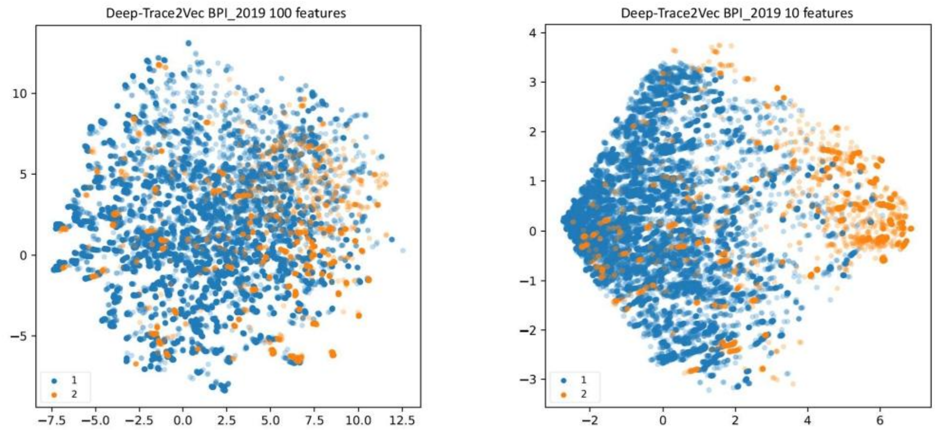


Fig. 8. PCA plots of the Deep-Trace2Vec embeddings before (left) and after (right) the feature selection task for the *BPIC*_19 dataset.

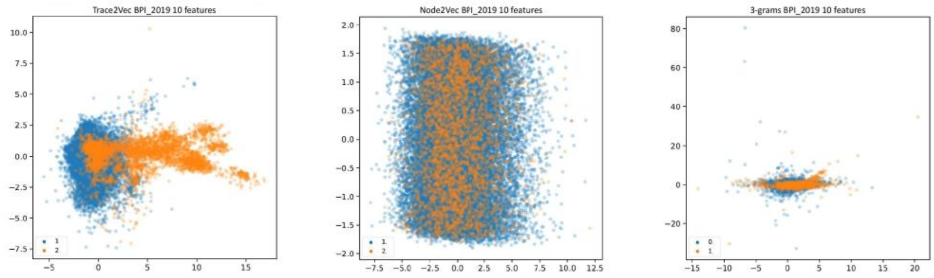


Fig. 9. PCA plots of the Trace2Vec (left), Node2Vec (middle) and 3-grams (right) embeddings after the feature selection task for the *BPIC_19* dataset.

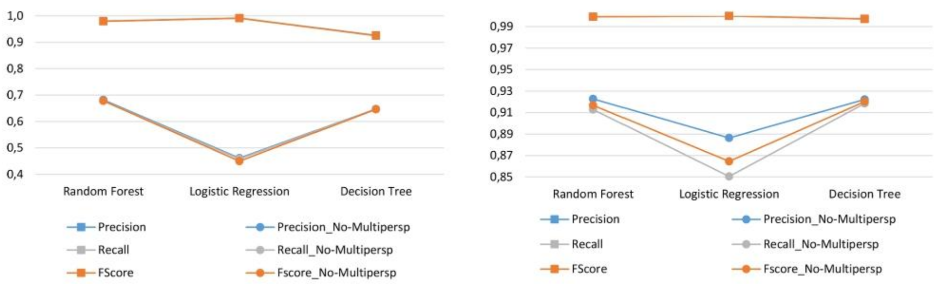


Fig. 10. Comparison of Deep-Trace2Vec performances with and w/o multi-perspective representation for the classification task applied at the *BPIC_15* (left) and the *BPIC_19* (right) datasets.

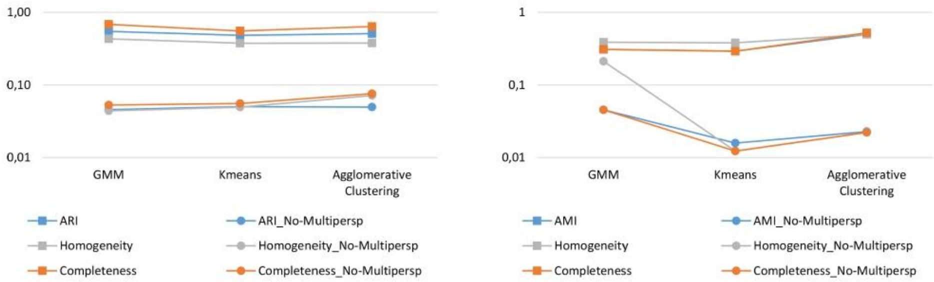


Fig. 11. Comparison of Deep-Trace2Vec performances with and w/o multi-perspective representation for the clustering task applied at the *BPIC_15* (left) and the *BPIC_19* (right) datasets.

