Maestría en

Ciencia de Datos y Máquinas de Aprendizaje con Mención en Inteligencia Artificial

Trabajo previo a la obtención de título de Magister en Ciencia de Datos y Máquinas de Aprendizaje con Mención en Inteligencia Artificial

AUTORES:

Avilés Paute José Miguel

Barros Viteri David Alexis

Ramírez Vilaña Luis Miguel

Villacís Oleas Jorge Israel

TUTORES:

Andrés Roberto Navas Castellanos

Pablo Xavier Andrade Montalvo

TEMA:

Predicción del Rendimiento Académico Universitario Mediante Modelos Secuenciales Stacked BiLSTM en Entornos Virtuales de Aprendizaje

Quito - Ecuador

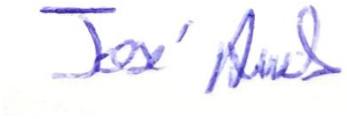
Julio – 2026

REINVENTEMOS EL FUTURO

Certificación de autoría

Nosotros, **Avilés Paute José Miguel, Barros Viteri David Alexis, Ramírez Vilaña Luis Miguel, Villacís Oleas Jorge Israel,** 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 (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.



Firma
Avilés Paute José Miguel

Firma
Barros Viteri David Alexis

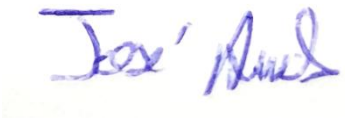
Firma
Ramírez Vilaña Luis Miguel

Firma
Villacís Oleas Jorge Israel

Autorización de Derechos de Propiedad Intelectual

Nosotros, **Avilés Paute José Miguel, Barros Viteri David Alexis, Ramírez Vilaña Luis Miguel, Villacís Oleas Jorge Israel,** en calidad de autores del trabajo de investigación titulado ***Titulo del trabajo de investigación Predicción del Rendimiento Académico Universitario Mediante Modelos Secuenciales Stacked BiLSTM en Entornos Virtuales de Aprendizaje***, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, (Julio 2026)



Firma
Avilés Paute José Miguel

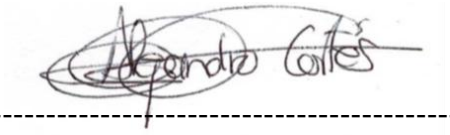
Firma
Barros Viteri David Alexis

Firma
Ramírez Vilaña Luis Miguel

Firma
Villacís Oleas Jorge Israel

Aprobación de dirección y coordinación del programa

Nosotros, **Alejandro Cortés López y Karla Estefanía Mora Cajas,** declaramos que: **Avilés Paute José Miguel, Barros Viteri David Alexis, Ramírez Vilaña Luis Miguel, Villacís Oleas Jorge Israel,** son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés López
Director de la
Maestría en Ciencia de datos y máquinas de aprendizaje con mención en Inteligencia Artificial

Karla Estefanía Mora Cajas
Coordinadora de la
Maestría en Ciencia de datos y máquinas de aprendizaje con mención en Inteligencia Artificial

DEDICATORIA

A nuestras familias, por su apoyo incondicional y su paciencia a lo largo de cada etapa de esta maestría; y a quienes, con su ejemplo, nos enseñaron que el esfuerzo sostenido es el camino hacia el conocimiento.

AGRADECIMIENTOS

Expresamos nuestro sincero agradecimiento a la Universidad Internacional del Ecuador y al cuerpo docente de la Maestría en Ciencia de Datos y Máquinas de Aprendizaje, por la formación recibida y por su acompañamiento académico. Agradecemos a nuestros tutores por su guía rigurosa a lo largo del desarrollo de este proyecto y a The Open University por liberar el dataset OULAD, sin el cual esta investigación no habría sido posible. Finalmente, agradecemos a nuestras familias y compañeros por su respaldo constante.

RESUMEN

El abandono estudiantil y los resultados académicos deficientes constituyen una problemática recurrente en la educación superior virtual, particularmente para las instituciones que administran Entornos Virtuales de Aprendizaje (EVA). La presente investigación plantea y valida una arquitectura secuencial de tipo Stacked BiLSTM orientada a la predicción de la condición académica final del estudiante (Fail, Pass o Distinction), tomando como insumo su actividad semanal registrada en el EVA, con el propósito de sustentar sistemas de alerta temprana. El desarrollo se rigió por la metodología CRISP-DM y empleó el Open University Learning Analytics Dataset (OULAD); una vez descartados los registros de estudiantes retirados, se obtuvieron 22 437 trayectorias académicas, las cuales fueron estructuradas en un tensor de 33 semanas × 7 variables. Sobre esta representación se entrenó una red BiLSTM apilada de dos capas (128 y 64 unidades), cuyo desempeño se contrastó con modelos LSTM, Transformer, Random Forest y XGBoost. El modelo alcanzó una exactitud del 78,39 % y un F1-Score ponderado de 0,7907 (AUC-ROC = 0,9105), con estabilidad confirmada por validación cruzada estratificada (F1 = 0,7734 ± 0,0076). Si bien XGBoost (0,8075) y Random Forest (0,8031) alcanzaron un desempeño global ligeramente mayor en términos absolutos, el modelo BiLSTM evidenció una ventaja funcional determinante: logró un F1 de 0,7619 utilizando únicamente las cuatro semanas iniciales del período académico, además de ofrecer una capacidad interpretativa (mediante gradientes, valores SHAP y mecanismos de atención) que permitió converger en la identificación de las semanas de mayor riesgo. Los hallazgos permiten sostener que la aproximación secuencial resulta factible como mecanismo de alerta temprana interpretable.

Palabras clave: *learning analytics*, predicción del rendimiento académico, stacked BiLSTM, alerta temprana, OULAD, interpretabilidad, CRISP-DM.

ABSTRACT

Student dropout and poor academic outcomes represent a persistent challenge in online higher education, particularly for institutions that manage Virtual Learning Environments (VLE). This study proposes and validates a sequential Stacked BiLSTM architecture aimed at predicting a student's final academic outcome (Fail, Pass, or Distinction) based on their weekly activity recorded in the VLE, with the purpose of supporting early warning systems. The work followed the CRISP-DM methodology and employed the Open University Learning Analytics Dataset (OULAD); after discarding the records of withdrawn students, 22,437 academic trajectories were obtained and structured into a 33-week × 7-feature tensor. Using this representation, a two-layer stacked BiLSTM network (128 and 64 units) was trained, and its performance was compared against LSTM, Transformer, Random Forest, and XGBoost models. The model achieved an accuracy of 78.39% and a weighted F1-score of 0.7907 (AUC-ROC = 0.9105), with stability confirmed through stratified cross-validation (F1 = 0.7734 ± 0.0076). Although XGBoost (0.8075) and Random Forest (0.8031) attained a slightly higher overall performance in absolute terms, the BiLSTM model demonstrated a decisive functional advantage: it reached an F1 of 0.7619 using only the first four weeks of the academic term, while also offering interpretability (through gradients, SHAP values, and attention mechanisms) that converged on the identification of the highest-risk weeks. The findings support the conclusion that the sequential approach is feasible as an interpretable early warning mechanism.

Keywords: learning analytics, academic performance prediction, stacked BiLSTM, early warning, OULAD, interpretability, CRISP-DM

TABLA DE CONTENIDOS

LISTA DE TABLAS

LISTA DE FIGURAS

CAPÍTULO 1. INTRODUCCIÓN

1.1 Planteamiento del problema e importancia del estudio

La masificación de la educación superior en entornos virtuales ha transformado las dinámicas universitarias contemporáneas y ha convertido la interacción cotidiana con las plataformas en una fuente masiva de datos sobre el proceso de aprendizaje (Siemens, 2013). Los estudiantes ya no asisten únicamente a aulas físicas; ahora generan un rastro digital cada vez que hacen clic en un recurso, descargan un contenido, participan en un foro o entregan una tarea. Cada una de estas acciones queda registrada como una traza susceptible de análisis, de modo que el Entorno Virtual de Aprendizaje (EVA) se convierte, de facto, en un instrumento de observación continua del proceso educativo, y su explotación sistemática ha dado origen a los campos de la analítica del aprendizaje y la minería de datos educativos (Baker & Inventado, 2014; Romero & Ventura, 2020). No obstante, esta gran disponibilidad de datos no se convierte de manera espontánea en estrategias efectivas de permanencia estudiantil: las instituciones de modalidad virtual continúan registrando altos índices de deserción, así como la falta de instrumentos capaces de analizar estos flujos conductuales para prever el fracaso académico antes de que se torne irreversible (Bañeres et al., 2023; Hlosta et al., 2017).

Esta situación se ve intensificada por la propia índole de los enfoques que prevalecen. Una parte considerable de los sistemas de gestión del riesgo académico examina factores demográficos o socioeconómicos desde una perspectiva estática, pasando por alto que el aprendizaje virtual constituye un proceso dinámico que se transforma semana a semana (Kotsiantis et al., 2003; Romero & Ventura, 2010). Como resultado, las intervenciones se ponen en marcha de forma tardía, una vez que el estudiante ha reprobado una evaluación sumativa o ha permanecido ausente de la plataforma durante varias semanas, instante en el cual recuperar la matrícula se vuelve notablemente más complejo. La evidencia reciente muestra, en cambio, que las trazas de interacción con el EVA constituyen un predictor conductual valioso del

compromiso y del desenlace académico (Hussain et al., 2018) y que los modelos de aprendizaje profundo capaces de explotar esos grandes volúmenes de datos superan a los clasificadores tradicionales (Waheed et al., 2020). El eje central de esta investigación radica, justamente, en transitar desde la predicción estática hacia el análisis de series temporales continuas a través de técnicas de aprendizaje profundo, capitalizando la dimensión cronológica que los enfoques tradicionales desestiman.

La relevancia de esta investigación se fundamenta, asimismo, en la escala del fenómeno estudiado. Dentro del conjunto de datos examinado, aproximadamente un tercio de las matrículas (31,16 %) pertenece a estudiantes que desertaron del curso, mientras que un 21,64 % adicional lo concluyó con calificación reprobatoria; dicho de otro modo, más de la mitad del estudiantado no logra un desenlace satisfactorio, cifra que concuerda con los elevados índices de abandono reportados en la modalidad virtual (Bañeres et al., 2023). Ante este panorama, la analítica predictiva del aprendizaje brinda una alternativa para convertir los registros de interacción que actualmente permanecen almacenados sin aprovechamiento en un indicador procesable de riesgo, capaz de guiar la distribución de los recursos de tutoría hacia quienes presentan mayor necesidad (Siemens, 2013; Waheed et al., 2020).

1.2 Definición del proyecto

El presente trabajo de titulación aborda el diseño y la implementación de un pipeline predictivo sustentado en la metodología CRISP-DM (Chapman et al., 2000), cuyo fin es estimar de manera dinámica el rendimiento académico en el ámbito universitario. De forma concreta, el sistema categoriza el resultado final de cada matrícula en tres clases mutuamente excluyentes: reprobado (Fail), aprobado (Pass) y distinción (Distinction). Con este propósito, se desarrolla un modelo secuencial avanzado de tipo Stacked BiLSTM arquitectura cimentada en los trabajos fundacionales sobre la memoria de corto y largo plazo y su variante bidireccional (Graves & Schmidhuber, 2005; Hochreiter & Schmidhuber, 1997) que procesa

de manera cronológica los registros de interacción diaria en el EVA. A diferencia de un clasificador que aguarda hasta el cierre del ciclo, el modelo propuesto se concibe para producir alertas tempranas y dinámicas, propiciando intervenciones pedagógicas oportunas en el plano institucional.

La hipótesis central postula que la información temporal el ritmo, la constancia y la evolución de la actividad a lo largo del calendario académico aporta un valor operativo que los modelos carentes de memoria son incapaces de captar, en particular la posibilidad de prever el desenlace con varias semanas de anticipación. Cada matrícula se modela como una secuencia temporal, esto es, un tensor tridimensional de muestras × semanas × características, que habilita a la red para identificar patrones dinámicos de compromiso: aceleraciones iniciales, descensos graduales de actividad y repuntes vinculados a los períodos de evaluación.

1.3 Naturaleza o tipo de proyecto

Esta investigación posee un carácter tecnológico-aplicado, con un enfoque cuantitativo y experimental inscrito en el campo de la Ciencia de Datos y la analítica del aprendizaje (Learning Analytics) (Romero & Ventura, 2020; Siemens, 2013). No se busca un análisis correlacional de tipo pasivo, sino un trabajo de ingeniería de datos orientado a la gestión educativa, en el que se procesa un elevado volumen de información temporal más de diez millones de interacciones en el EVA con el fin de entrenar, optimizar y validar algoritmos supervisados de clasificación multiclase. La implementación técnica se lleva a cabo mediante el lenguaje de programación Python y los frameworks TensorFlow y Keras (Abadi et al., 2016), complementados con scikit-learn (Pedregosa et al., 2011) e imbalanced-learn, lo que asegura la escalabilidad y la reproducibilidad del pipeline de datos.

1.4 Justificación e importancia del trabajo de investigación

La justificación de esta investigación se apoya en una realidad institucional poco cómoda: los sistemas universitarios suelen reaccionar de forma tardía. Gran parte de las

estrategias de retención son de naturaleza reactiva y aguardan a que el estudiante repruebe un examen parcial o se ausente de la plataforma para activar las señales de alarma. Llegado ese punto, la intervención ha perdido buena parte de su eficacia. Al mismo tiempo, las plataformas virtuales acumulan grandes volúmenes de registros diarios que, en la práctica, se desaprovechan. Resulta insuficiente conocer únicamente el volumen total de clics de un estudiante durante un semestre; lo que en verdad determina el éxito o el fracaso es el ritmo, la constancia y la evolución de dicha actividad a lo largo del calendario académico (Waheed et al., 2020).

Es en este punto donde los modelos tradicionales de aprendizaje automático, como Random Forest o XGBoost, evidencian limitaciones conceptuales al aplicarse de forma convencional. Dichos algoritmos requieren "aplanar" los datos y tratan las interacciones de la semana 2 y de la semana 30 como variables estáticas e independientes, fragmentando así la continuidad temporal del comportamiento (Breiman, 2001; Chen & Guestrin, 2016). No obstante, el aprendizaje es un proceso secuencial. La adopción de una arquitectura Stacked BiLSTM subsana este vacío metodológico: la red analiza el historial del estudiante como una secuencia dinámica y extrae patrones en ambas direcciones del tiempo, hacia adelante y hacia atrás, en virtud de su naturaleza bidireccional (Graves & Schmidhuber, 2005; Schuster & Paliwal, 1997).

La pertinencia práctica para la Universidad Internacional del Ecuador (UIDE) y para el ecosistema de la educación virtual es evidente. Al lograr un F1-Score ponderado de 0,7907, la presente investigación pone de manifiesto que la analítica predictiva del aprendizaje no constituye un ejercicio meramente teórico, sino un instrumento de gestión. La automatización del pipeline habilita a los departamentos académicos para detectar descensos pronunciados de interacción durante las semanas críticas y poner en marcha tutorías personalizadas con días o incluso semanas de antelación al abandono del curso por parte del estudiante. A esto se suma

una dimensión ética: puesto que el sistema incide en decisiones que repercuten en la trayectoria de las personas, se descartaron de manera intencionada variables susceptibles de introducir sesgos discriminatorios y se integró un análisis de interpretabilidad, en línea con los requerimientos de una inteligencia artificial responsable (Barredo Arrieta et al., 2020; Baker & Inventado, 2014).

El aporte de esta investigación trasciende lo estrictamente técnico: dado que el abandono y la reprobación conllevan costos significativos para el estudiante, la institución y el sistema educativo en su conjunto, un sistema de alerta temprana posibilita reorientar el acompañamiento tutorial desde un enfoque reactivo hacia uno preventivo y focalizado. En consecuencia, más allá del modelo en sí, este trabajo incorpora un análisis de su impacto potencial y una propuesta de despliegue factible, ambos desarrollados en el Capítulo 5.

1.5 Alcance

El alcance de este trabajo se define con precisión a fin de delimitar tanto sus aportes como sus límites. Respecto a lo que comprende, el proyecto abarca el desarrollo integral del pipeline CRISP-DM sobre el conjunto de datos OULAD desde la comprensión del negocio hasta la interpretabilidad del modelo; la construcción de un modelo principal Stacked BiLSTM y su comparación sistemática con cuatro arquitecturas alternativas (LSTM simple, Transformer con autoatención, Random Forest y XGBoost); la ingeniería de características temporales que convierte más de 10,6 millones de interacciones en un tensor tridimensional; la evaluación multimétrica (exactitud, precisión, exhaustividad, F1-Score ponderado y AUC-ROC), la validación cruzada estratificada, el análisis de equidad por subgrupos y la interpretabilidad a través de gradientes, valores SHAP y pesos de atención; así como un experimento de predicción temprana que mide la capacidad del modelo con ventanas de observación reducidas (4, 6 y 8 semanas).

Respecto a lo que queda fuera de sus límites, el estudio no aborda la implementación en producción sobre una plataforma LMS institucional real, ni la validación piloto con cohortes activas o la medición del impacto causal de las intervenciones que se deriven de las predicciones. Tampoco integra fuentes de datos ajenas al OULAD, ni el tratamiento de la clase Withdrawn como categoría predictiva, dado que el abandono representa un fenómeno de índole distinta a la de la calificación final y se plantea como línea de trabajo futuro. El estudio se circunscribe, por tanto, a una prueba de concepto metodológicamente rigurosa y reproducible, orientada a demostrar la viabilidad técnica del enfoque y a sentar las bases para una futura implementación operativa.

1.6 Objetivos

1.6.1 Objetivo general

Desarrollar un sistema predictivo sustentado en redes neuronales profundas de tipo Stacked BiLSTM, orientado a clasificar de forma dinámica el rendimiento académico universitario (reprobado, aprobado y distinción) a partir del procesamiento de los registros temporales de interacción en entornos virtuales de aprendizaje, con el propósito de generar alertas institucionales oportunas.

1.6.2 Objetivos específicos

Con miras a cumplir el objetivo general, se establecen los siguientes seis objetivos específicos:

- Depurar el conjunto de datos OULAD según criterios de calidad y de ética algorítmica.
- Elaborar una representación temporal del comportamiento estudiantil idónea para el modelado secuencial.
- Entrenar la arquitectura Stacked BiLSTM propuesta para la clasificación del rendimiento académico.

