



Maestría en

INTELIGENCIA ARTIFICIAL APLICADA

Trabajo previo a la obtención de título de Magister en Inteligencia Artificial Aplicada

Autores:

Gabriela Alexandra Changotasig Ayala

Javier Sebastián Horra Ramos

Carlos Enrique Miño Flores

TUTOR/ES:

Karla Estefanía Mora Cajas

Fernanda Paulina Vizcaino Imacaña

Tema:

Modelo predictivo de demanda horaria para optimizar el horneado de pan en puntos de venta de una cadena de panadería en Ecuador

Certificación de autoría

Nosotros, **Gabriela Alexandra Changotasig Ayala, Javier Sebastián Horra Ramos y Carlos Enrique Miño Flores**, 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.

GABRIELA
ALEXANDRA
CHANGOTASIG AYALA

Firmado digitalmente por
GABRIELA ALEXANDRA
CHANGOTASIG AYALA
Fecha: 2026.08.24 17:36:08 -05'00'

Firma

Gabriela Alexandra Changotasig Ayala

Javier Horra R.

Firmado digitalmente por
Javier Horra R.
Fecha: 2026.08.24 12:37:52
-05'00'

Firma

Javier Sebastián Horra Ramos

CARLOS
ENRIQUE
MIÑO FLORES

Firmado digitalmente
por CARLOS ENRIQUE
MIÑO FLORES
Fecha: 2026.08.24
12:51:56 -05'00'

Firma

Carlos Enrique Miño Flores

Autorización de Derechos de Propiedad Intelectual

Nosotros, **Gabriela Alexandra Changotasig Ayala, Javier Sebastián Horra Ramos y Carlos Enrique Miño Flores**, en calidad de autores del trabajo de investigación titulado **Modelo predictivo de demanda horaria para optimizar el horneado de pan en puntos de venta de una cadena de panadería en Ecuador**, 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

GABRIELA
ALEXANDRA
CHANGOTASIG
AYALA

Firmado digitalmente por
GABRIELA ALEXANDRA
CHANGOTASIG AYALA
Fecha: 2026.08.24
17:38:29 -05'00'

Firma

Gabriela Alexandra Changotasig Ayala

Javier Horra R.

Firmado digitalmente por Javier
Horra R.
Fecha: 2026.08.24 12:38:21 -05'00'

Firma

Javier Sebastián Horra Ramos

CARLOS
ENRIQUE
MIÑO FLORES

Firmado digitalmente
por CARLOS ENRIQUE
MIÑO FLORES
Fecha: 2026.08.24
12:52:13 -05'00'

Firma

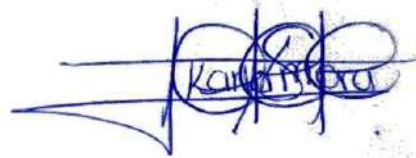
Carlos Enrique Miño Flores

Aprobación de dirección y coordinación del programa

Nosotros, **Alejandro Cortés López** Director EIG y **Karla Estefanía Mora Cajas** Coordinadora UIDE, declaramos que: **Gabriela Alexandra Changotasig Ayala, Javier Sebastián Horra Ramos y Carlos Enrique Miño Flores** 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 Inteligencia Artificial
Aplicada



Karla Estefanía Mora Cajas
Coordinadora de la
Maestría en Inteligencia Artificial Aplicada

DEDICATORIA

Gabriela Alexandra Changotasig Ayala

Dedico este trabajo a Dios, por darme fortaleza, sabiduría y perseverancia durante este proceso. A mi hija, Mía, quien ha sido mi mayor inspiración y la razón que me impulsa a crecer, aprender y esforzarme cada día. A mi esposo Edgar y familia, por su apoyo, comprensión y compañía en cada etapa de este camino.

Javier Sebastián Horra Ramos

Dedico este trabajo a Dios, por darme la fortaleza, la sabiduría y la oportunidad de llegar hasta este momento, permitiéndome dar lo mejor de mí en cada paso de este camino. A mi madre, por creer siempre en mí y enseñarme el valor del esfuerzo y la perseverancia. A mi hijo, por ser mi mayor motivación y la razón que me impulsa a seguir adelante cada día. También a mi novia, por acompañarme durante este proceso, por su paciencia, comprensión y apoyo constante.