field of process mining, as it allows to get a deep insight into the process behavior. In fact, for their very nature process executions are a heterogeneous collection of attributes, that confer the embeddings higher representational capabilities compared when such attributes are omitted from the computation.

5.6. Conclusive discussion

The main results of Section 5.5 are summarized below to provide a clear answer to the research questions defined at the beginning of this section.

RQ1 Our embedding in comparison with the state-of-art techniques usually performs better in both trace classification and trace clustering tasks. In particular, Deep-Trace2Vec significantly outperforms all the baselines in the classification task when tested with *BPIC_19*, and in the clustering task when tested with *BPIC_15*, while it results to be comparable with the n-gram-based approach in the classification task when tested with *BPIC_15*, and with Trace2Vec in the clustering task when tested with *BPIC_19*. In addition, from the qualitative visualization analysis shown in Figs. 6–8, we observe that Deep-Trace2Vec produces the most homogeneous and separable regions characterizing the classes provided in the ground truth, meaning that our embedding technique has more representative power than the other ones.

RQ2 We demonstrated that taking into account the multiple perspectives of the process increases the representative power of the trace embeddings. This is revealed by the ablation study conducted in Section 5.5 where Deep-Trace2Vec drastically outperforms the single-perspective approach for both the trace classification and trace clustering tasks. We can thus conclude that modeling the multi-dimensional nature of processes allows a better representation of the processes behavior.

6. Conclusions

Nowadays, there is an enormous amount of data storing process executions containing high-dimensional and heterogenous information. This data can be formatted as event log and analyzed with process mining techniques in order to extract useful knowledge for the analysis of the multiple and various characteristics of processes. However, current literature primarily focuses on the control-flow perspective, or in a better case considers additional perspectives separately. Moreover, there is very little prior process mining research on exploiting neural network-based embedding. To overcome these issues, in this paper we have addressed the problem of computing a concise and informative representation of execution traces that includes the multiple perspectives of the process behavior. We proposed Deep-TRace2Vec, a deep learning-based approach that is able to encode complex process behavior in the form of a vectorial representation of the traces. Deep-TRace2Vec is able to learn robust and compact representations automatically in an usupervised way, that is, it does not rely on any prior knowledge about the process. We used several deep recurrent multilayer neural networks in order to embed the multidimensional characteristic of traces along with their connections. We evaluated our embedding with two real world datasets taken from the BPI Challenge 2015 and the BPI Challenge 2019 competitions. Results have shown that Deep-TRace2Vec compared with the state of art techniques, provides a richer and more informative representation space that is better suited in trace classification and clustering tasks. Moreover, the performed ablation study showed how the multi-perspective feature that also includes the temporal dimension of a trace, drastically improves the performance in both the trace classification and clustering tasks.

As future work we aim at extending the proposed model in order to obtain good performances also when used for different data mining tasks like anomaly detection and novelty detection. Moreover, we will also seek to investigate how effectively combine dimensionality reduction of the embedding architecture with surrogate modelling (Alizadeh et al., 2019; Alizadeh, Allen, & Mistree, 2020; Jia et al., 2020), with the aim of improving the performances of simulation tasks for engineering design. We will also investigate extensions of our approach with transformer neural networks (Devlin, Chang, Lee, & Toutanova, 2018) that are another suitable processing unit for sequential and temporal data.

CRediT authorship contribution statement

Antonella Guzzo: Conceptualization, Methodology, Writing - review & editing. **Antonino Rullo:** Software, Methodology, Writing - review & editing. **Edoardo Serra:** Conceptualization, Methodology, Writing - review & editing. **Mikel Joaristi:** Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* (pp. 265–283).

Alizadeh, R., Allen, J. K., & Mistree, F. (2020). Managing computational complexity using surrogate models: a critical review. *Research in Engineering Design, 31*(3), 275–298.