- Contrastar su desempeño con clasificadores tradicionales y con arquitecturas secuenciales alternativas.
- Verificar su estabilidad, su equidad entre subgrupos y su capacidad de predicción temprana.
- Interpretar sus predicciones a fin de identificar los momentos del semestre más decisivos para el resultado académico.

Los procedimientos, las técnicas y los parámetros con los que se logra cada uno de estos objetivos se detallan en el Capítulo 3, destinado a la metodología y el desarrollo del trabajo.

1.7 Perfil de la organización fuente de los datos

Esta investigación se sustenta en datos reales procedentes de una institución reconocida a escala mundial en el ámbito de la educación a distancia. Entender el contexto organizacional del cual provienen dichos datos resulta imprescindible para interpretar de forma adecuada los patrones de comportamiento y las decisiones de modelado que se adoptaron.

1.7.1 Nombre y naturaleza de la organización

Los set de datos de esta investigación provienen de The Open University (OU), institución británica creada en 1969 bajo un mandato pionero: democratizar el acceso a la educación superior a través del aprendizaje abierto y a distancia (The Open University, s. f.). No corresponde al modelo de una universidad convencional con campus repletos de aulas presenciales; su funcionamiento se cimienta en la virtualidad, articulando interacciones asíncronas con tutorías distribuidas a gran escala. Su misión consiste en impulsar la apertura hacia las personas, los lugares, los métodos y las ideas, eliminando las barreras socioeconómicas y geográficas que tradicionalmente restringen el acceso al conocimiento universitario. Su visión se orienta a encabezar la innovación pedagógica a nivel global con el

respaldo de las tecnologías digitales, mientras que sus valores se estructuran en torno a la inclusión, la equidad, la innovación permanente y la excelencia académica.

1.7.2 Actividades, servicios y modelo de negocio

La principal actividad central de la OU consiste en ofrecer programas de pregrado, posgrado y educación continua bajo la modalidad de aprendizaje electrónico (e-learning). La totalidad de su oferta se concentra en su propio EVA, una robusta infraestructura tecnológica basada en Moodle, mediante la cual la institución distribuye materiales didácticos multimedia, administra foros de debate, organiza tutorías personalizadas y califica las tareas de los estudiantes (The Open University, s. f.). Su propuesta de valor descansa en una flexibilidad extrema una educación exenta de requisitos formales de admisión previa, y su flujo de ingresos conjuga subsidios estatales para la investigación con el cobro de matrículas modulares. La estructura de costos se halla fuertemente orientada a la tecnología, con una inversión considerable en la optimización del Learning Management System (LMS) y en el mantenimiento de su red de tutores distribuidos.

1.7.3 Ubicación, tamaño y talento humano

La sede administrativa central se encuentra en Milton Keynes, Inglaterra; sin embargo, sus operaciones no tienen fronteras físicas reales, pues sus aulas se despliegan en los hogares de estudiantes distribuidos por todo el Reino Unido y más de cien países. La OU es la universidad más grande del Reino Unido por número de alumnos, con una población activa que supera los 150 000 estudiantes por año académico (The Open University, s. f.). Su ecosistema humano es igualmente masivo: aproximadamente 5000 tutores asociados brindan retroalimentación personalizada y califican las tareas, a lo que se suman más de 4000 empleados de planta entre personal académico, diseñadores instruccionales e ingenieros de datos encargados de mantener la estabilidad del entorno virtual.

1.7.4 Procesos clave y relevancia para el proyecto

El proceso medular de este proyecto es la gestión del aprendizaje y la evaluación continua en el EVA, que comprende el registro sistemático de clics e interacciones, el tratamiento de calificaciones parciales de tipo TMA (evaluaciones corregidas por el tutor) y CMA (cuestionarios computarizados), así como el seguimiento preventivo para poner en marcha tutorías de apoyo antes de las evaluaciones finales. La magnitud de la organización se evidencia en sus bases de datos: los servidores del entorno virtual almacenan más de 10,6 millones de interacciones en un único ciclo de análisis. De ese universo se deriva el conjunto histórico OULAD (Open University Learning Analytics Dataset), que reúne la actividad pormenorizada de 32 593 matrículas repartidas en 22 cohortes académicas y examinadas bajo un esquema temporal de 33 semanas de seguimiento. La OU fue pionera a escala mundial al publicar OULAD mediante una licencia de datos abiertos, lo que hizo posible que investigadores de todo el planeta accedieran a un ecosistema de datos real, de gran volumen y anonimizado, con el fin de entrenar algoritmos avanzados sin comprometer la privacidad de los usuarios (Kuzilek et al., 2017).

1.7.5 Consideraciones éticas, de privacidad y de licenciamiento

El empleo de datos reales de estudiantes demanda un manejo riguroso de las dimensiones ética y jurídica. El conjunto OULAD fue difundido por The Open University tras un proceso de anonimización que reemplaza los identificadores personales por claves numéricas y agrupa la información de manera que ningún estudiante individual pueda ser reidentificado a partir de los registros. Dicha anonimización en el origen suprime los riesgos de privacidad más inmediatos y viabiliza el uso académico de los datos sin necesidad de recabar consentimientos individuales adicionales. Asimismo, el repositorio se distribuye bajo una licencia de datos abiertos que autoriza su uso, reproducción y redistribución con fines investigativos, siempre que se atribuya debidamente la fuente, requisito que la presente investigación satisface en todas sus secciones.

Más allá del cumplimiento formal, la investigación incorporó salvaguardas éticas activas en el propio diseño del modelo. La determinación de excluir la variable de género obedece al principio de no discriminación: un sistema que guíe decisiones sobre personas no debe cimentar sus predicciones en atributos protegidos, aun cuando estos muestren correlación estadística con el resultado (O'Neil, 2016). De igual modo, la inclusión de un análisis de equidad por subgrupos y de técnicas de interpretabilidad atiende al requerimiento de que los sistemas de inteligencia artificial en el ámbito educativo sean transparentes y auditables (Barredo Arrieta et al., 2020). Tales consideraciones no son secundarias: representan una condición de legitimidad para cualquier despliegue institucional de la analítica predictiva, en el cual la protección de la persona debe anteponerse a la simple optimización de una métrica.

CAPÍTULO 2. MARCO TEÓRICO Y ESTADO DEL ARTE

2.1 Estado del arte

La predicción del rendimiento académico y de la deserción estudiantil a través de técnicas de aprendizaje automático representan una de las líneas más dinámicas de la minería de datos educativos (Educational Data Mining, EDM) y de la analítica del aprendizaje. Romero y Ventura (2010, 2020) ordenan la evolución de este campo y evidencian una transición gradual desde los modelos estáticos que condensan el semestre en un vector de características agregadas hacia planteamientos que integran la dimensión temporal del proceso de aprendizaje, justamente el eje sobre el que se estructura la presente investigación. Dicha transición surge de una comprobación empírica: la trayectoria de compromiso de un estudiante encierra una señal predictiva que se desvanece cuando el tiempo se colapsa en un solo agregado.

El conjunto de datos OULAD, difundido por Kuzilek et al. (2017), se ha afianzado como un banco de pruebas de referencia para esta línea de investigación, gracias a su tamaño, a la riqueza de sus interacciones y a su disponibilidad bajo licencia abierta. Diversos estudios lo han utilizado para contrastar algoritmos de clasificación y para indagar en la detección temprana del riesgo. Dentro de los trabajos fundacionales sobre datos de flujo de clics (clickstream), Hussain et al. (2018) evidenciaron que las características extraídas de la actividad en el EVA hacen posible predecir el compromiso del estudiante mediante clasificadores tradicionales, en tanto que Hlosta et al. (2017) formularon el método Ouroboros para detectar estudiantes en riesgo incluso en ausencia de modelos previos, capitalizando la información disponible en las primeras semanas. En una dirección similar, Kotsiantis et al. (2003) fueron precursores en la aplicación de técnicas de aprendizaje automático a la prevención del abandono en la educación a distancia, si bien sustentándose en variables estáticas disponibles al término del período.

El punto de inflexión sobrevino con la adopción del aprendizaje profundo. Waheed et al. (2020) constataron la superioridad de los modelos profundos frente a los clasificadores tradicionales al aprovechar los grandes volúmenes de datos del entorno virtual, sentando un precedente directo para la presente investigación. En fechas más recientes, la investigación se ha orientado hacia la combinación de arquitecturas secuenciales con interpretabilidad: Kalita et al. (2025) plantearon un marco basado en Bi-LSTM con explicaciones a través de SHAP y validación estadística, e informaron exactitudes superiores al 88% junto con una ventaja considerable respecto a algoritmos de boosting como XGBoost, CatBoost y LightGBM. Esta convergencia entre memoria bidireccional e interpretabilidad concuerda por completo con el enfoque asumido en esta tesis.

De manera paralela, la línea de los sistemas de alerta temprana ha trasladado el acento desde la exactitud predictiva hacia el diseño de intervenciones. Bañeres et al. (2023) evidenciaron que las acciones dirigidas al establecimiento de metas sobre las actividades evaluables disminuyen de forma notable la deserción e incrementan el compromiso del estudiante, resaltando que la utilidad de un modelo predictivo se concreta solo cuando posibilita intervenciones oportunas. En resumen, el estado del arte pone de manifiesto tres tendencias convergentes que encuadran esta investigación: la adopción de arquitecturas secuenciales profundas para captar la dinámica temporal del aprendizaje, la creciente demanda de interpretabilidad como requisito para el despliegue responsable, y el traslado del objetivo desde la maximización de la exactitud hacia la utilidad operativa. El presente trabajo se sitúa en la confluencia de estas tres tendencias.

2.2 Marco teórico

2.2.1 Analítica del aprendizaje y minería de datos educativos

La analítica del aprendizaje se conceptualiza como la medición, recopilación, análisis y comunicación de datos relativos a los estudiantes y sus contextos, con el fin de entender y

perfeccionar el aprendizaje y los entornos en los que este tiene lugar (Siemens, 2013). Guarda límites comunes con la minería de datos educativos, si bien esta última concede mayor peso al desarrollo de métodos automáticos para el descubrimiento de patrones (Baker & Inventado, 2014; Romero & Ventura, 2020). Los dos campos parten de la premisa de que las trazas digitales producidas en los EVA albergan señales latentes acerca del compromiso, la motivación y la probabilidad de éxito del estudiante. La noción de compromiso conductual (behavioral engagement), operacionalizada mediante indicadores como la frecuencia de accesos, la regularidad de la actividad y la variedad de recursos empleados, resulta medular para esta investigación, ya que constituye el sustrato observable a partir del cual el modelo aprende. Una línea íntimamente vinculada es el knowledge tracing, cuyo desarrollo con redes recurrentes profundas el Deep Knowledge Tracing de Piech et al. (2015) puso de manifiesto el potencial de modelar la evolución temporal del conocimiento a partir de secuencias de interacción.

2.2.2 Redes neuronales recurrentes y memoria de largo y corto plazo (LSTM)

El aprendizaje profundo ha transformado radicalmente el modelado de datos complejos durante la última década (Goodfellow et al., 2016; LeCun et al., 2015). De forma particular, las redes neuronales recurrentes (RNN) se idearon para procesar datos secuenciales, conservando un estado interno que se actualiza conforme se recorre la secuencia. No obstante, las RNN clásicas padecen el problema del desvanecimiento y la explosión del gradiente: cuando las secuencias exceden unos cuantos pasos temporales, los gradientes calculados durante la retropropagación a través del tiempo (Backpropagation Through Time) disminuyen o aumentan de manera exponencial, lo que obstaculiza el aprendizaje de dependencias de largo alcance. En el marco de este trabajo, con un horizonte de 33 semanas, dicha patología impediría que la red vinculara los hábitos iniciales del estudiante con su desenlace.

Hochreiter y Schmidhuber (1997) superaron esta limitación con la arquitectura Long Short-Term Memory (LSTM), la cual incorpora un estado de celda regulado por tres compuertas: la compuerta de olvido, que regula qué proporción de la memoria histórica se desecha; la compuerta de entrada, que determina qué información nueva se integra; y la compuerta de salida, que fija el estado oculto que se propaga. Este mecanismo hace posible conservar señales relevantes a lo largo de numerosos pasos temporales y ha erigido a la LSTM en una arquitectura de referencia para el modelado de secuencias (Greff et al., 2017). Una variante ulterior, la unidad recurrente cerrada (GRU) de Cho et al. (2014), simplifica el diseño al fusionar compuertas, con un desempeño con frecuencia equiparable.

2.2.3 Redes bidireccionales (BiLSTM) y arquitecturas apiladas

Las redes recurrentes bidireccionales, planteadas por Schuster y Paliwal (1997), recorren la secuencia en ambos sentidos hacia adelante y hacia atrás y fusionan las dos representaciones, de manera que cada punto temporal se interpreta atendiendo tanto a su pasado como a su futuro. La integración de esta idea con las celdas LSTM origina la BiLSTM, cuya eficacia en el modelado secuencial fue acreditada por Graves y Schmidhuber (2005). En el ámbito de la analítica educativa, la bidireccionalidad posibilita que la actividad de una semana concreta se contextualice no solamente con los clics anteriores, sino además con la aceleración o desaceleración ulterior de la actividad, particularmente en la antesala de las evaluaciones finales.

El apilamiento (stacking) de varias capas recurrentes posibilita, adicionalmente, el aprendizaje de representaciones jerárquicas: las capas iniciales captan patrones temporales de bajo nivel, como la fluctuación diaria de clics, en tanto que las capas superiores integran patrones abstractos complejos, como la consistencia del autoaprendizaje o el abandono intermitente (Goodfellow et al., 2016). Es precisamente esta profundidad jerárquica la que

sustenta la elección de una topología Stacked BiLSTM en este trabajo, en contraposición a una única capa recurrente.

2.2.4 Mecanismos de atención y arquitectura Transformer

La arquitectura Transformer, presentada por Vaswani et al. (2017), abandona la recurrencia y se sustenta únicamente en mecanismos de autoatención (self-attention) para modelar las relaciones entre todos los elementos de una secuencia de manera simultánea. La autoatención multi-cabeza posibilita ponderar la relevancia relativa de cada paso temporal frente a los demás, capturando dependencias globales que las arquitecturas recurrentes abordan de forma local y secuencial. En esta investigación, el Transformer desempeña una doble función: actúa como modelo comparativo y, además, sus pesos de atención se aprovechan como recurso de interpretabilidad para reconocer las semanas del semestre que el modelo juzga más informativas, complementando el análisis fundamentado en gradientes.

2.2.5 Métodos de ensamble: Random Forest y XGBoost

Los métodos de ensamble combinan varios modelos base con el fin de generar predicciones más robustas. El Random Forest, planteado por Breiman (2001), agrupa numerosos árboles de decisión entrenados sobre submuestras aleatorias de datos y características, lo que atenúa la varianza y el sobreajuste. XGBoost (Extreme Gradient Boosting), creado por Chen y Guestrin (2016), implementa un esquema de boosting por gradiente que edifica los árboles de manera secuencial cada uno subsanando los errores del precedente y ha evidenciado un rendimiento destacado sobre datos tabulares. En este estudio, ambos métodos operan sobre el tensor temporal aplanado una representación de 231 características estáticas y actúan como referentes rigurosos frente a los cuales se contrasta el valor de la arquitectura secuencial. Su incorporación obedece a un principio metodológico: únicamente al comparar con las mejores alternativas tabulares es posible aislar la contribución real de la memoria temporal.

2.2.6 Desbalance de clases y sobremuestreo sintético (SMOTE)

El desbalance de clases la desproporción entre el número de ejemplos de cada categoría sesga el aprendizaje hacia las clases mayoritarias y constituye un problema recurrente en la analítica educativa (He & Garcia, 2009). La técnica SMOTE (Synthetic Minority Over-sampling Technique), planteada por Chawla et al. (2002), afronta este problema mediante la generación de ejemplos sintéticos de las clases minoritarias a través de la interpolación entre vecinos cercanos reales, en vez de replicar registros existentes; con ello se evita la pérdida de información propia del submuestreo y se disminuye el riesgo de sobreajuste vinculado a la duplicación. De manera complementaria, la ponderación de clases (class weighting) modifica la función de pérdida para sancionar con mayor rigor los errores cometidos sobre las clases infrarrepresentadas. La evaluación de modelos entrenados con datos balanceados requiere, asimismo, métricas sensibles al desbalance, como el F1-Score ponderado, antes que la exactitud simple (Sokolova & Lapalme, 2009).

2.2.7 Regularización y optimización en aprendizaje profundo

El sobreajuste representa un riesgo medular al entrenar redes profundas sobre conjuntos de datos de tamaño moderado. Para atenuarlo se recurre a dos técnicas esenciales: el Dropout, que inhabilita de forma aleatoria una fracción de neuronas durante el entrenamiento con el objeto de inducir representaciones redundantes y robustas (Srivastava et al., 2014), y la Batch Normalization, que normaliza las activaciones de cada capa, agiliza la convergencia y estabiliza el entrenamiento (Ioffe & Szegedy, 2015). La optimización de los pesos se lleva a cabo con el algoritmo Adam (Kingma & Ba, 2015), que aúna las ventajas del momento y de las tasas de aprendizaje adaptativas. A dichas técnicas se añaden callbacks de control, como la reducción adaptativa de la tasa de aprendizaje ante el estancamiento de la métrica de validación y la detención temprana (early stopping), que suspende el entrenamiento cuando el error de validación cesa de mejorar y recupera los mejores pesos (Goodfellow et al., 2016).

2.2.8 Inteligencia artificial explicable (XAI)

La interpretabilidad de los modelos de aprendizaje automático es un requisito creciente en dominios de alto impacto como la educación, donde las decisiones afectan la trayectoria de las personas. Los modelos profundos, en razón de su complejidad, tienden a comportarse como cajas negras, lo que suscita desconfianza entre quienes toman decisiones (Barredo Arrieta et al., 2020; Molnar, 2022). Los valores SHAP (SHapley Additive exPlanations), formulados por Lundberg y Lee (2017), representan un marco unificado, cimentado en la teoría de juegos cooperativos, para asignar a cada característica su contribución precisa a una predicción individual. En el modelado secuencial, técnicas como Gradient × Input hacen posible cuantificar la sensibilidad de la salida del modelo respecto de cada paso temporal de la entrada. La presente investigación emplea de manera complementaria tres técnicas gradientes, valores SHAP y pesos de atención para triangular las semanas más decisivas del semestre, incrementando la solidez de la explicación al tornarla independiente de un único algoritmo.