Carlos Enrique Miño Flores

Este trabajo lo dedico, en primer lugar, a Dios, por darme la fuerza para dar lo mejor de mí y superarme cada día. A mi esposa, por ser el apoyo en cada paso que doy y por estar siempre conmigo, en las buenas y en las malas, pese a todo. A mis hijos Nicolás y Joaquín, para que sepan que son el motor por el que cada día me levanto, que los amo y que entiendan que, si se esfuerzan en la vida, todo lo bueno llega por añadidura. Esto es por ustedes, los amo con toda mi alma y corazón.

AGRADECIMIENTOS

Gabriela Alexandra Changotasig Ayala

Agradezco a Dios por acompañarme durante esta etapa y permitirme culminar este objetivo académico. A mi esposo e hija, por su amor, paciencia, apoyo. A mi madre, hermana y tía Julieta, por estar siempre pendientes de mí, por su cariño y ánimo en cada etapa. A mis compañeros, Javier y Carlos, por su compromiso, colaboración y trabajo compartido durante el desarrollo de esta investigación. También agradezco a todas las personas que aportaron con su orientación, conocimiento y apoyo para alcanzar este objetivo.

Javier Sebastián Horra Ramos

Agradecimientos

Agradezco a mi madre, a mi hijo y a mi novia por su apoyo constante, paciencia y motivación durante este proceso. A mis compañeros Gabriela Alexandra Changotasig Ayala y Carlos Enrique Miño Flores, por compartir conocimientos, experiencias y aprendizajes que enriquecieron esta etapa académica. Finalmente, agradezco a mis docentes y a todas las personas que contribuyeron con su orientación y apoyo para la culminación de este trabajo.

Carlos Enrique Miño Flores

Quiero agradecer a toda mi familia, mis padres y hermano por estar siempre apoyándome en todo lo que hago. A mis cuñadas por siempre darme ánimos para no decaer.

A mis compañeros Gaby y Javier por hacer un excelente grupo, siempre pendiente y juntos lograr este trabajo de la mejor manera.

RESUMEN

La planificación del horneado en los puntos de venta de una cadena de panaderías de Quito se realiza en un contexto de alta variabilidad de la demanda entre locales, productos, días y horas. El objetivo fue desarrollar un modelo predictivo para estimar la demanda horaria observada por local y producto, como apoyo a la planificación del horneado.

Se aplicó un enfoque cuantitativo de aprendizaje automático supervisado. A partir de 2.910.313 registros de ventas POS del período 2024–2026, se construyó una base final de 498.750 observaciones horarias correspondientes a diez locales y quince productos. Se evaluaron Ridge Regression, Random Forest, XGBoost, LightGBM, Support Vector Regression y Holt-Winters mediante una partición temporal. El MAE fue el criterio principal; RMSE y R^2 se utilizaron como métricas complementarias y MAPE como referencia descriptiva.

XGBoost y LightGBM presentaron el mejor desempeño agregado entre los modelos horarios, con valores de MAE cercanos a 3 unidades por hora y un R^2 aproximado de 0,31. Ningún algoritmo fue superior en todos los locales: LightGBM obtuvo el menor MAE en cuatro, XGBoost en tres y SVR en tres. Las devoluciones de producto no vendido se utilizaron únicamente como información complementaria y no formaron parte de la variable objetivo ni de las características predictoras.

Se concluye que los registros históricos de ventas POS permiten generar estimaciones horarias reproducibles de la demanda observada para apoyar la planificación del horneado. La hipótesis se considera respaldada dentro del alcance predictivo del estudio, sin demostrar efectos directos sobre devoluciones, ventas, disponibilidad o cantidades efectivamente horneadas.

Palabras clave: predicción de demanda, aprendizaje automático, productos perecibles, panadería, horneo, series temporales.

ABSTRACT

Baking planning at the points of sale of a bakery chain in Quito takes place in a context of high demand variability across stores, products, days, and hours. The objective was to develop a predictive model to estimate observed hourly demand by store and product, as support for baking planning.

A quantitative supervised machine learning approach was applied. From 2,910,313 POS sales records covering the 2024–2026 period, a final dataset of 498,750 hourly observations corresponding to ten stores and fifteen products was constructed. Ridge Regression, Random Forest, XGBoost, LightGBM, Support Vector Regression, and Holt-Winters were evaluated using a temporal data split. MAE was used as the primary criterion; RMSE and R^2 were used as complementary metrics, and MAPE as a descriptive reference.