Alizadeh, R., Jia, L., Nellippallil, A. B., Wang, G., Hao, J., Allen, J. K., & Mistree, F. (2019). Ensemble of surrogates and cross-validation for rapid and accurate predictions using small data sets. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AI EDAM, 33*(4), 484–501.

Augustyniak, P., & Slusarczyk, G. (2018). Graph-based representation of behavior in detection and prediction of daily living activities. *Computers in Biology and Medicine, 95*, 261–270.

Baeza-Yates, R. A. (1992). Text-retrieval: Theory and practice. In *IFIP congress (1)*, Vol. 12, pp. 465–476.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. CoRR abs/1409.0473.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(8), 1798–1828.

Bezerra, F., Wainer, J., & van der Aalst, W. M. (2009). Anomaly detection using process mining. In *Enterprise, business-process and information systems modeling* (pp. 149–161). Springer.

Böhmer, K., & Rinderle-Ma, S. (2017). Multi instance anomaly detection in business process executions, in *International conference on business process management* (pp. 77–93). Springer.

Böhmer, K., & Rinderle-Ma, S. (2016). Multi-perspective anomaly detection in business process execution events. In *OTM confederated international conferences on the move to meaningful internet systems* (pp. 80–98). Springer.

Bose, R. P. J. C., & van der Aalst, W. M. P. (2009). Context aware trace clustering: Towards improving process mining results. In *Proc of the SIAM international conference on data mining (SDM 2009)* (pp. 401–412).

Chollet, F., et al. (2015). *Keras*. URL: https://github.com/fchollet/keras.

Chorowski, J. K., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y. (2015). Attention-based models for speech recognition. In *Advances in neural information processing systems* (pp. 577–585).

De Koninck, P., vanden Broucke, S., & De Weerdt, J. (2018). act2vec, trace2vec, log2vec, and model2vec: Representation learning for business processes. In International conference on business process management. (pp. 305–321). Springer.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv: 1810.04805.

De Weerdt, J., Schupp, A., Vanderloock, A., & Baesens, B. (2013). Process mining for the multi-faceted analysis of business processes—a case study in a financial services organization. *Computers in Industry, 64*(1), 57–67.

dos Santos Garcia, C., Meincheim, A., Faria Junior, E. R., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., Santos, E. A. P., & Scalabrin, E. E. (2019). Process mining techniques and applications – a systematic mapping study. *Expert Systems with Applications, 133*, 260–295. URL: https://www.sciencedirect.com/science/article/pii/S0957417419303161.

Evermann, J., Rehse, J., & Fettke, P. (2017). Predicting process behaviour using deep learning. *Decision Support Systems, 100*, 129–140.

Folino, F., Guarascio, M., & Pontieri, L. (2012). Context-aware predictions on business processes: An ensemble-based solution. *New frontiers in mining complex patterns – first international workshop, NFMCP 2012 revised selected papers*, 215–229.

Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems, 151*, 78–94.

Grover, A., & Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. In *Proceedings of the 22Nd ACM SIGKDD international conference on knowledge discovery and data mining, KDD '16* (pp. 855–864). ACM.

Horita, H., Hirayama, H., Hayase, T., Tahara, Y., & Ohsuga, A. (2016). Business process verification and restructuring ltl formula based on machine learning approach. In *Computer and information science* (pp. 89–102). Springer.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(1), 193–218.

Jans, M., Van Der Werf, J. M., Lybaert, N., & Vanhoof, K. (2011). A business process mining application for internal transaction fraud mitigation. *Expert Systems with Applications, 38*(10), 13351–13359.

Jia, L., Alizadeh, R., Hao, J., Wang, G., Allen, J. K., & Mistree, F. (2020). A rule-based method for automated surrogate model selection. *Advanced Engineering Informatics, 45*, Article 101123.

Jolliffe, I. (2011). *Principal component analysis*. Springer.

Kim, J. Y., & Shawe-Taylor, J. (1994). Fast string matching using an n-gram algorithm. *Software: Practice and Experience, 24*(1), 79–88.

Luca, S. E., Pimentel, M. A., Watkinson, P. J., & Clifton, D. A. (2018). Point process models for novelty detection on spatial point patterns and their extremes. *Computational Statistics and Data Analysis, 125*, 86–103.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA, USA: MIT Press.

Mehdiyev, N., Evermann, J., & Fettke, P. (2018). A novel business process prediction model using a deep learning method. *Business & Information Systems Engineering*, 1–15.