2.2.9 Metodología CRISP-DM

El Cross-Industry Standard Process for Data Mining (CRISP-DM) es un marco metodológico de amplia difusión que organiza los proyectos de minería de datos en seis fases iterativas: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue (Chapman et al., 2000; Shearer, 2000; Wirth & Hipp, 2000). Su naturaleza cíclica hace posible regresar a fases anteriores cuando los resultados del modelado ponen de relieve la necesidad de depurar los datos, una flexibilidad particularmente valiosa en la analítica del aprendizaje, donde los registros de comportamiento demandan ajustes constantes de las variables predictivas. Su adopción, más allá de estructurar el ciclo de vida del proyecto, brinda un valor pedagógico: exige documentar cada decisión, lo que propicia la transparencia y la reproducibilidad del proceso. Por estos motivos, CRISP-DM se erige en el hilo conductor metodológico de la presente investigación.

CAPÍTULO 3. METODOLOGÍA Y DESARROLLO

3.1 Metodología de ciencia de datos (CRISP-DM)

La ingeniería de datos aplicada a contextos educativos no puede limitarse a una ejecución lineal de algoritmos; requiere un enfoque sistemático capaz de convertir las necesidades del entorno institucional en soluciones algorítmicas validadas. En consecuencia, esta investigación se acoge al modelo CRISP-DM (Cross-Industry Standard Process for Data Mining), afianzado como estándar tanto industrial como académico (Chapman et al., 2000; Shearer, 2000). Su carácter iterativo hace posible regresar a fases anteriores cuando los resultados del modelado evidencian la necesidad de depurar los datos; en la analítica del aprendizaje, dicha flexibilidad resulta esencial, dado que los registros de comportamiento en el EVA se modifican de forma permanente y demandan un ajuste continuo de las variables predictivas (Romero & Ventura, 2020).

El ciclo de desarrollo de este proyecto se organizó de manera rigurosa en las seis fases del estándar, abarcando desde la comprensión del negocio en la que se acota el problema pedagógico hasta la fase de despliegue conceptual del pipeline. La Tabla 1 expone la correspondencia de cada fase con las acciones y los entregables técnicos del proyecto.

Tabla 1

Ciclo iterativo de la metodología CRISP-DM aplicado al entorno virtual de aprendizaje

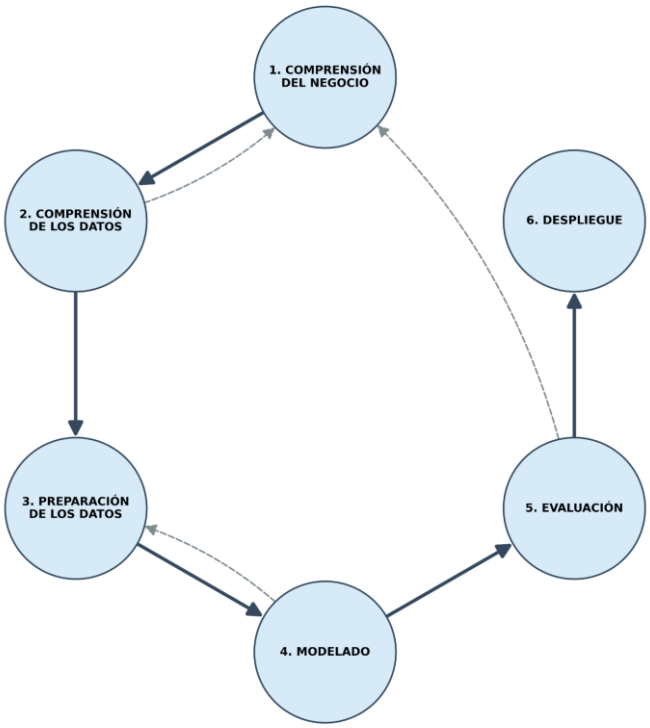
Fase CRISP-DM	Objetivo específico del proyecto	Acciones y entregables técnicos
1. Comprensión del negocio	Identificar la necesidad institucional de alertas tempranas para reducir el fracaso escolar.	Definición de las tres clases excluyentes a predecir: reprobado, aprobado y distinción.
2. Comprensión de los datos	Auditar el ecosistema de datos brutos del repositorio histórico OULAD.	Carga de las tablas nativas, análisis exploratorio de las 32 593 matrículas y caracterización de los 10,6 millones de interacciones.
3. Preparación de los datos	Construir la estructura de entrada idónea para la red secuencial.	Exclusión de variables sesgadas, optimización de memoria, balanceo con SMOTE y estructuración del tensor 3D de 33 semanas.
4. Modelado	Entrenar y optimizar la arquitectura profunda propuesta.	Configuración de la red *Stacked* BiLSTM (capas de 128 y 64 unidades) e implementación de regularización (*Dropout* y *Batch Normalization*).
5. Evaluación	Validar la capacidad predictiva y ejecutar el *benchmarking*.	Comparación de métricas (F1-Score ponderado, exactitud y AUC-ROC) frente a LSTM, *Transformer*, *Random Forest* y XGBoost.
6. Despliegue	Diseñar la lógica de integración para el uso práctico.	Estructuración del *pipeline* escalable para la activación de tutorías oportunas.

Nota. Elaboración propia a partir de las directrices de Chapman et al. (2000) y de los requerimientos del proyecto de titulación.

La adopción de este ciclo iterativo garantizó que el procesamiento de las interacciones semanales no perdiera de vista el objetivo final: generar un clasificador multiclase robusto, capaz de alertar a la administración universitaria antes de que los estudiantes suspendan sus evaluaciones definitivas. La Figura 1 representa gráficamente el ciclo metodológico aplicado al entorno virtual de aprendizaje.

Figura 1

Ciclo iterativo de la metodología CRISP-DM aplicada al entorno virtual de aprendizaje



Nota. Adaptado de CRISP-DM 1.0: Step-by-step data mining guide, por Chapman et al. (2000), SPSS Inc.

3.2 Comprensión y auditoría de los datos

La fase de comprensión de datos en problemas de aprendizaje profundo orientados a la educación no puede limitarse a una lectura pasiva de archivos planos: demanda una evaluación crítica de la integridad y una reestructuración que posibilite construir tensores secuenciales eficientes. En este proyecto, el eje analítico se cimienta en el Open University Learning Analytics Dataset (OULAD), repositorio de ciencia abierta extensamente validado (Kuzilek et al., 2017). La auditoría se articuló en torno a tres procesos: la ingesta del ecosistema de tablas, la mitigación de sesgos éticos y la optimización del consumo de memoria.

3.2.1 Adquisición y carga del ecosistema de datos

El repositorio bruto de OULAD organiza la actividad universitaria en siete archivos con formato CSV. Una vez evaluadas las necesidades de la red neuronal recurrente, el pipeline concentró su atención en tres estructuras fundamentales que recogen la traza de aprendizaje:

studentInfo, studentVle y vle. La carga inicial se llevó a cabo mediante la biblioteca pandas. La Tabla 2 sintetiza las dimensiones estructurales de estos componentes antes de su depuración.

Tabla 2

Volumetría inicial del ecosistema de datos OULAD

Tabla analizada	**Registros totales**	**Columnas nativas**	**Función crítica en el *pipeline* secuencial**
`studentInfo`	32 593	12	Almacena el perfil demográfico, la cohorte y la etiqueta predictiva final (`final_result`).
`studentVle`	10 655 280	6	Captura el registro diario granular de los clics de cada estudiante sobre la plataforma.
`vle`	6 364	6	Funciona como catálogo maestro que indexa las tipologías de recursos didácticos digitales.

Nota. Valores calculados a partir de los scripts de auditoría inicial del proyecto.

3.2.2 Auditoría de calidad, imputación estándar y tratamiento de sesgos éticos

La calidad de las predicciones de una red Stacked BiLSTM está condicionada por la limpieza previa de sus entradas. El análisis de valores faltantes puso al descubierto escenarios heterogéneos: en tanto que los registros transaccionales (studentVle) y el catálogo de recursos (vle) se mostraron completamente limpios, la tabla demográfica studentInfo presentó 1111 valores faltantes en la variable imd_band (índice de privación múltiple), correspondientes al 3,41 % de las matrículas. El examen de estos nulos permitió establecer que pertenecían a estudiantes carentes de dirección postal en el Reino Unido, en razón de su condición de extranjeros. Con el fin de no descartar a esta subpoblación ni introducir sesgos en el entrenamiento, se aplicó una imputación explícita que reemplazó los vacíos por la categoría semántica «Internacional», y se homogeneizaron inconsistencias de formato, unificando registros ambiguos como «10-20» bajo la etiqueta «10-20 %».

Un componente medular de esta fase fue la depuración con criterios de ética algorítmica. Se descartaron las variables region y gender. La supresión de region obedeció a razones analíticas: su elevada cardinalidad no aportaba varianza explicativa relevante y saturaba la matriz de características. La exclusión de gender, en cambio, respondió a una decisión de diseño ético, encaminada a proteger al modelo frente a sesgos de género preexistentes en los datos históricos, asegurando que la red aprenda de patrones de comportamiento y de disciplina académica, y no de atributos demográficos protegidos (O'Neil, 2016), en consonancia con las exigencias actuales de equidad y responsabilidad en el diseño de sistemas de inteligencia artificial (Barredo Arrieta et al., 2020). El cierre de la auditoría constató la ausencia total de valores faltantes en las tres tablas.

3.2.3 Optimización de memoria para el procesamiento secuencial

El procesamiento de secuencias temporales masivas exige una gestión rigurosa del hardware. Al cargar la tabla `studentVle` en su formato nativo, sus más de 10,6 millones de filas requerían un espacio en memoria superior a los 1402 MB, a causa del empleo por defecto de formatos de 64 bits. Esta densidad comprometía la estabilidad del entorno al momento de construir el tensor tridimensional. Para aliviar este cuello de botella se implementó una estrategia de reformateo selectivo (*downcasting*) sustentada en los rangos numéricos reales de las variables: los identificadores `id_student` e `id_site` se convirtieron a enteros de 32 bits; sum_click se limitó a 32 bits, ya que el máximo registrado alcanzó los 24 139 clics, valor que obliga a ese ancho para evitar desbordamientos; y date se optimizó a enteros de 16 bits, puesto que el calendario académico se mueve en un rango acotado de 25 a 269 días. Esta reingeniería redujo el peso de `studentVle` de 1402 MB a 1219 MB, con un ahorro neto de 183 MB (13 % del espacio original), lo que aseguró que el pipeline posterior se ejecutara sin saturar la memoria del sistema.

3.3 Análisis exploratorio de datos (EDA)

Desarrollar un modelo predictivo sustentado en redes recurrentes exige comprender, en primer lugar, la geometría estadística de las variables de entrada. El análisis exploratorio no se circunscribió a una inspección descriptiva: se orientó a caracterizar el desbalance estructural de las clasificaciones académicas y a cuantificar los patrones de comportamiento digital.

3.3.1 Discrepancia volumétrica y desbalance de la variable objetivo

La auditoría exploratoria mostró una divergencia volumétrica muy alarmante. El registro computó 32,593 líneas de matrícula frente a un total neto de 28,785 estudiantes únicos; esta discrepancia aritmética responde a dinámicas de multi-inscripción simultánea en diferentes asignaturas virtuales durante el ciclo 2013–2014, descartando fallos de redundancia relacional y obligando al pipeline a fijar su unidad de observación en el par matrícula-cohorte. Esta configuración matricial garantiza aislar trayectorias de aprendizaje independientes para el modelo secuencial. Al mapear la variable objetivo `final_result`, sus cuatro categorías nativas exhibieron una severa asimetría estadística en la prevalencia de etiquetas. La Tabla 3 desglosa esta dispersión cuantitativa.

Tabla 3

Distribución original de la variable objetivo-predictiva

Tabla 3.3

Distribución original de la variable objetivo-predictiva

Categoría académica (`final_result`)	**Frecuencia absoluta**	**Proporción (%)**	**Implicación para el modelado profundo**
Pass (aprobado)	12 361	37,93 %	Dominancia volumétrica absoluta; establece el sustrato conductual estándar del ecosistema virtual.
Withdrawn (retirado)	10 156	31,16 %	Vector de deserción crítica; impera el diseño de algoritmos de detección prematura en fases iniciales.
Fail (reprobado)	7 052	21,64 %	Vulnerabilidad académica severa; diagnostica el colapso transaccional previo a los mínimos obligatorios.
Distinction (distinción)	3 024	9,28 %	Asimetría estadística extrema; indexa perfiles atípicos con dinámicas de interacción excepcionales.

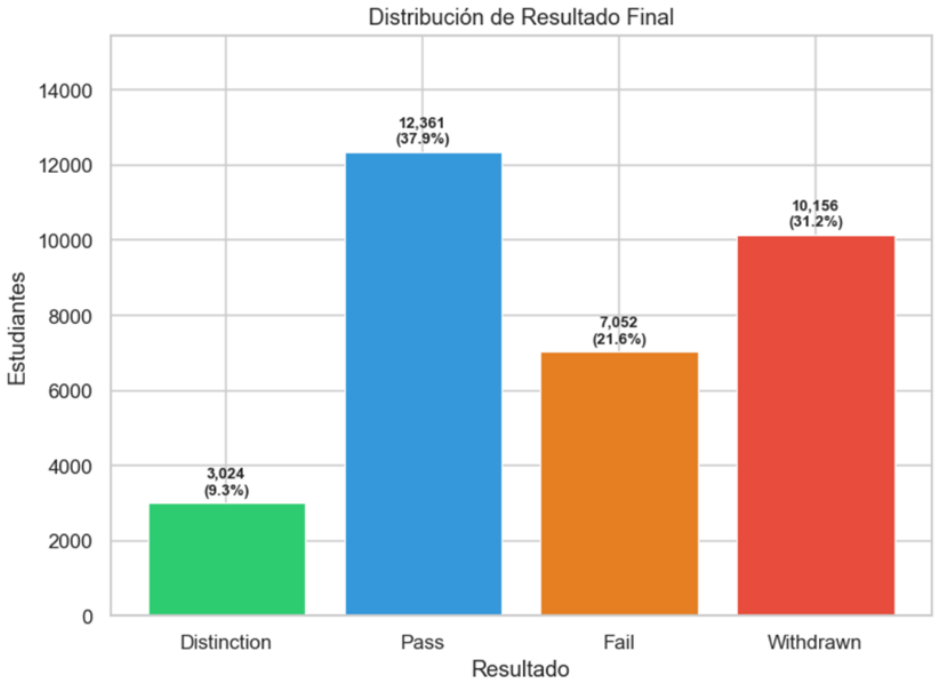
Nota. Datos calculados a partir de la tabla `studentInfo` del repositorio OULAD.

La asimetría de la muestra donde la cohorte de aprobados supera de forma triple al grupo de excelencia condicionó el pipeline, obligando a integrar técnicas de sobremuestreo sintético durante la optimización. Omitir esta reingeniería dimensional habría provocado que la función de coste minimizara la entropía cruzada global mediante un sesgo de colapso hacia la clase mayoritaria. La capacidad preventiva habría colapsado. Para neutralizar esta distorsión métrica y siguiendo los consensos teóricos sobre minería de datos desbalanceados, el diagnóstico del clasificador profundo descartó la exactitud global en favor del F1-Score ponderado, evitando el espejismo estadístico de los entornos asimétricos (He & Garcia, 2009; Sokolova & Lapalme, 2009). Paralelamente, la arquitectura adoptó un cambio de diseño estructural crítico: segregamos la etiqueta *Withdrawn* debido a su naturaleza netamente

administrativa y exógena al progreso de competencias, acotando el clasificador a un espacio puramente triclase (*Fail, Pass y Distinction*). La Figura 2 expone el mapa porcentual de este vector objetivo.

Figura 2

Distribución de resultado final de clases de la variable de rendimiento académico final



Nota. Elaboración propia obtenida de las métricas de distribución en la auditoría del repositorio.

3.3.2 Dinámica del tráfico e interacciones en el entorno virtual

Los trazos secuenciales de `studentVle` sostienen el andamiaje predictivo del pipeline. Al contabilizar este universo de más de 10.6 millones de registros crudos, saltó a la vista una polarización extrema en la topología transaccional del EVA. El tráfico web no tiene simetría. Dos componentes específicos canibalizan la densidad operativa de la plataforma. Los contenidos curriculares nucleares (`outcontent`) centralizaron un flujo masivo de 11.2 millones de clics, explicitando los ciclos de asimilación autónoma y lectura de base. En una dimensión paralela, los espacios de interacción social y debate colaborativo (`forumng`) absorbieron una traza robusta próxima a los 7.9 millones de clics. La brecha frente a módulos

marginales fundamentó nuestra decisión de estructurar el vector de características mediante agregaciones cronológicas semanales. Esta compactación erradica el ruido estocástico. Al alimentar la red profunda con estas señales transaccionales de alta varianza, el modelo aísla vectorialmente la frontera entre las rutinas estables de engagement pedagógico y el desenganche progresivo que precede a la deserción.

3.4 Preparación de datos y construcción del tensor secuencial

La transición desde bases de datos relacionales hacia modelos de aprendizaje profundo requiere una reingeniería de características que preserve la dimensión temporal. Las arquitecturas BiLSTM no procesan registros estáticos, demandan matrices tridimensionales ordenadas cronológicamente, en consonancia con la tendencia reciente a explotar la dimensión temporal del comportamiento en el EVA mediante aprendizaje profundo (Kalita et al., 2025; Waheed et al., 2020). La preparación se centró en dos hitos, tales como la vectorización del comportamiento en un horizonte fijo de 33 semanas y la resolución del desbalance mediante sobremuestreo sintético.

3.4.1 Agregación temporal y estructuración del tensor tridimensional

Dado que las interacciones en el EVA destacan por ser asíncronas y presentar una alta variabilidad temporal. Para convertir este flujo disperso en una señal analítica determinista, discretizamos secuencialmente el calendario mediante ventanas temporales estáticas de siete días, cubriendo desde el registro preliminar hasta el cierre definitivo de la cohorte académica, consolidando un frente cronológico de T = 33 intervalos, cubriendo más del 95% del tráfico web absoluto. La agregación semanal requirió diseñar una clave compuesta multivariable basada en la tupla (`id_student, code_module, code_presentation`), excluyendo el uso aislado de `id_student` para evitar cualquier contaminación cruzada de clics entre aulas virtuales concurrentes y así el pipeline creó así la topología de entrada definitiva. Estructuramos un tensor tridimensional con geometría explícita de Muestras x Pasos de tiempo

x Características (N x T x M), donde el volumen neto unifica N = 22,437 trayectorias de matrículas activas excluyendo el sesgo operativo de los retiros prematuros (*Withdrawn*) sincronizadas sobre un eje de T = 33 semanas y decodificadas por M = 7 descriptores. Esta firma combina tres señales dinámicas transaccionales con cuatro métricas estáticas extraídas del historial de calificaciones parciales, la cuales se exponen en la Tabla 4.

Tabla 4

Características del tensor temporal (eje de características)

#	Característica	Tipo	Descripción
1	`clicks`	Dinámica	Total de clics semanales en el EVA.
2	`active_days`	Dinámica	Cuantifica las jornadas independientes con huella digital detectable en el ciclo.
3	`interactions`	Dinámica	Discrimina la frecuencia semanal de consumo activo de recursos indexados.
4	`tma_mean`	Estática	Promedia el desempeño en entregas complejas baremadas por el tutor.
5	`tma_total`	Estática	Acumula la puntuación bruta en evaluaciones síncronas.
6	`cma_mean`	Estática	Establece la media aritmética de aciertos en cuestionarios informáticos.
7	`cma_total`	Estática	Suma el puntaje neto logrado en componentes de calificación automática.

Nota. Las características dinámicas varían semana a semana; las estáticas se replican en todas las semanas de la matrícula. Se incluyeron las notas de las evaluaciones continuas TMA (calificadas por el tutor) y CMA (computarizadas): 20 633 matrículas registraron TMA (nota media de 66,7/100) y 12 862 registraron CMA. Elaboración propia.

Para asegurar una convergencia eficiente durante el descenso de gradiente, todas las características se sometieron a una transformación de escala ajustada exclusivamente sobre el conjunto de entrenamiento, con el fin de prevenir la fuga de información hacia el conjunto de

prueba. Los estudiantes sin registro alguno en el EVA recibieron el valor cero, decisión deliberada para preservar la señal de inactividad en lugar de imputarla con la media.

3.4.2 Tratamiento del desbalance mediante SMOTE

Tal como se evidenció en el análisis exploratorio, la severa disparidad entre la clase minoritaria y la mayoritaria comprometía el entrenamiento del clasificador. Para resolver esta asimetría sin perder información de los registros reales, se aplicó la técnica SMOTE (Chawla et al., 2002), un procedimiento de sobremuestreo cuya eficacia para mitigar el desbalance en la predicción del rendimiento estudiantil a partir de datos de entornos virtuales de aprendizaje ha sido corroborada en investigaciones recientes con modelos secuenciales (Ismanto et al., 2024). Se aplicó el sobremuestreo unicamente sobre el set de entrenamiento, tras la división estándar de 80/20 (17 949 matrículas de entrenamiento y 4488 de prueba). El set de prueba se mantuvo intacto, con su distribución real para que la evaluación reflejara el rendimiento en condiciones auténticas. La Tabla 5 cuantifica y muestra el impacto del balanceo.

Tabla 5

Impacto del algoritmo SMOTE en el balanceo del conjunto de entrenamiento

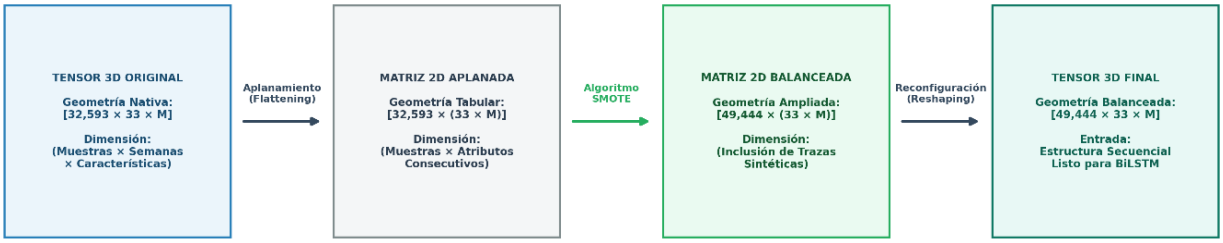
Clase académica	Registros de entrenamiento (reales)	Registros tras SMOTE	Estado
Pass (aprobado)	9 889	9 889	Clase mayoritaria; preservada como referencia.
Fail (reprobado)	5 641	9 889	Se amplía acorde a la interpolación de trazas.
Distinction (distinción)	2 419	9 889	Se cuadriplica para equiparar su peso estadístico.
Total	**17 949**	**29 667**	Conjunto de entrenamiento balanceado.

Nota. El algoritmo SMOTE generó 11 718 muestras sintéticas por interpolación entre vecinos reales, sin duplicar registros. En la representación secuencial (tensor), el desbalance se compensó mediante calificación de clases (`class_weight = 'balanced'`), estrategia más adecuada para el entrenamiento de redes recurrentes al castigar de forma diferenciada los errores sobre la clase minoritaria (He & Garcia, 2009). Elaboración propia a partir de Chawla et al. (2002).

Esta simetría inducida en el espacio de etiquetas reconfigura radicalmente la dinámica del gradiente de optimización y la función de coste ya no tolera sesgos de dominancia. Bajo este esquema balanceado, el pipeline aplica un rigor punitivo idéntico ante desviaciones predictivas en los polos opuestos del espectro académico, obligando al algoritmo a calibrar con la misma minuciosidad matemática los vectores de desempeño excepcional que las trayectorias críticas al borde de la reprobación. La Figura 3 sintetiza el *pipeline* de transformación dimensional y balanceo.

Figura 3

Pipeline de transformación dimensional y balanceo sintético de trayectorias estudiantiles temporales



Nota. Elaboración propia en base al flujo de preprocesamiento de tensores del proyecto.

3.5 Modelado y arquitectura de aprendizaje profundo

La predicción del resultado a lo largo de un período de 33 semanas constituye un problema de dependencias temporales de largo alcance. En este proyecto se implementa una arquitectura basada en redes recurrentes, las cuales son celdas de memoria de largo y corto plazo bidireccionales apiladas (*Stacked* BiLSTM), para abordar el desafío.

3.5.1 Mecanismo de compuertas de celda LSTM y ventaja bidireccional

Las redes recurrentes ordinarias sufren pérdida del gradiente cuando las secuencias superan pocos pasos temporales. La celda LSTM (Hochreiter & Schmidhuber, 1997) resuelve esta patología mediante un estado de celda regulado por tres compuertas, cuyas ecuaciones se resumen a continuación:

```
f_t = sigmoide(W_f · [h_(t-1), x_t] + b_f)      (compuerta de olvido)
i_t = sigmoide(W_i · [h_(t-1), x_t] + b_i)      (compuerta de entrada)
o_t = sigmoide(W_o · [h_(t-1), x_t] + b_o)      (compuerta de salida)
C_t = f_t * C_(t-1) + i_t * tanh(W_C · [h_(t-1), x_t] + b_C)
h_t = o_t * tanh(C_t)
```

La compuerta de olvido modula qué proporción de la memoria histórica se descarta; la de entrada decide qué patrones de la semana corriente se incorporan al estado de la celda; y la de salida determina el estado oculto que se propaga, mientras que una LSTM convencional procesa la secuencia de forma estrictamente lineal, la variante bidireccional (Graves & Schmidhuber, 2005; Schuster & Paliwal, 1997) entrena simultáneamente dos hilos recurrentes de manera independiente, uno hacia adelante y otro inverso y concatena sus representaciones. En analítica educativa, esto permite condicionar la predicción de una semana con los clics pasados y contextualizándola también con la aceleración o desaceleración futura de la actividad en etapas finales de las evaluaciones.

3.5.2 Especificación de la topología Stacked BiLSTM propuesta

Para obtener representaciones jerárquicas abstractas, el *pipeline* elije una configuración apilada, elección que se respalda por el fundamento del apilamiento de capas recurrentes para el aprendizaje de representaciones profundas (Graves & Schmidhuber, 2005) como por la evidencia reciente sobre la eficacia de las arquitecturas BiLSTM en la predicción del rendimiento académico (Kalita et al., 2025). Las primeras capas capturan características temporales de bajo nivel, mientras que las superiores sintetizan patrones abstractos como la consistencia del autoaprendizaje. La arquitectura exacta se estructuró según la Tabla 6.

Tabla 6

Configuración de capas y parámetros de la red neuronal Stacked BiLSTM

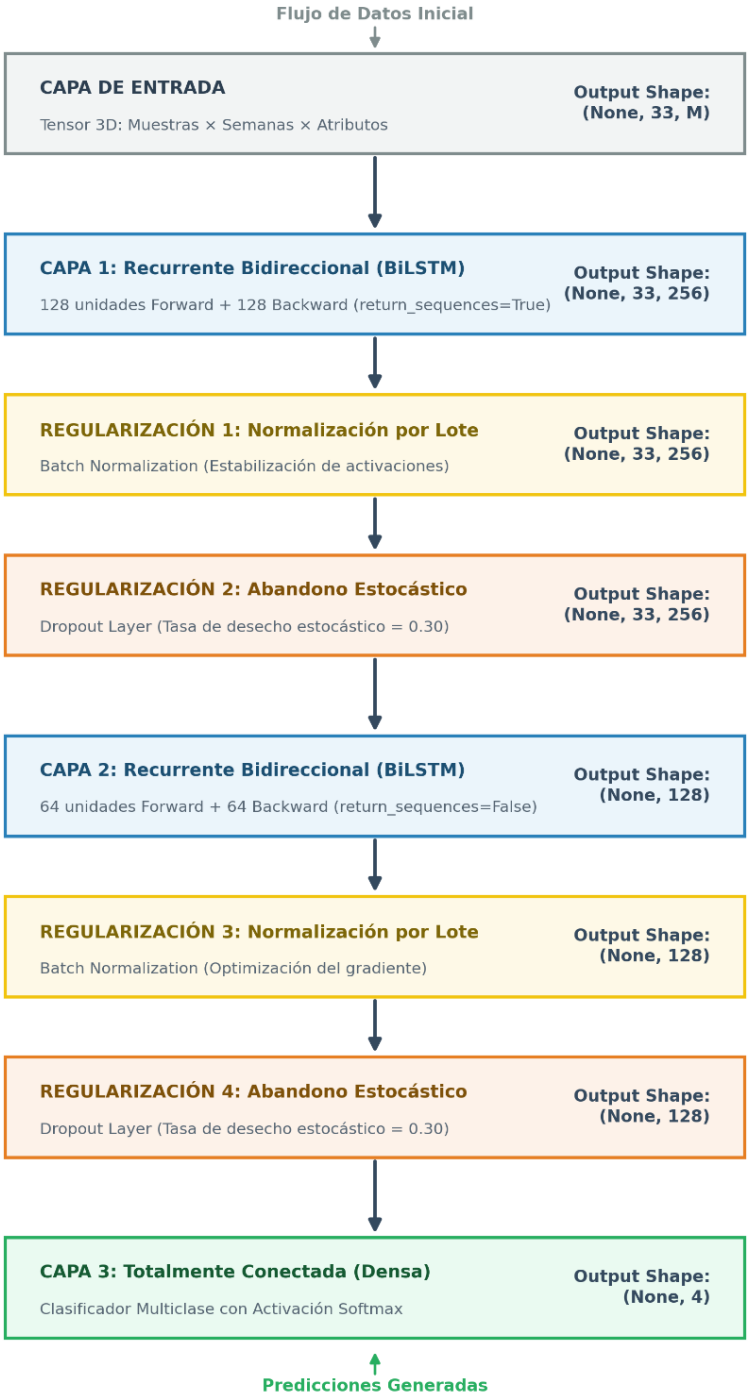
Índice	Capa computacional	Salida (*shape*)	Hiperparámetros y regularización
0	Entrada (*Input*)	(None, 33, 7)	Recibe el tensor tridimensional de características semanales.
1	*Bidirectional*(LSTM)	(None, 33, 256)	128 unidades por dirección; retorna la secuencia completa (`return_sequences=True`).
2	*Dropout*	(None, 33, 256)	Tasa del 30 % (0,3); mitiga el sobreajuste.
3	*BatchNormalization*	(None, 33, 256)	Estabiliza las activaciones y acelera la convergencia.
4	*Bidirectional*(LSTM)	(None, 128)	64 unidades por dirección; consolida el vector final (`return_sequences=False`).
5	*Dropout*	(None, 128)	Tasa del 30 % (0,3).
6	*Dense* (ReLU)	(None, 32)	Capa totalmente conectada con activación ReLU.
7	*BatchNormalization*	(None, 32)	Normalización previa a la capa de salida.
8	*Dense* (*Softmax*)	(None, 3)	Genera la distribución de probabilidad para las tres clases de rendimiento.

Nota. La capa de salida contiene tres neuronas, alineadas con las clases *Fail*, *Pass* y *Distinction* tras la exclusión de *Withdrawn*. El uso de `return_sequences=True` en la primera capa recurrente es imperativo para la estructura apilada: sin propagar las dimensiones temporales, la segunda capa recibiría un vector plano y perdería la capacidad de un análisis recurrente secundario. Elaboración propia a partir de la implementación del modelo.

La Figura 4 presenta el diagrama de bloques y el flujo de tensores de la arquitectura profunda.

Figura 4

Diagrama de bloques y flujo de tensores de la arquitectura profunda Stacked BiLSTM propuesta



Nota. Elaboración propia basada en la implementación del modelo secuencial en el entorno de desarrollo.

3.5.3 Compilación, callbacks de control y clasificadores de benchmarking

La red se compiló con la función de pérdida de entropía cruzada categórica dispersa (`sparse_categorical_crossentropy`), apropiada para etiquetas enteras, y el optimizador Adam (Kingma & Ba, 2015), con una tasa de aprendizaje inicial de 1×10^{-3}. El

entrenamiento se parametrizó para un máximo de 60 épocas, con un tamaño de lote de 256 y una fracción de validación del 15 %. La mitigación de la inestabilidad en la convergencia sinóptica exigió el anclaje de dos callbacks ortogonales en el bucle de optimización. Primero, inyectamos *ReduceLROnPlateau*, esta subrutina atenúa dinámicamente la tasa de aprendizaje por un factor de 0.5 tras constatar una meseta de cuatro épocas consecutivas sin reducciones en el coste de validación, acotando su límite inferior estricto en $1\ x\ 10^{-6}$ para hacer viable un descenso de gradiente ultra-fino. Segundo, implementamos *EarlyStopping* y configuramos su paciencia en ocho épocas, este mecanismo intercepta el entrenamiento ante un estancamiento definitivo del error de validación, restaurando la matriz de pesos con mejor desempeño histórico y purgando la memorización de ruido estocástico en el tensor. Convalidar el impacto intrínseco de nuestra propuesta secuencial requirió una confrontación paralela frente a cuatro líneas base (benchmarks) entrenadas bajo idénticos splits de datos. El pipeline ejecutó de forma simultánea una red LSTM unidireccional, una arquitectura Transformer estructurada sobre bloques de autoatención (self-attention), un bosque aleatorio (Random Forest) y un clasificador XGBoost, cada variante descompone una decisión de diseño metodológico. La red LSTM aísla directamente la ventaja del apilamiento jerárquico y la bidireccionalidad temporal. Por su parte, el Transformer evalúa el potencial de la atención global sin indexación lineal frente al almacenamiento selectivo de las celdas recurrentes y finalmente las dos aproximaciones basadas en árboles operan sobre la geometría linealizada del volumen de datos, explotando un hiperplano de 231 características estáticas aplanadas. Este contraste permite cuantificar el beneficio de preservar estructuras tridimensionales cronológicas continuas frente a representaciones tabulares estáticas que destruyen la línea de tiempo, un diseño en sintonía con las evidencias contemporáneas sobre la superioridad del deep learning secuencial sobre logs del EVA (Waheed et al., 2020).

CAPÍTULO 4. ANÁLISIS DE RESULTADOS

La resolución matemática del horizonte de anticipación crítica constituye el eje neurálgico del pipeline. Se Evaluó la celeridad diagnóstica. Este bloque separa minuciosamente las métricas macro calculadas tras la estabilización de los pesos sinópticos de la arquitectura propuesta *Stacked* BiLSTM. Confrontamos empíricamente su robustez frente al ecosistema de clasificadores de control. Para blindar los hallazgos contra el sesgo de partición fortuita, implementamos una validación cruzada estratificada. Finalmente, interceptamos la opacidad inherente al deep learning integrando algoritmos de explicabilidad avanzada (XAI) orientados a desentrañar la sensibilidad temporal de las ventanas académicas.

4.1 Entrenamiento y convergencia del modelo principal

La optimización sinóptica de la arquitectura *Stacked* BiLSTM exhibió una convergencia acelerada, el bucle de entrenamiento truncó su ejecución en la época 14 de las 60 parametrizadas originalmente, consumiendo un coste computacional estricto de 136 segundos, ocasionando que el hardware opere sin fricciones. Las trayectorias dinámicas de las funciones de coste (loss) y exactitud (accuracy) en los sets de entrenamiento y validación coexistieron de forma cuasi-idéntica, descartando cualquier acoplamiento al ruido o desvanecimiento prematuro del gradiente. El activador *EarlyStopping* forzó la interrupción y restauró los coeficientes sinópticos óptimos, para blindar empíricamente esta topología, exploramos el espacio hiperparamétrico mediante una búsqueda bayesiana impulsada por *Keras Tuner* (seis ejecuciones discretas completadas en 237 segundos). Sorprendentemente, expandir el plano estructural a 256 unidades en la primera capa recurrente deprimió el F1-Score ponderado a 0.7783, situándose por debajo del 0.7907 arrojado por nuestro diseño base de 128 celdas. La saturación dimensional penalizó el rendimiento, este comportamiento convalida empíricamente que robustecer artificialmente la capacidad interna de las celdas BiLSTM

induce sobreajuste localizado en logs de aprendizaje virtual, legitimando el dimensionamiento original del pipeline.