Meng, Q., Catchpoole, D., Skillicom, D., & Kennedy, P. J. (2017). Relational autoencoder for feature extraction. In *in '2017 International Joint Conference on Neural Networks (IJCNN)', IEEE* (pp. 364–371). IEEE.

Mhaskar, H., & Poggio, T. A. (2016). *Deep vs. shallow networks: An approximation theory perspective*. CoRR abs/1608.03287.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality*. CoRR abs/1310.4546.

Navarin, N., Vincenzi, B., Polato, M., & Sperduti, A. (2017). Lstm networks for data-aware remaining time prediction of business process instances. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1–7).

Nolle, T., Luettgen, S., Seeliger, A., & Mühlhäuser, M. (2018). Analyzing business process anomalies using autoencoders. *Machine Learning* (pp. 1–19).

Nolle, T., Seeliger, A., & Mühlhäuser, M. (2018). Binet: multivariate business process anomaly detection using deep learning. In *International conference on business process management* (pp. 271–287). Springer.

Okoye Dr, K., Naeem Dr, U., Islam Dr, S., Tawil Dr, A.-R. H., & Lamine Dr, E. (2017). Process models discovery and traces classification: A fuzzy-bpmn mining approach. *Journal of International Technology and Information Management, 26*(4), 2–50.

Polato M., Sperduti A., Burattin, A., & de Leoni, M. (2018). Time and activity sequence prediction of business process instances. *Computing 100*(9), 1005–1031.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association, 66*(336), 846–850.

Romano, S., Vinh, N. X., Bailey, J., & Verspoor, K. (2016). Adjusting for chance clustering comparison measures. *The Journal of Machine Learning Research, 17*(1), 4635–4666.

Rullo, A., Guzzo, A., Serra, E., & Tirrito, E. (2020). A framework for the multi-modal analysis of novel behavior in business processes. In *International conference on intelligent data engineering and automated learning* (pp. 51–63). Springer.

Sonawane, S. S., & Kulkarni, P. A. (2014). Graph based representation and analysis of text document: A survey of techniques. *International Journal of Computer Applications, 96*(19).

Staudemeyer, R. C., & Morris, E. R. (2019). *Understanding lstm–a tutorial into long short-term memory recurrent neural networks*. arXiv preprint arXiv:1909.09586.

Tavares, G. M., Barbon Junior, S., Ceravolo, P., & Damiani, E. (2020). Evaluating trace encoding methods in process mining. In *9th International symposium DATAMOD 2020*. Springer.

van der Aalst, W. M. P. (2016). *Process mining – Data science in action* (2nd Ed.). Springer.

Van Dongen, B.F., & Boudewijn (2015). BPI Challenge 2015, Eindhoven University of Technology. URL: https://data.4tu.nl/repository/uuid:31a308ef-c844-48da-948c-305d167a0ec1 [DATASET].

Van Dongen, B.F. Boudewijn (2019). BPI Challenge 2019, 4TU.Centre for Research Data. URL: https://data.4tu.nl/repository/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1 [DATASET] .

Van Dongen, B. F., & Van der Aalst, W. M. (2004). Multi-phase process mining: Building instance graphs. In *International conference on conceptual modeling* (pp. 362–376). Springer.

van Eck, M. L., Sidorova, N., & van der Aalst, W. M. P. (2016). Discovering and exploring state-based models for multi-perspective processes. In M. L. Rosa, P. Loos, & O. Pastor (Eds.), *Business process management – 14th international conference, BPM 2016, Rio de Janeiro, Brazil, September 18–22, 2016. Proceedings', Vol. 9850 of Lecture notes in computer science* (pp. 142–157). Springer.

Vinh, N. X., Epps, J., & Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research, 11*(Oct), 2837–2854.

Witten, I. H., Bray, Z., Mahoui, M., & Teahan, W. J. (1999). Text mining: A new frontier for lossless compression. In *Data compression conference, DCC 1999, Snowbird, Utah, USA, March 29–31, 1999* (pp. 198–207). doi: 10.1109/DCC.1999.755669.

Wombacher, A. (2006). Evaluation of technical measures for workflow similarity based on a pilot study. In *OTM confederated international conferences on the move to meaningful internet systems* (pp. 255–272). Springer.

Yang, W.-S., & Hwang, S.-Y. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert Systems with Applications, 31*(1), 56–68.