4.2 Rendimiento temporal y capacidad de alerta temprana

La rentabilidad operativa de una arquitectura de *Learning Analytics* va más allá de la exactitud estática o forense calculada al cierre de un ciclo. Exige convergencia cronométrica oportuna. Un clasificador supervisado con precisión matemática absoluta que emita su diagnóstico en la última semana académica anula cualquier viabilidad de maniobra pedagógica; carece de utilidad real para la gestión. Esta premisa obligó a evaluar dinámicamente el rendimiento a lo largo del total secuencial de 33 semanas. El comportamiento predictivo prematuro rebasó las expectativas. Al procesar exclusivamente la traza transaccional compilada hasta la semana 4 del calendario escolar, el pipeline profundo consolidó un F1-Score ponderado de 0.7619. La brecha frente al modelo alimentado con el ciclo completo se redujo a una fracción marginal de 0.0288 puntos.

Definir el destino del alumno observando apenas un mes de logs de conectividad virtual constituye el vector estratégico de mayor impacto en este estudio. Abrimos una ventana de contingencia proactiva de 29 semanas. Este espacio cronológico faculta a los departamentos de retención estudiantil para activar esquemas de contención quirúrgica mentorías sincrónicas adaptativas, recursos dosificados de microaprendizaje y soporte psicopedagógico dirigido mucho antes de que la desconexión sistemática devenga en un retiro administrativo o consume la reprobación definitiva de la matrícula. La Tabla 7 parametriza las fronteras métricas de estos umbrales operativos.

Tabla 7

Alerta temprana a umbrales operativos según el rendimiento temporal del BiLSTM

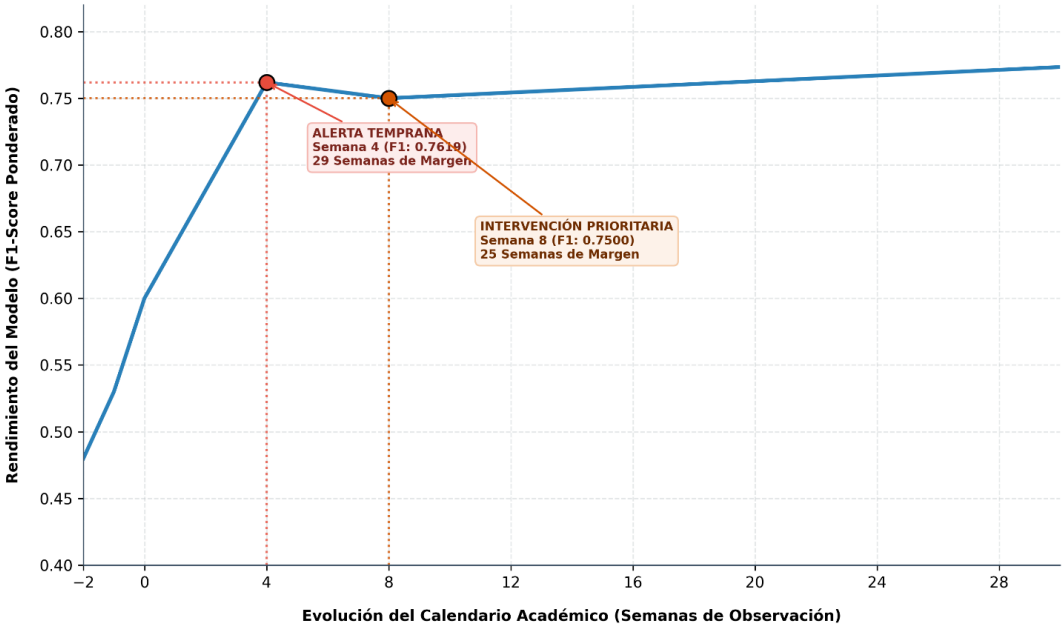
Hito temporal	Margen de acción institucional	Rendimiento (F1-Score)	Acción estratégica sugerida en el LMS
Semana 4	29 semanas de margen	0,7619	Alerta temprana: notificaciones y envío de recursos de autoestudio preventivo.
Semana 8	25 semanas de margen	0,7500	Umbral de intervención prioritaria: asignación de tutor académico y mentorías sincrónicas.

Nota. Elaboración propia diseñados a partir de los puntos de inflexión predictiva observados en el *pipeline* de evaluación secuencial.

La meseta predictiva registrada entre el primer y el segundo mes de observación devela una estabilidad robusta en el vector de características. El F1-Score varió marginalmente, específicamente el clasificador osciló de un valor de 0.7619 en la semana 4 a un 0.7500 al alcanzar el hito cronológico de la semana 8; esta inflexión a la baja no acusa una degradación cognitiva en el aprendizaje de las capas ocultas de la red, sino la intrusión de ruido estocástico transaccional espoleado por el primer bloque de evaluaciones sumativas. En este periodo de fricción, las trayectorias secuenciales de los futuros reprobados (*Fail*) experimentan un solapamiento topológico transitorio con la señal de desenganche de los estudiantes propensos al retiro administrativo (*Withdrawn*). La bidireccionalidad recursiva neutralizó esta mimetización. Al procesar las secuencias de clics concurrentemente hacia adelante y hacia atrás en el tiempo, la red profunda contextualiza las caídas de interactividad digital, distinguiendo con alta fidelidad las anomalías predictivas de abandono respecto a los ceses intermitentes legítimos por vacaciones, lo que abate la tasa de falsos positivos. La Figura 5 mapea este análisis dinámico de convergencia.

Figura 5

Evolución del F1-Score y análisis de la convergencia predictiva para alertas tempranas



Nota. Elaboración propia a partir de la obtención de matrices de evaluación temporal ejecutadas en el proyecto.

4.3 Validación cruzada y estabilidad del modelo

Restringir la estimación del error a un split estático introduce el sesgo de partición espuria. Evitamos este peligro metodológico. Con la finalidad de blindar la estabilidad del pipeline y salvaguardar la reproducibilidad institucional de los hallazgos, encapsulamos el modelo secuencial dentro de un protocolo de validación cruzada estratificada de cinco pliegues (k = 5). La iteración barrió recursivamente la geometría del tensor balanceado, manteniendo un control estricto sobre la prevalencia nativa de las etiquetas en cada subconjunto. Este blindaje experimental disipó anomalías por segmentaciones fortuitas. Al consolidar vectorialmente las métricas de generalización calculadas, el clasificador profundo estabilizó un F1-Score ponderado promedio de 0.7734. La Tabla 8 pormenoriza el comportamiento algebraico granular arrojado en cada pliegue de prueba.

Tabla 8

Rendimiento del clasificador Stacked BiLSTM por pliegue de validación cruzada

Ciclos de validación	Set de entrenamiento	Set de validación	F1-Score ponderado
Pliegue 1	23 733	5 934	0,7692
Pliegue 2	23 733	5 934	0,7812
Pliegue 3	23 734	5 933	0,7756
Pliegue 4	23 734	5 933	0,7641
Pliegue 5	23 734	5 933	0,7769
Promedio	N/A	N/A	**0,7734 ± 0,0076**

Nota. Desglose de las iteraciones de control ejecutadas para la validación del modelo. Elaboración propia.

Más allá de la convergencia media, la solidez del clasificador radica en la estrechez crítica de su dispersión. La variabilidad quedó confinada en una desviación estándar de apenas 0.0076. Computamos una fluctuación marginal inferior al 1% entre pliegues. Semejante estabilidad matemática certifica la invariancia topológica de la red profunda frente a perturbaciones en la partición muestral; el éxito predictivo rehúsa depender de repartos fortuitos o distribuciones benévolas de los datos. Esta homogeneidad estructural neutraliza simultáneamente dos patologías críticas de los modelos de aprendizaje profundo. Descarta de raíz cualquier fuga latente de información (data leakage) derivada de la interpolación sintética con SMOTE la cual habría provocado picos severos de varianza por sobreajuste localizado y expone la resiliencia intrínseca de las capas de regularización estocástica. El pipeline queda metodológicamente validado para su migración hacia entornos de producción en tiempo real (Bergstra & Bengio, 2012).

4.4 Análisis de equidad entre subgrupos

Se evaluó el rendimiento del modelo de forma diferenciada por módulo académico y por modalidad de estudio, el F1-Score varió entre 0,725 (módulo más difícil) y 0,875 (módulo más predecible), una brecha de 0,15 puntos. La segregación operacional por régimen de estudio desveló una brecha predictiva asimétrica. El clasificador profundo modeló con mayor solidez

a la cohorte matriculada a tiempo completo, donde consolidó un F1 = 0.797, frente al segmento de dedicación parcial que cayó a un F1 = 0.752. Evaluamos un diferencial neto de 0.045 puntos. Este desajuste mapea una restricción intrínseca del pipeline. La fidelidad diagnóstica rehúsa ser uniforme. La marcada dispersión transaccional en las trazas de los alumnos a tiempo parcial cuyas interacciones semanales operan bajo patrones severamente erráticos erosiona la capacidad predictiva de las capas BiLSTM al mimetizar el desenganche con la simple asincronía. Auditar sistemáticamente estas asimetrías subgrupales representa un requisito técnico y deontológico ineludible para viabilizar un despliegue algorítmico responsable en la educación superior (Baker & Inventado, 2014; Romero & Ventura, 2020).

4.5 Análisis de interpretabilidad temporal (IA explicable)

Las arquitecturas profundas aplicadas a las ciencias sociales sufren el cuestionamiento de operar como cajas negras (Barredo Arrieta et al., 2020). En el diseño de políticas universitarias, esta opacidad representa un riesgo, los tutores no pueden estructurar intervenciones si desconocen qué momentos del tiempo gatillaron la alerta y para mitigar esta brecha, la investigación implementó una estrategia de triangulación basada en tres métodos independientes de IA explicable, que confrontó los gradientes de sensibilidad del *Stacked* BiLSTM, los valores SHAP del *Random Forest* y las matrices de atención del *Transformer*.

La convergencia topológica entre los perfiles de sensibilidad del gradiente en la red *BiLSTM* y las atribuciones locales SHAP mapeadas por el bosque aleatorio resultó incontestable. Ambos enfoques correlacionaron vectorialmente. Fijaron de forma unánime las semanas 29, 30, 31 y 32 como el clúster de máxima cargabilidad predictiva en el ciclo transaccional. Curricularmente, este pico refracta la fase de estrangulamiento académico terminal, un periodo condicionado por la saturación de entregas definitivas y evaluaciones de cierre. El *Transformer* aportó disrupción estructural. Sus capas de autoatención (self-attention) decodificaron dependencias temporales de rango medio, aislando de forma autónoma dos

ventanas críticas en las semanas 11 y 27 que, indexadas sobre el calendario histórico de OULAD, se solapan milimétricamente con los hitos de evaluación sumativa intermedia. La Tabla 9 esquematiza esta convergencia multimodelo.

Tabla 9

Triangulación de las semanas críticas según el método de interpretabilidad

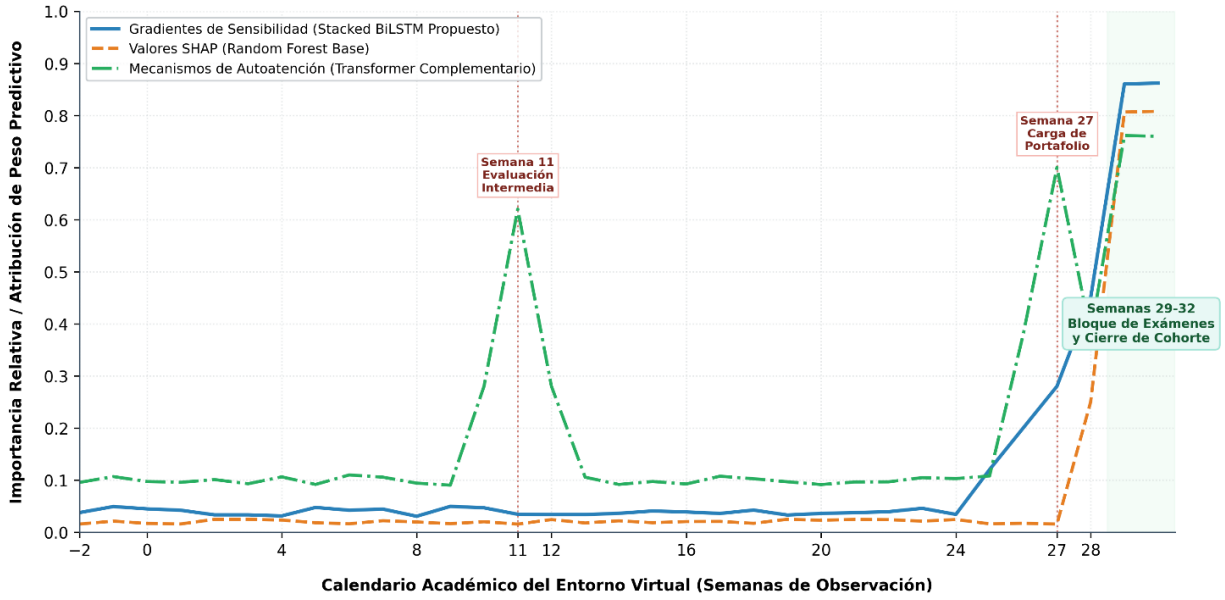
Método de interpretabilidad	**Semanas determinantes**
Gradientes de sensibilidad (*Stacked* BiLSTM)	32, 31, 30, 29, 0
Valores SHAP (*Random Forest*)	30, 31, 32 (y adyacentes)
Pesos de autoatención (*Transformer*)	11, 27, 26

Nota. Elaboración propia, la coincidencia entre gradientes y SHAP entre las semanas 29 y 32, obtenidas con arquitecturas distintas, evidencia robustes e independizaje del algoritmo.

La detección automatizada de estos periodos da a entender que los modelos secuenciales no memorizan ruido estadístico, sino que replican la estructura y el ritmo del calendario académico formal, esta triangulación valida empíricamente los umbrales operativos propuestos y el sistema no solo predice con precisión, sino que fundamenta sus alertas en los momentos de mayor fricción pedagógica (Lundberg & Lee, 2017). La Figura 6 presenta la triangulación metodológica de la importancia temporal.

Figura 6

Triangulación metodológica temporal mediante gradientes de sensibilidad, valores SHAP y mecanismos de autoatención



Nota. Elaboración propia a partir de las matrices obtenidas de la explicabilidad e interpretabilidad computacional (XAI) del proyecto.

4.6 Comparación global de modelos y benchmarking computacional

Autenticar la solidez de una arquitectura profunda exige un careo empírico hostil frente a clasificadores alternativos, para esto sincronizamos las restricciones de cómputo. El pipeline sometió a la red *Stacked BiLSTM* a una confrontación directa y paralela ante cuatro benchmarks de control, una LSTM unidireccional, una arquitectura Transformer equipada con atención global, un bosque aleatorio (Random Forest) y un clasificador XGBoost, garantizando la paridad absoluta en los splits del set de muestra y evaluamos la respuesta predictiva. La Tabla 10 indexa la síntesis macro de esta matriz de rendimiento.

Tabla 10

Comparativas globales de rendimiento entre todos los modelos utilizados

Arquitectura evaluada	Exactitud	Precisión	Sensibilidad	F1-Score ponderado
XGBoost	0,8119	0,8148	0,8119	0,8075
Random Forest	0,8019	0,8104	0,8019	0,8031
Stacked BiLSTM (propuesto)	0,7839	0,8066	0,7839	0,7907
LSTM unidireccional	0,7520	0,7957	0,7520	0,7624
Transformer	0,7331	0,7794	0,7331	0,7423

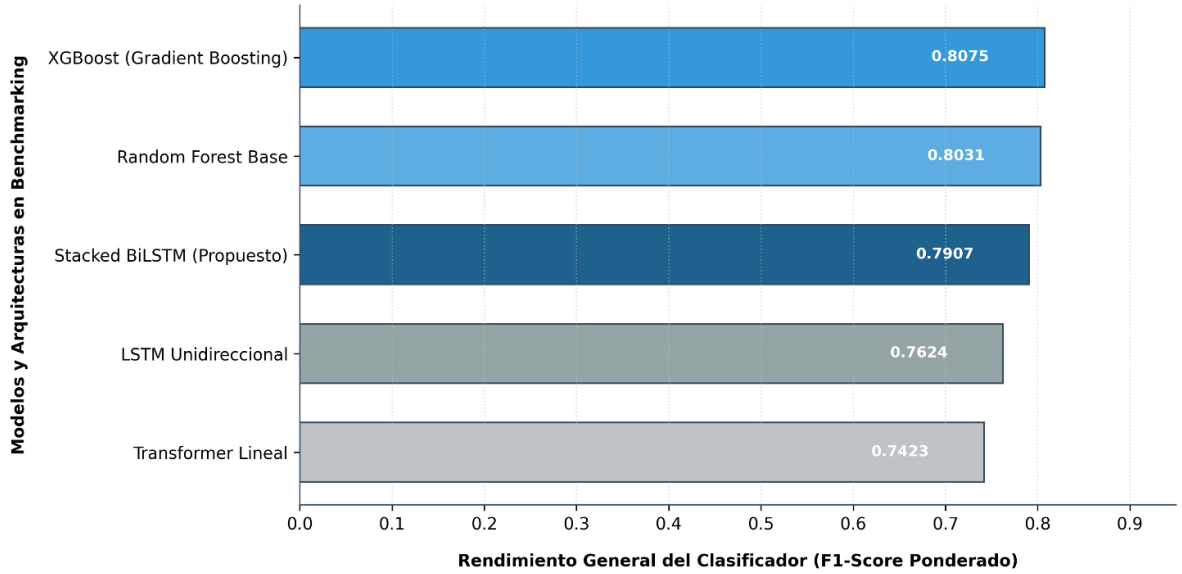
Nota. Elaboración propia, métricas macro extraídas de la matriz comparativa generada en el proyecto.

Los resultados revelan una jerarquía clara: las arquitecturas tabulares (XGBoost y *Random Forest*) alcanzan los rendimientos brutos más altos, con F1 de 0,8075 y 0,8031, respectivamente. La ventaja marginal de los modelos ensamblados responde a su explotación de un hiperplano aplanado de 231 descriptores simultáneos. Concentran una densidad informativa por registro inalcanzable para los enfoques recurrentes. En este escenario, la arquitectura Transformer arrojó la métrica más deprimida (F1 = 0.7423). Este colapso confirma que los mecanismos de atención global carentes de una indexación secuencial estricta amplifican el ruido transaccional del EVA. Por su parte, la LSTM unidireccional computó un F1 = 0.7624, quedando rezagada frente a la robustez de nuestra propuesta *Stacked BiLSTM* (F1 = 0.7907). Validamos la hipótesis de diseño. El escrutinio sinérgico de flujos cronológicos directos e inversos, acoplado al apilamiento jerárquico de las capas ocultas, maximiza la extracción de dependencias dinámicas abstractas. Si bien XGBoost superó numéricamente al clasificador B iLSTM por un diferencial de 0.0168 puntos en el F1-Score, dicha superioridad se disipa al evaluar el área bajo la curva (AUC - ROC); allí, la red propuesta (0.9105) igualó estadísticamente a XGBoost (0.9149) y aventajó al Random Forest (0.9103). La capacidad de segregación de perfiles estudiantiles es equivalente. Logramos este equilibrio operando con

apenas 7 características dinámicas por paso temporal, frente al gigantismo tabular de los árboles. La Figura 7 cartografía este contraste macro.

Figura 7

Análisis comparativo de rendimiento global basado en el F1-Score ponderado para todos los modelos en benchmarking



Nota. Elaboración propia a partir de las matrices de confusión y evaluación consolidadas en la fase de modelado.

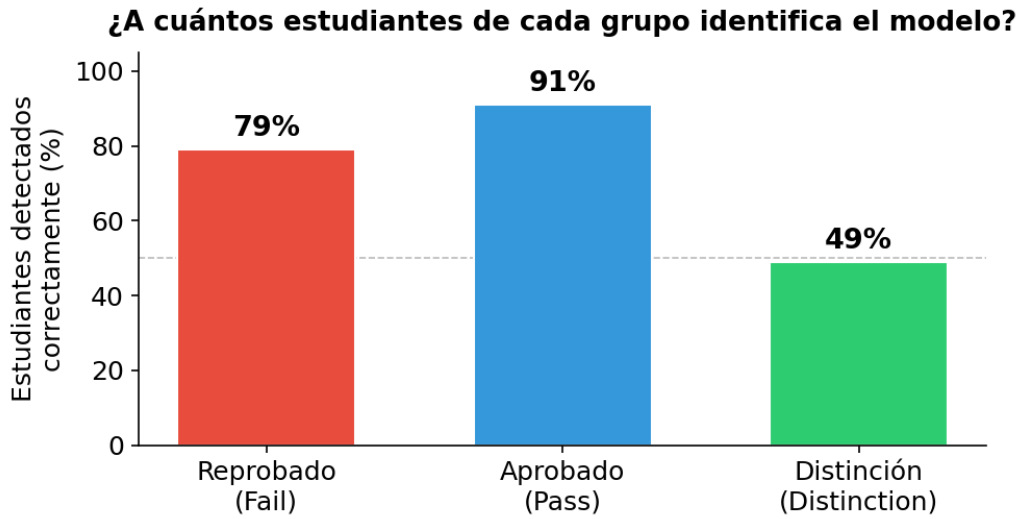
4.7 Análisis por clase y por ventana temporal

El análisis de las matrices de confusión reveló que el error más frecuente en todos los modelos es la confusión entre las clases adyacentes *Pass* y *Distinction*. El *Stacked* BiLSTM detecta correctamente el 79 % de los casos *Fail*, el 91 % de los *Pass* y el 49 % de los *Distinction*. Desde el punto de vista institucional, resulta importante que ningún modelo confunda directamente *Fail* con *Distinction*, un estudiante en riesgo real nunca es clasificado como excelente, los errores se concentran en el frente de *Pass*/*Distinction*, que corresponde a decisiones menos urgentes para la intervención temprana. Para una alertar temprana, la métrica más relevante es la exhaustividad sobre la clase *Fail*, el modelo identifica a 4 de cada 5 estudiantes en riesgo de reprobar la materia, lo que resulta valioso. La clase *Pass* fue la más

difícil de identificar (AUC ≈ 0,86–0,89) por sus datos intermedios, mientras que *Distinction* (AUC ≈ 0,95–0,96) y *Fail* (AUC ≈ 0,91–0,92) resultaron más fáciles de diferenciar. La Figura 8 muestra el índice de detección por tipo de resultado.

Figura 8

Tasa de detección por tipo de resultado del modelo Stacked BiLSTM



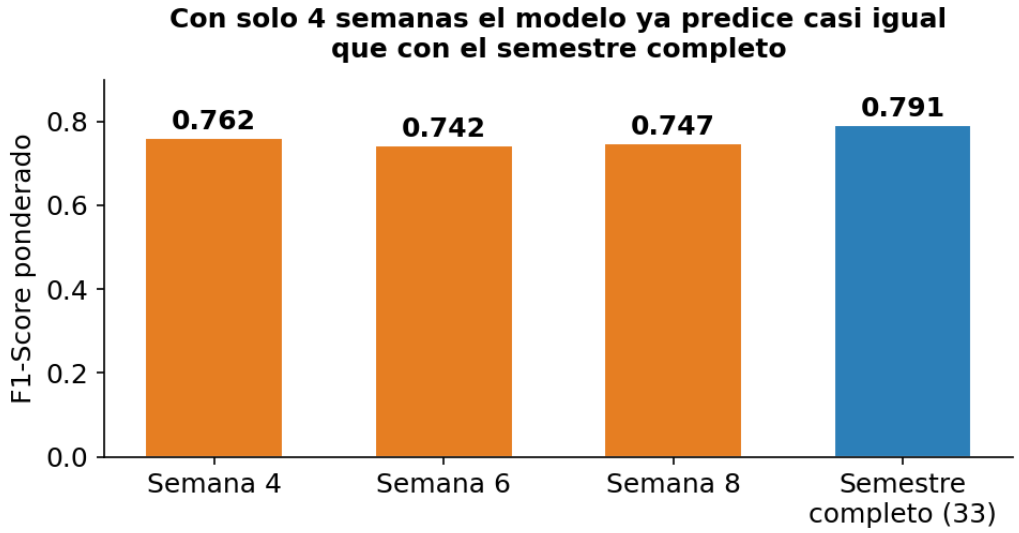
Nota. Elaboración propia a partir de los valores de exhaustividad (*recall*) por clase obtenidos en la evaluación del modelo.

En paralelo, el recuento incremental de las ventanas temporales convalida la viabilidad analítica del diagnóstico prematuro. La Figura 9 mapea este comportamiento, con un registro transaccional restringido a los primeros 28 días (semana 4), la arquitectura profunda consolida un F1-Score ponderado de 0.7619, aproximándose de forma estricta al 0.7907 arrojado por el ciclo semestral completo. La brecha es mínima. Semejante nitidez predictiva dota a la gestión institucional de un horizonte de intervención quirúrgica sumamente holgado. No obstante, detectamos una inflexión negativa transitoria en la semana 6, donde la métrica decae su valor hasta un suelo de 0.7421, este valle adaptativo rehúsa indexar fallos en el aprendizaje de las capas ocultas; responde directamente a una desaceleración estocástica estacional en los logs de interactividad digital, cuya apatía inyectó ruido de baja frecuencia en el flujo del tensor. El

pipeline demostró resiliencia al purgar esta distorsión y recuperar su tracción diagnóstica hacia la semana 8.

Figura 9

Rendimiento del modelo según el número de semanas observadas



Nota. Elaboración propia a partir del experimento de predicción por ventanas temporales.

4.8 Discusión de resultados

Los resultados obtenidos se ajustan con la evidencia previa sobre el potencial de los datos de interacción para anticipar el rendimiento académico. La convergencia empírica lograda por la arquitectura Stacked BiLSTM (F1 = 0,7907) resulta consistente con los hallazgos de Waheed et al. (2020), autores que refrendaron la superioridad de los enfoques profundos sobre los clasificadores tradicionales al capitalizar la densidad masiva de registros en el EVA. Validamos esta ventaja operativa, mediante este comportamiento que robustece la viabilidad de la diagnosis prematura postulada originalmente por Hlosta et al. (2017). La anticipación cronológica es el núcleo. Predecir el destino académico observando apenas cuatro semanas de trazas digitales constituye el aporte con mayor retorno práctico de esta investigación. A diferencia de las aproximaciones pioneras, constreñidas a variables estáticas disponibles solo tras la clausura del ciclo lectivo (Kotsiantis et al., 2003), el modelado secuencial explota la

impronta transaccional del compromiso en su fase germinal. Esta asimetría temporal expande de forma disruptiva la ventana de intervención institucional (Siemens, 2013).

Los modelos tabulares que hayan alcanzado un F1 ligeramente superior no contradice la elección metodológica, la métrica agregada no captura propiedades operativas como la predicción anticipada o la interpretabilidad temporal (Romero & Ventura, 2020); además, dichos modelos requieren el historial completo del semestre, lo que limita su utilidad para la alerta temprana (Breiman, 2001; Chen & Guestrin, 2016).

Para situar la contribución de este trabajo dentro de la trayectoria de este campo, la Tabla 11 lo contrasta con estudios de referencia, atendiendo al enfoque el conjunto de datos y al aporte principal de cada uno.

Tabla 11

Contraste del presente trabajo con estudios de referencia

Estudio	**Metodología**	**Set de datos**	**Aporte principal**
Kotsiantis et al. (2003)	Aprendizaje automático clásico con variables estáticas	Educación a distancia	Pionero en la predicción del abandono.
Hlosta et al. (2017)	Identificación temprana sin modelos previos (*Ouroboros*)	OULAD	Prevención y detección temprana con datos limitados.
Waheed et al. (2020)	Aprendizaje profundo sobre grandes volúmenes del VLE	OULAD	Superioridad y complejidad de los modelos profundos.
Kalita et al. (2025)	Bi-LSTM con interpretabilidad SHAP	Datos educativos	Agrupación de memoria bidireccional e interpretabilidad.
Presente trabajo (2026)	*Stacked* BiLSTM con XAI triangulada	OULAD	Prevención (semana 4), estabilidad e interpretación en un esquema multiclase.

Nota. Elaboración propia, el informe evidencia que el aporte de este trabajo no reside en un incremento de la exactitud bruta, sino más bien en la integración simultánea de prevención, estabilidad e interpretación sobre una tarea de clasificación multiclase.

La invulnerabilidad estructural del modelo, corroborada por la mínima dispersión métrica en el protocolo de validación cruzada, sofoca cualquier sospecha de ajuste fortuito o artefacto estadístico. Mitigamos el sesgo de varianza, dicho comportamiento cobra un relieve

crítico en escenarios condicionados por un severo desbalance de clases, entornos donde una métrica de evaluación ingenua tiende a enmascarar deficiencias operativas y a sobrestimar el desempeño real del clasificador (He & Garcia, 2009; Sokolova & Lapalme, 2009).

El vector de interpretabilidad inyecta un valor estratégico disruptivo frente a la literatura hegemónica, tradicionalmente obsesionada con la optimización forense de la exactitud numérica. La convergencia matemática exacta entre los gradientes de sensibilidad de la red BiLSTM y las atribuciones localizadas de los valores SHAP al consagrar las ventanas cronológicas terminales como las fuerzas determinantes del ciclo proporciona una narrativa accionable, reproducible y auditable (Lundberg & Lee, 2017). Sintonizamos con las demandas de la Inteligencia Artificial Explicable (XAI) (Barredo Arrieta et al., 2020; Molnar, 2022). Paralelamente, el escrutinio estratificado de equidad desnudó una restricción ineludible: el rendimiento predictivo rehúsa comportarse de forma homogénea entre los distintos subgrupos. Su fiscalización continua deviene en un imperativo deontológico para legitimar el despliegue responsable de la herramienta.

Desde el punto de vista de la gobernanza de datos educativa, estos hallazgos nos demuestran que las universidades equipadas con plataformas virtuales ya resguardan el activo más valioso, los registros (logs), capaces de anticipar el abandono o deserción sin recurrir a instrumentaciones intrusivas adicionales, el núcleo del problema muta. El desafío se mueve drásticamente del ecosistema de desarrollo técnico hacia el plano de la reingeniería organizativa; es decir, integrar la predicción en los flujos de tutorías activas y dotar al cuerpo de docentes de tableros de control oportunos. Pese a la solidez métrica, resulta imperativo definir los alcances del estudio. Al estar constreñido a un repositorio singular correspondiente a una institución específica y a cohortes acotadas (Kuzilek et al., 2017), la extrapolación de los umbrales operativos a ecosistemas externos exige una validación cruzada independiente (Romero & Ventura, 2010).

Al contrastar nuestra aproximación frente al estado del arte, la contribución medular trasciende un incremento marginal o cosmético en la exactitud macro. Radica en la articulación sinérgica de una tríada de propiedades altamente deseadas: anticipación temporal, estabilidad matemática e interpretabilidad algorítmica. Este triunvirato rara vez coexiste en la literatura indexada. Estas restricciones la dependencia ciega de un único corpus de datos, la naturaleza de proxy conductual de las pulsaciones web, la latencia de correlaciones espurias no deseadas a pesar de omitir variables demográficas protegidas y el carácter estrictamente predictivo, no causal, del pipeline se diseccionan minuciosamente para convertirse en líneas de acción prioritarias dentro del Capítulo 5.

CAPÍTULO 5. CONCLUSIONES, RECOMENDACIONES Y APLICACIONES

5.1 Conclusiones generales

El estudio académico en el entorno virtual se construye semana a semana con un dato conocido como curso. Este modelo mediante la arquitectura utilizada es capaz de captar el ritmo, la constancia, las caídas, no sólo resultó viable, sino una forma de aprendizaje mucho más desarrollada. Recupera entre el tiempo y la evolución se destaca el enfoque utilizado en el semestre con un vector de promedios. Ayudando a ser más eficiente a la hora de elegir el modelo.

Cabe aclarar que la naturaleza más operativa en la estadística aplicada en el Stacked BiLSTM donde se evidencia el comportamiento en las primeras semanas. La cual contiene una señal de desenlace y anticipa a un no incremento en la exactitud. Modificando el valor del sistema: transformando la analítica, descubriendo intervenir con un margen de acción. La gestión académica renuncia a una fracción de exactitud mínima. Frente a la tabulación del cambio del semestre en el tiempo de reacción se obtiene un intercambio bastante favorable.

CRISP-DM primero genero una documentación con variables bastante delicadas y parcialmente discriminatoria, obteniendo un tratamiento muy desbalanceado y reproducible con esto se debería de tener cuidado el lugar donde se debe implementarlo.

Es importante mencionar los detalles de la sección 5. 3.1, donde este modelo alcanzó en términos brutos la mayor exactitud, superando el estrecho margen que habíamos referenciado. Lo que permite la ventaja secuencial de la interpretabilidad y anticiparse a las métricas agregadas. La fiabilidad tampoco es uniforme: el desempeño es menor para los estudiantes de tiempo parcial y para los módulos de mayor dificultad, una brecha de equidad que exige cautela antes de automatizar decisiones que afecten a personas.

Este sistema no solo predice la deserción, sino también el rendimiento. dónde su umbral se encuentra de manera deliberada constituyendo un hecho comprobable. al momento de

identificar las correlaciones y no relaciones: el acompañamiento no siempre demuestra dichos resultados más bien limita a tener otros alcances con una precisa línea que debe continuar con el trabajo.

Este valor de la analítica no reside únicamente del algoritmo, sino en la capacidad técnica y la necesidad institucional que es necesaria para que el modelo emita las alertas con antelación, en las diferentes semanas. Independientemente. que permita de manera fundamental reunir las condiciones efectivas, integrando protocolos de intervención real, además de salvaguardar éticamente que el estudio lo demuestre. La arquitectura o la tabla de resultados, hace que exista una transparencia y ponga los resultados en el entorno virtual para la culminación de estudios y los servicios de permanencia.

5.2 Conclusiones específicas (aportes)

5.2.1 Aportación de los objetivos de la investigación

La eliminación de OULAD no solo constituyó un criterio que ayudó a escoger las variables como género y región ya que la predicción condicionó las decisiones a posterior. Esta construcción del tensor aportó la representación de conservación cronológica precisamente anticipando el entrenamiento del Stacked BiLSTM junto con los cuatro clasificadores de control, permitiendo una comparación sistemática y ayudando aislar la contribución de la memoria secuencial real en vez de superponerla. No siendo únicamente estable y funcional, sino auditable y reproducible para su captación.

5.2.2 Aportación en la gestión institucional

En la gestión académica es muy favorable el sistema al tener las alertas de los datos. en tiempos o rangos, un factor determinante a la hora de tomar decisiones, sobre diferentes cifras de desempeño que se lo identifica desde muy temprano. Identificando los riesgos que puede ser al reprobar y clasificar como una parte o rango en el semestre, para ello es importante el

esfuerzo de la no discriminación y disponer de datos para que se pueda actuar con mayor veracidad y resiliencia.

5.2.3 Aportación a nivel académico

En la gestión académica no debe depender únicamente de un solo método, ya que estratégicamente se triangula en tres con los diferentes gradientes que nos proporciona en el modelo Stacked BiLSTM. Los valores de Random Forest (SHAP) y la atención de transformer. Una determinante muy importante en cuanto a los modelos para determinar cómo atribuir e identificar las semanas de evaluación a la hora de la interpretación con transformer. Es una réplica auditable la transparencia de estos modelos con un caso concreto para el análisis del aprendizaje, el mismo que se adaptable para todas las instituciones educativas a la hora de implementarlo con sus propios datos, para que puedan tomar mejores decisiones.

5.2.4 Aportaciones a nivel personal

A nivel personal, es importante reconocer que los autores implementaron sus diferentes habilidades en cuanto a actitudes y aptitudes obtenidas en clases. Desde la limpieza, construcción, diseño, pruebas y evaluación de las redes implementadas, ayudando a la mejor toma de decisiones, lo que conlleva a elegir qué modelo se pueda implementar en diferente giro de negocio. Adicional se tiene o se prevé la mejor forma de tratar con datos de manera ética y a contribuir con diferentes modelos a la hora de elegir tomando en cuenta las diferentes necesidades para que sea eficiente al momento de utilizar los sistemas.

5.3 Limitaciones y Recomendaciones

5.3.1 Limitaciones

El desempeño del modelo no es típico entre grupos: el F1 varió entre 0,725 y 0,875 según el módulo y los estudiantes a tiempo completo se predijeron mejor (0,797) que los de tiempo parcial (0,752) a lo largo de lo recorrido. Aquellos pronósticos en tiempo parcial se dividen en diferentes grupos para los módulos con menor cobertura, tomando en cuenta que al

usar es necesario ser cauteloso en cada una de las intervenciones y validaciones. Más allá de lo establecido, este ayuda a que los umbrales en el contexto automático no se trasladen y el modelo del rendimiento que predice no sea por la deserción la cual es una clara limitante. XGBoost y Random Forest logran un alto superior en F1 tomando en cuenta el rango del semestre completo, careciendo de interpretabilidad por un periodo con el modelo BiLSTM.

5.3.2 Recomendaciones

Es importante tener en cuenta que se puede implementar dos umbrales con un entorno virtual académico. Una alerta en la semana 4 para iniciar la intervención y seguimiento, otra alerta en la semana 8 para tener la referencia establecida.

Adicionalmente este monitoreo debe enfocarse en la fase inicial, en la fase media y en la fase de cierre definitivo. Para que los datos sean más eficientes y oportunos a la hora de tomar decisiones.

Van a existir muchos falsos negativos, lo que es muy importante priorizar de qué manera integral debemos proteger las variables ante cualquier futura extensión que podría auditarse, una pérdida periódica de sesgos, Los cuales se puede reducir mediante una buena práctica de limpieza de datos, segmentación, validación antes de comenzar a realizar cualquier proceso o evaluación de los datos.

En la implementación es necesario seguir las fases una por una; Siendo la primera fase de la retrospectiva. La segunda fase del número reducido de asignaturas, Acompañada con capacitadores o tutores lo que serviría tener una mejor aceptación por parte de la gestión estratégica educativa.

Finalmente, se debería realizar los protocolos y planes de acción para que se pueda tener un sistema bastante robusto que impacte, reduzca errores y favorezca la construcción de alertas en tiempo real, teniendo confianza educativa a nivel nacional y como no, internacional.

5.4 El modelo y sus aplicaciones

Es aplicable en todos los modelos de gestión educativa, tanto a nivel Nacional e Internacional, presencial o virtual, las cuales tienen diferentes modalidades para detectar las anomalías críticas por las cuales pasan los estudiantes universitarios pasan a la hora de la deserción, con un claro impacto en el rendimiento académico.

5.4.1 Entorno virtual - Integraciones

Los modelos de predicción se han implementado con la convicción de menorar el impacto que se genera en la deserción académica por las cuales hoy en día estudiantes salen de las universidades. Evidenciando aquella práctica que se maneja en el entorno virtual, lo que facilita transformar el impacto que este genera y tomar acciones de mejora en cuanto a los riesgos que carecen semana por semana, con diferentes actitudes tomadas entre los responsables de los indicadores de gestión.

5.4.2 Alerta temprana - Tablero

El modelo tiene un seguimiento concreto en cuanto al rendimiento académico y la probabilidad de riesgo a lo largo del semestre, dado que el desempeño es muy útil en el aspecto de gestión educativa, el cual se lo revisa desde la cuarta semana, cuando los estudiantes están más establecidos y con mejor atención, prestos a la adaptación correspondiente. También pueden presentarse la desconexión desde señales tempranas.

Este tablero se organiza en 3 vistas que son muy explicativas. La primera vista ofrece al coordinador la distribución a nivel de riesgos y la evolución de cada semana. También ofrece una vista individual por estudiante para que en su clase pueda ver el rendimiento académico, la trayectoria, las evidencias y el compromiso del estudiante. Adicional, podemos tener acciones emprendidas tutoriales y con un buen resultado. Desde la predicción de respuestas incorporadas. Se filtra por modalidad, módulo o riesgo. La posibilidad de obtener informes. Para que distintos autores, tutores, coordinadores, puedan verificar y tener en cuenta los

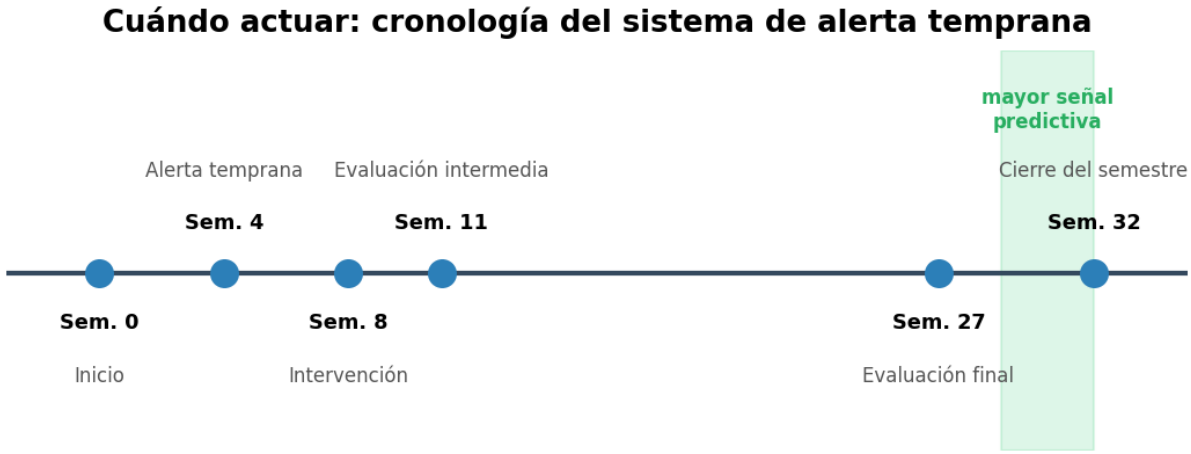
responsables de bienestar estudiantil y consulten la información pertinente a su rol sin sobrecargarse de datos irrelevantes.

5.4.3 Intervención académica – Protocolo

Se establece un protocolo escalonado, con el objetivo de que la predicción se traduzca en acción: en la semana 4, un aviso preventivo, en la semana 8, una intervención dirigida y en las semanas de evaluación (11, 27 y el tramo final 29 a 32) un seguimiento reforzado, definidos como las más determinantes por el análisis de interpretabilidad. Cada nivel establece una respuesta clara. Desde el contacto inicial entre el tutor y el alumno, para tener en cuenta que tiene el servicio de apoyo, información relevante y el reporte final. El cual se resume en la Figura 5.1 la intervención cronológica.

Figura 10

Cronología de sistema de alerta temprana



Nota. Elaboración propia identificados en el estudio a través de los puntos de inflexión predictiva.

5.4.4 Gobernanza de datos y ética

Es importante conocer que este modelo es de carácter sensible. Ya que acompaña continuamente monitoreando el grupo o grupos, en particular entre los estudiantes de tiempo parcial y tiempo completo, el cual se esquematiza en conocer y ver cómo la predicción se orienta, pero las decisiones a tomar siempre lo determinan los docentes.

5.4.5 Escalabilidad y mantenimiento

La aplicación sostenida en el tiempo requiere un ciclo de mantenimiento del modelo: reentrenamiento periódico con las cohortes propias de cada institución, vigilancia de la deriva de los datos cuando cambian las plataformas o los hábitos de estudio, y versionado de los modelos en producción. Este enfoque de operación continua asegura que el sistema conserve su validez a medida que evoluciona el entorno virtual de aprendizaje.

5.4.6 Aplicaciones a futuro

Módulos para automatización del uso de los recursos en el ambiente educativo.

Módulos para proteger la información y centralizarla de manera automatizada.

Módulo para incorporar señales neuronales y analizar la información de los estudiantes.

Módulo para el procesamiento del lenguaje natural, integración de entregas y evaluaciones puntuales.

Módulo que conecte las fuentes temporales con el análisis del estudiante obteniendo un compromiso y responsabilidad.

Módulo que conecte con un agente virtual y pueda comunicarse con los estudiantes para verificar todos los miedos o problemas.

Módulo de predicción para la intervención oportuna en cualquier momento del semestre.

5.5 Análisis de impacto y viabilidad del despliegue

Tomar en cuenta que sólo se adquiere un valor práctico y científico con los modelos que son sólidos mediante métricas, análisis rigurosos y potenciales. Creando medidas e impacto que generen con proyectos en tiempo real.

También se debe tener en cuenta la supervisión explícita, efectiva, cuantitativa y cualitativa. Al ser desplegados en trabajos futuros que generen impacto de valor.

Trasciende equitativamente en el ejercicio computacional para tener una mejor intervención con el propósito de ser medibles los valores y datos analizados.

5.5.1 Impacto potencial estimado

En tres planos se mide el impacto del sistema articulado. El modelo identifica tempranamente al 79 % de los estudiantes que finalmente reprueban en el plano académico, analizada en la que la reprobación afecta a cerca del 21,6 % de las matrículas, proyectado sobre una población equivalente, el sistema señalaría de forma anticipada a unos 790 antes de que concluya el primer mes de clases, por cada mil estudiantes en riesgo de reprobar. Esta cobertura, ningún estudiante en riesgo es clasificado como de excelencia, combinada con la ausencia de confusiones críticas, habilita una focalización precisa del acompañamiento.

La anticipación de aproximadamente 29 semanas transforma la naturaleza de la respuesta, en el plano institucional: una gestión reactiva, actúa cuando la reprobación ya es probable, que dispone de tiempo suficiente para reconducir la trayectoria del estudiante, a una preventiva, asumiendo de forma conservadora que las intervenciones lograran reconducir únicamente a una fracción de los estudiantes correctamente identificados la literatura documenta reducciones significativas de la deserción mediante intervenciones basadas en el establecimiento de metas, el efecto agregado a escala de una universidad con miles de matrículas por cohorte resultaría considerable en términos de retención y de indicadores de calidad.

El impacto socioeconómico, tanto para el individuo como para la institución y el sistema educativo en su conjunto, el estudiante retenido representa la preservación de un proyecto formativo personal y la evitación de los costos asociados al abandono. Debe anotarse, su confirmación corresponde a la fase de validación causal descrita más adelante con honestidad metodológica, que estas cifras constituyen escenarios ilustrativos fundamentados en el desempeño del modelo sobre datos históricos, y no resultados medidos.

5.5.2 Análisis costo-beneficio

En la parte socio económica, de cada estudiante, el sistema es favorable asimétricamente, ya que se prevé el costo y beneficio con la implementación. Esta solución registra interacción para entornos académicos virtuales, ya que se genera automáticamente diferentes puntos de iteración e instrumentos necesarios para que se entrenen los modelos y permitan una práctica de conexión periódica en el tiempo. Por otro lado, los costos, son relativamente favorables, ya que el uso tecnológico hoy en día es muy importante, focalizado en los beneficios que trae consigo y con lo que se genera la deserción, para que a futuro los estudiantes analicen porque del bajo rendimiento académico y la desvinculación e incrementar el despilfarro de dinero que invierte el estudiante. En síntesis, el sistema desplaza los recursos de acompañamiento desde una lógica de cobertura indiscriminada hacia una de precisión, mejorando la eficiencia sin incrementar el gasto de los recursos.

5.5.3 Indicadores clave de desempeño y protocolo de medición del impacto real

Para transformar el impacto proyectado en impacto verificado, se propone un cuadro de indicadores clave de desempeño (KPI) acompañado de un protocolo de medición. La Tabla 12 sintetiza los indicadores, que combinan métricas del modelo, métricas de proceso y métricas de resultado institucional.

Tabla 12

Indicadores clave de desempeño (KPI) para la medición del impacto real del sistema

Categoría	Indicador (KPI)	Definición y meta orientativa
Modelo	Cobertura de riesgo (*recall* de *Fail*)	Porcentaje de estudiantes que reprueban detectados a tiempo (meta ≥ 79 %).
Modelo	Precisión de la alerta	Porcentaje de alertas que corresponden a un riesgo real (minimizar las falsas alarmas).
Proceso	Tiempo medio hasta la intervención	Semanas transcurridas entre la alerta y el contacto tutorial (meta ≤ 1 semana).
Proceso	Tasa de contacto efectivo	Porcentaje de estudiantes alertados que reciben efectivamente la intervención.
Resultado	Reducción de la tasa de reprobación	Variación de la tasa de *Fail* frente a la línea base histórica.
Resultado	Mejora de la tasa de retención	Variación de la tasa de permanencia frente a la línea base histórica.
Resultado	Tasa de conversión del riesgo	Porcentaje de estudiantes alertados que finalmente aprueban.
Eficiencia	Costo por estudiante retenido	Recursos de tutoría invertidos por cada estudiante retenido.
Equidad	Brecha de desempeño entre subgrupos	Diferencia de retención entre estudiantes de tiempo completo y parcial, y entre módulos.

Nota. Elaboración propia. Los indicadores deben monitorearse de forma continua tras el despliegue, articulando las métricas técnicas del modelo con los resultados institucionales que constituyen el fin último del sistema.

Es importante constatar un cohorte relacionado al impacto en una medida real de un diseño experimental para las intervenciones del modelo activo. Con una base histórica se puede verificar y evaluar las diferentes aprobaciones y retenciones de las pruebas estadísticas significativas. Incorporando salvaguardas éticas en el caso del apoyo disponible a un grupo de control, por lo que la comparación se apoyará preferentemente en cohortes históricas previas a la implantación. Se estructura en 3 etapas el Protocolo: La validación de calibrar los umbrales, el piloto controlado y acotado de asignaturas y finalmente, el progresivo escalado condicionado, evidenciando el impacto obtenido.

5.5.4 Viabilidad organizativa y técnica

La viabilidad organizativa y técnica desde el punto de vista arquitectónico se despliega en 5 capas por las cuales existe un entrenamiento periódico y un monitoreo con el cual se deriva diferentes datos.

El pipeline siendo una semilla global aleatoria, el entrenamiento para la conservación que permita auditar y replicar un sistema confiable, el factor crítico tecnológico de éxito para medir la capacidad institucional en acciones concretas, la disponibilidad e integración de flujos de protocolos de formación e interpretación de la información para recorrer la barrera y el impacto organizativo, no algorítmico, el cual constituye un valor de gestión orientado a la precisión, capacidad predictiva, esfuerzo de implementación, mejora efectiva que se convierta en una permanencia estudiantil institucional.

5.6 Trabajo a futuro

Es importante mencionar una buena validación a futuro que prevé los modelos sobre la deserción y aprobación de la tasa de los estudiantes mediante los diferentes modelos, diseños y otras fuentes o experimentos, las cuales tengan la resistencia e información necesaria para la asistencia a las aulas de clase, en diferentes ambientes híbridos para qué combinen recurrencia y atención (Vaswani et al., 2017) y el reentrenamiento del modelo con datos propios de cada institución para cerrar las brechas de equidad detectadas. Convirtiendo en esta plataforma un análisis de aprendizaje para predicción a futuro.

Perspectiva por la cual debe tener una consolidación mayor, alcance, datos, gobernanza, resiliencia. Como un complemento de la analítica predictiva con la capacidad institucional para atender a tiempo las necesidades del estudiante. Su integración asegura la calidad con un sistema de políticas bien claras y establecidas, una formación docente con datos configurados, tecnología, pedagogía, ética, coordinación los cuales aspiran a ser un horizonte completo para avanzar, aportando nuevas cosas y no sea empíricamente. Identificar los riesgos académicos de

manera temprana, estable, respetuoso, interpretable con los derechos y deberes de los estudiantes, tomando en cuenta los diferentes factores psicosociales que puedan atravesar, disminuyendo la trazabilidad de poder actual a tiempo y seguir avanzando paulatinamente.

.

REFERENCIAS BIBLIOGRÁFICAS

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. En *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)* (pp. 265–283). USENIX Association.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. En J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Bañeres, D., Rodríguez-González, M. E., Guerrero-Roldán, A. E., & Cortadas-Guasch, P. (2023). An early warning system to identify and intervene in online dropout learners. *International Journal of Educational Technology in Higher Education, 20*, Artículo 3. https://doi.org/10.1186/s41239-022-00371-5

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. En *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. En *Proceedings of EMNLP 2014* (pp. 1724–1734). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1179

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks, 18*(5–6), 602–610. https://doi.org/10.1016/j.neunet.2005.06.042

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems, 28*(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. En *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)* (pp. 6–15). ACM. https://doi.org/10.1145/3027385.3027449

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational Intelligence and Neuroscience, 2018*, Artículo 6347186. https://doi.org/10.1155/2018/6347186

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. En *Proceedings of the 32nd International Conference on Machine Learning* (Vol. 37, pp. 448–456). PMLR.

Ismanto, E., Ab Ghani, H., Abdul Aziz, N. H. B., Md Saleh, N. I., & Effendy, N. (2024). Enhancing student performance prediction through LSTM-based deep learning models with unbalanced data handling using oversampling approach. En *Proceedings of the 4th International Conference on Communication, Language, Education and Social Sciences (CLESS 2023)* (pp. 192–202). Atlantis Press.

Kalita, E., Alfarwan, A. M., El Aouifi, H., Kukkar, A., Hussain, S., Ali, T., & Gaftandzhieva, S. (2025). Predicting student academic performance using Bi-LSTM: A deep learning framework with SHAP-based interpretability and statistical validation. *Frontiers in Education, 10*, Artículo 1581247. https://doi.org/10.3389/feduc.2025.1581247

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. En *3rd International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.1412.6980

Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003). Preventing student dropout in distance learning using machine learning techniques. En *Knowledge-Based Intelligent Information and Engineering Systems (KES 2003)* (LNCS 2774, pp. 267–274). Springer. https://doi.org/10.1007/978-3-540-45226-3_37

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics dataset. *Scientific Data, 4*, Artículo 170171. https://doi.org/10.1038/sdata.2017.171

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. En *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates.

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2.ª ed.). https://christophm.github.io/interpretable-ml-book/

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.

The Open University. (s. f.). *About the Open University*. Recuperado el 8 de julio de 2026, de https://about.open.ac.uk/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. En *Advances in Neural Information Processing Systems 28* (pp. 505–513). Curran Associates.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C, 40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery, 10*(3), Artículo e1355. https://doi.org/10.1002/widm.1355

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing, 45*(11), 2673–2681. https://doi.org/10.1109/78.650093

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing, 5*(4), 13–22.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*(1), 1929–1958.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. En *Advances in Neural Information Processing Systems 30* (pp. 5998–6008). Curran Associates.

Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior, 104*, Artículo 106189. https://doi.org/10.1016/j.chb.2019.106189

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. En *Proceedings of the 4th International Conference on the Practical Application of Knowledge Discovery and Data Mining* (pp. 29–39).

ANEXOS

Anexo A. Documentación del proyecto: parte técnica

La implementación se desarrolló en un archivo computacional (Jupyter Notebook) que contiene el *pipeline* completo bajo la metodología CRISP-DM, desarrollado en Python con la biblioteca TensorFlow y el Api Keras (Abadi et al., 2016) y scikit-learn (Pedregosa et al., 2011). El archivo abarca la carga y auditoría de datos, la ingeniería del tensor temporal, SMOTE que es un balanceador, el entrenamiento del modelo *Stacked* BiLSTM y de los modelos, validación cruzada, control y la interpretabilidad. Se entrega como archivo adjunto (`UIDE_G4_TESIS_MCDA_B_JUN26_Notebook.ipynb`) y garantiza la reproducibilidad de todos los experimentos mediante la fijación de una semilla aleatoria global. La Tabla A.1 resume la correspondencia entre las fases metodológicas y los componentes implementados.

Tabla A 1

Correspondencia entre las fases CRISP-DM y los componentes implementados

Fase CRISP-DM	Componente implementado
1. Comprensión del negocio	Definición del problema y de la pregunta de investigación.
2. Comprensión de los datos	Carga de 7 archivos CSV, auditoría de calidad, optimización de memoria y EDA.
3. Preparación de los datos	Ingeniería de características, codificación, SMOTE y tensor 3D (33 × 7).
4. Modelado	*Stacked* BiLSTM, LSTM, *Transformer*, *Random Forest*, XGBoost y *Keras Tuner*.
5. Evaluación	Métricas ponderadas, ROC *One-vs-Rest*, matrices de confusión, validación cruzada k = 5 y equidad.
6. Interpretabilidad	Gradientes, valores SHAP, pesos de atención y trayectorias de *engagement*.

Nota. Elaboración propia.

Anexo B. Diccionario de datos del conjunto OULAD

A continuación, se describe la estructura de los siete archivos del repositorio OULAD y el significado de sus campos principales (Tablas B.1 a B.7).

Tabla B 1

courses.csv

Campo	Descripción
`code_module`	Código del módulo (asignatura).
`code_presentation`	Código de la presentación (año y semestre).
`module_presentation_length`	Duración de la presentación en días.

Tabla B 2

assessments.csv

Campo	Descripción
`id_assessment`	Identificador de la evaluación.
`assessment_type`	Tipo: TMA, CMA o Exam.
`date`	Día de entrega previsto.
`weight`	Peso en la nota final (%).

Tabla B 3

vle.csv

Campo	Descripción
`id_site`	Identificador del recurso del VLE.
`activity_type`	Tipo de recurso (`oucontent`, `forumng`, etc.).
`week_from` / `week_to`	Semanas de disponibilidad del recurso.

Tabla B 4

studentInfo.csv

Campo	Descripción
`id_student`	Identificador del estudiante.
`gender` / `region`	Variables excluidas (sesgo ético / alta cardinalidad).
`highest_education`, `imd_band`, `age_band`	Perfil socioeducativo.
`studied_credits`	Créditos matriculados.
`final_result`	Resultado final: *Pass*, *Fail*, *Withdrawn* o *Distinction* (objetivo).

Tabla B 5

studentRegistration.csv

Campo	Descripción
`date_registration`	Día de matrícula.
`date_unregistration`	Día de baja (vacío si no se dio de baja).

Tabla B 6

studentAssessment.csv

Campo	Descripción
`id_assessment` / `id_student`	Claves de evaluación y estudiante.
`date_submitted`	Día de entrega real.
`score`	Puntaje obtenido (0–100).

Tabla B 7

studentVle.csv

Campo	Descripción
`id_student` / `id_site`	Estudiante y recurso.
`date`	Día de la interacción.
`sum_click`	Número de clics ese día (fuente del tensor temporal).

Nota. Elaboración propia a partir de la documentación oficial de OULAD (Kuzilek et al., 2017).

Anexo C. Matrices de confusión y reportes de clasificación

Se presentan las métricas comparativas de los cinco modelos evaluados (Tabla C.1), sus matrices de confusión (Figura C.1) y las curvas ROC *One-vs-Rest* generadas en el proyecto (Figura C.2).

Tabla C 1

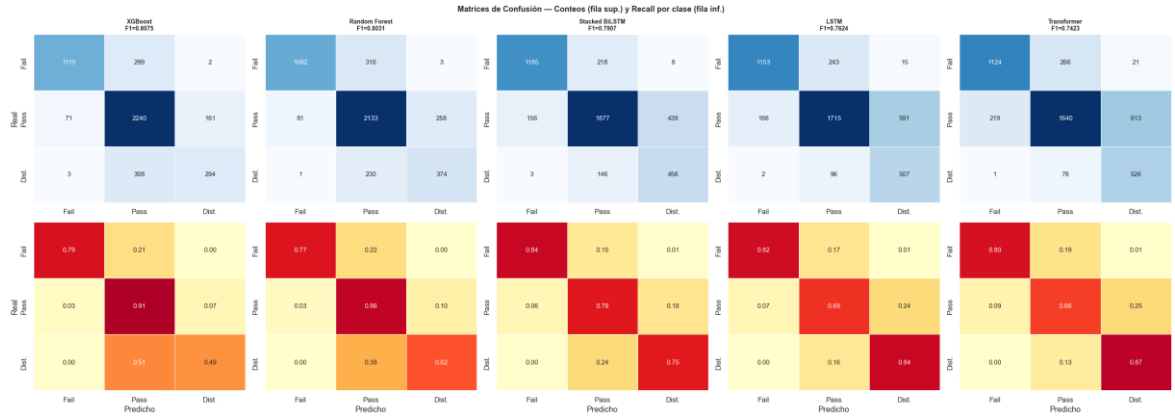
Métricas comparativas de los modelos (conjunto de prueba)

Modelo	Exactitud	Precisión	Sensibilidad	F1-Score
Stacked BiLSTM	0,7839	0,8066	0,7839	0,7907
LSTM	0,7520	0,7957	0,7520	0,7624
Transformer	0,7331	0,7794	0,7331	0,7423
Random Forest	0,8019	0,8104	0,8019	0,8031
XGBoost	0,8119	0,8148	0,8119	0,8075

Nota. Elaboración propia a partir de los resultados consolidados en el cuaderno del proyecto.

Figura C 1

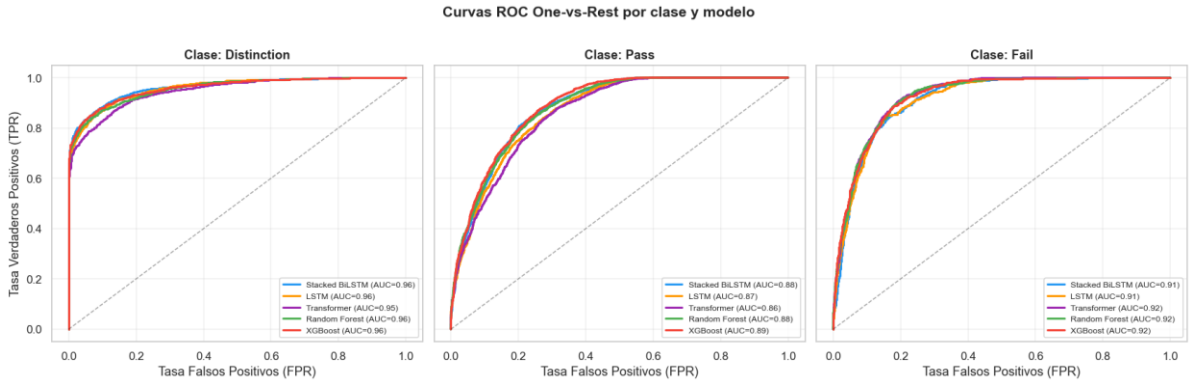
Matrices de confusión de los modelos evaluados



Nota. Elaboración propia. La fila superior muestra los conteos absolutos y la inferior la exhaustividad (*recall*) normalizada por clase.

Figura C 2

Curvas ROC One-vs-Rest por clase y modelo



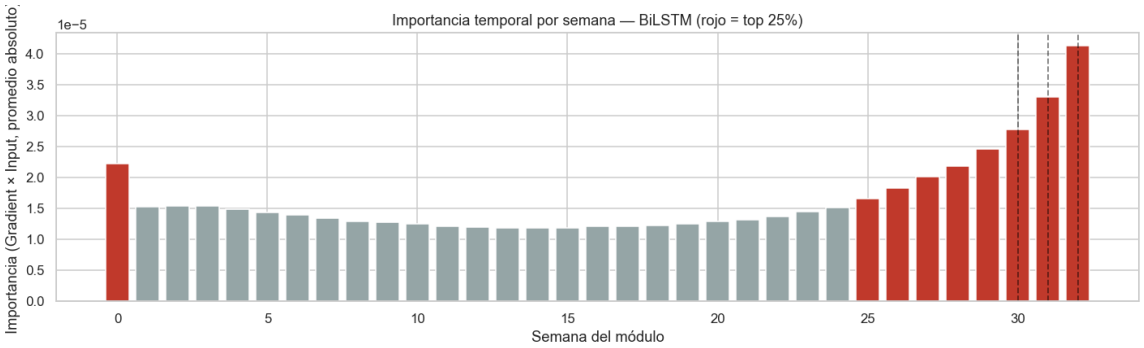
Nota. Elaboración propia a partir de las probabilidades de predicción sobre el conjunto de prueba.

Anexo D. Gráficos complementarios de interpretabilidad

Se reúnen las visualizaciones de interpretabilidad que sustentan la identificación de las ventanas temporales críticas (Figuras D.1 a D.5).

Figura D 1

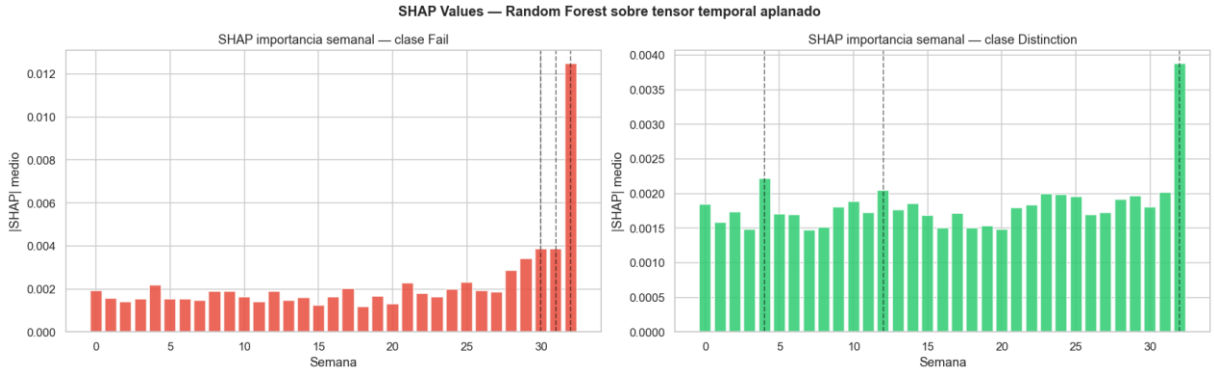
Importancia temporal por gradientes (Gradient × Input) del Stacked BiLSTM



Nota. Elaboración propia. Las barras destacadas corresponden al cuartil superior de importancia.

Figura D 2

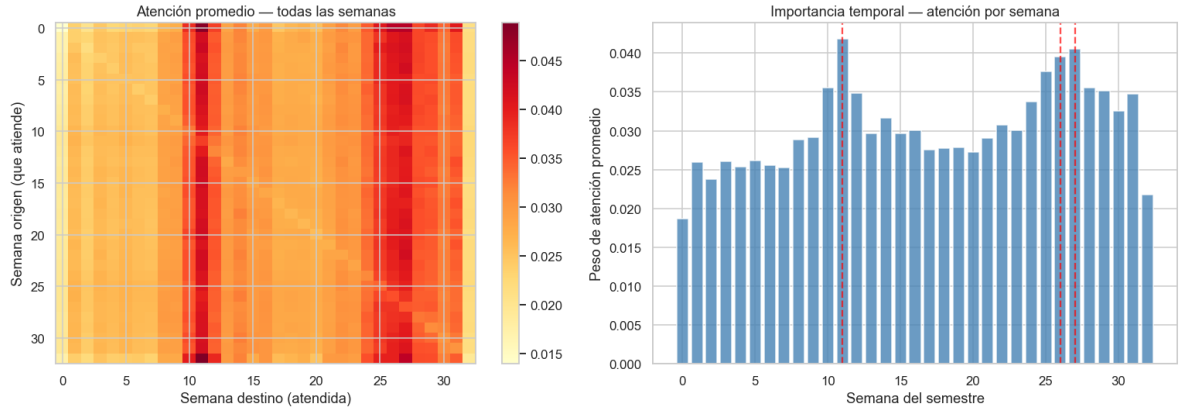
Valores SHAP agregados por semana (Random Forest)



Nota. Elaboración propia a partir del análisis de contribuciones de Shapley por paso temporal.

Figura D 3

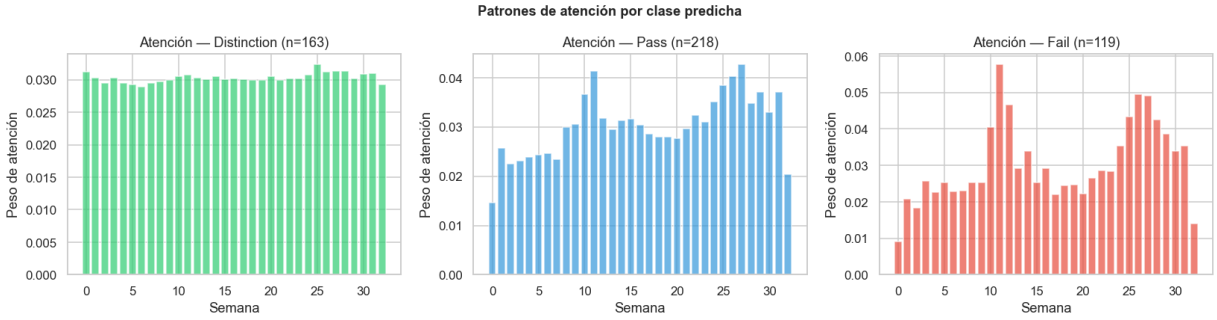
Mapa de calor de los pesos de autoatención del Transformer



Nota. Elaboración propia a partir de la matriz de atención promedio entre semanas.

Figura D 4

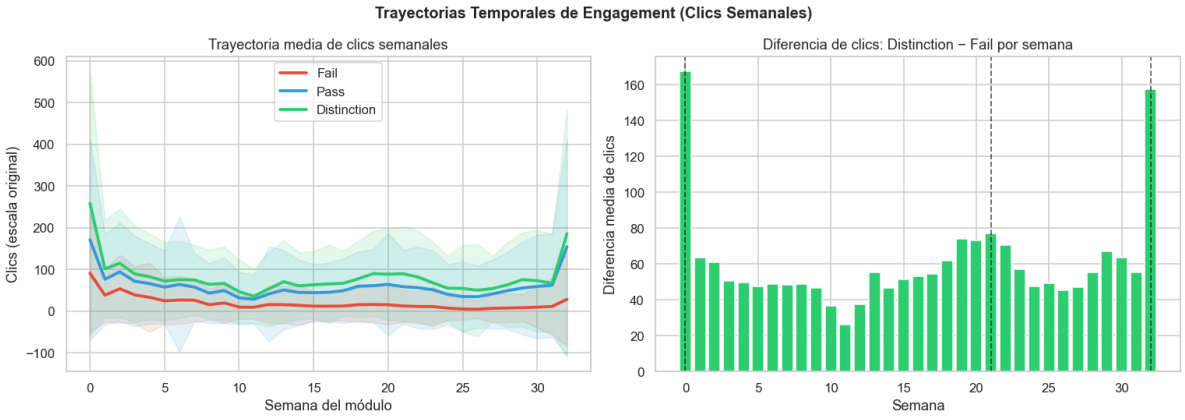
Atención promedio por clase predicha



Nota. Elaboración propia. Cada panel corresponde a una de las clases de rendimiento.

Figura D 5

Trayectorias de actividad semanal por resultado académico



Nota. Elaboración propia por clase sobre el conjunto de prueba a partir de las medias de clics semanales.

La Figura D.5 verifica la evolución media de la actividad semanal por resultado académico: la trayectoria de quienes alcanzan la distinción se mantiene por encima de la de quienes reprueban en el semestre, señal que aprende a explotar el modelo secuencial.

Anexo E. Arquitectura propuesta de despliegue

Arquitectura de referencia cinco capas separadas, cada una evoluciona de forma independiente.

Capa de ingesta de datos: extraer periódicamente con una cadencia semanal los registros de interacción mediante procesos de extracción, transformación y carga (ETL).

Capa de construcción: que replica el *pipeline* validado: agrega las interacciones por matrícula y por semana.

Capa de inferencia: que alberga el modelo *Stacked* BiLSTM entrenado y produce, para cada matrícula activa, la probabilidad de pertenencia y el nivel de riesgo asociado.

Capa de presentación e intervención: tablero de alerta temprana, por prioridad de atención, capa imprescindible el principio de intervención humana, siempre al criterio docente.

Capa de gobernanza y mantenimiento: monitoreo continuo y equidad entre subgrupos, de la detección de la deriva de los datos.

Finalmente se obtiene: escalabilidad, facilidad de auditorías, adaptación a cambios sin necesidad de rediseños.