XGBoost and LightGBM showed the best overall performance among the hourly models, with MAE values close to 3 units per hour and an approximate R^2 of 0.31. No algorithm was superior across all stores: LightGBM achieved the lowest MAE in four stores, XGBoost in three, and SVR in three. Unsold product returns were used only as complementary information and were not included in the target variable or the predictive features.

It is concluded that historical POS sales records make it possible to generate reproducible hourly estimates of observed demand to support baking planning. The hypothesis is considered supported within the predictive scope of the study, without demonstrating direct effects on returns, sales, availability, or actual quantities baked.

Keywords: demand forecasting, machine learning, perishable products, bakery, baking planning, time series.

Índice

Capítulo 1	1
1. Introducción	1
1.1 Definición del Proyecto.....	3
1.2 Justificación e Importancia del Trabajo de Investigación	6
1.3 Alcance del Proyecto.....	9
1.4 Objetivos	12
Capítulo 2	14
2. Revisión de Literatura.....	14
2.1 Estado del Arte.....	14
2.2 Marco Teórico	22
2.2.1 Inteligencia artificial aplicada a la predicción de demanda	22
2.2.2 Predicción de demanda en productos perecibles y en el contexto de la panadería	23
2.2.3 Series temporales y representación de las variables temporales	25
2.2.4 Modelos predictivos aplicados a la demanda.....	28
2.2.5 Heterogeneidad de la demanda por local y producto	31
2.2.6 Devoluciones de producto no vendido como capa analítica complementaria	33
2.2.7 Evaluación, selección e interpretación operativa de los modelos	36
2.3 Terminología	40
Capítulo 3	47
3. Desarrollo.....	47
3.1 Diseño metodológico de la investigación	47
3.2 Desarrollo del Trabajo	49
3.3 Fuente y características de los datos	52
3.3.1 Fuente de los datos y justificación de la elección	52
3.3.2 Descripción y estructura del conjunto de datos.....	53
3.3.3 Volumen y período de información	56
3.3.4 Consideraciones de calidad, alcance y posibles sesgos de los datos.....	58
3.3.5 Procedimiento de extracción y consolidación de los datos	59
3.4 Preparación y procesamiento de datos	60
3.4.1 Limpieza y validación de los datos	60
3.4.2 Transformación y construcción de variables temporales	62
3.4.3 Tratamiento analítico de las devoluciones	63

3.4.4 Selección de locales y productos.....	65
3.4.5 Construcción del conjunto de datos horario.....	67
3.4.6 Ingeniería de características para el modelado.....	69
3.4.7 Partición temporal de los datos	73
3.5 Metodología de modelado predictivo	75
3.5.1 Enfoque de modelado.....	76
3.5.2 Modelos predictivos evaluados	78
3.5.3 Configuración e hiperparámetros de los modelos	81
3.5.4 Métricas de evaluación.....	85
3.5.5 Criterio de comparación y selección de modelos.....	89
Capítulo 4	91
4. Resultados y Discusión	91
4.1 Análisis exploratorio de los datos	91
4.2 Resultados del modelo predictivo.....	101
4.2.1. Desempeño en entrenamiento, validación y prueba.....	102
4.2.2. Comparación de modelos en el conjunto de prueba.....	105
4.2.3. Resultados por local en conjunto de prueba.....	109
4.2.4 Validación con datos posteriores	112
4.2.5 Síntesis comparativa de los modelos.....	114
4.3 Interpretación de resultados	116
4.4 Discusión	119
4.4.1 Contraste con la literatura	119
4.4.2 Explicación de los resultados obtenidos	123
4.4.3 Hallazgos inesperados	127
4.4.4 Implicaciones	129
4.4.5 Discusión de las limitaciones	131
4.4.6 Grado de validación de la hipótesis	135
Capítulo 5	139
5. Conclusiones y Recomendaciones.....	139
5.1 Conclusiones	139
5.2 Recomendaciones	142
6. Referencias Bibliográficas.....	147
Anexo 1 – Prototipo académico para consulta de predicciones de demanda	152

Índice de Tablas

Tabla 1 Principales variables de la capa de ventas POS.....	54
Tabla 2 Principales variables de la capa de devoluciones	55
Tabla 3 Volumen inicial de registros por fuente y año	57
Tabla 4 Franjas horarias utilizadas en el análisis.....	62
Tabla 5 Variables predictoras utilizadas en los modelos	71
Tabla 6 Distribución temporal del conjunto de modelado.....	74
Tabla 7 Modelos predictivos evaluados en la investigación.....	80
Tabla 8 Configuración e hiperparámetros de los modelos predictivos.....	84
Tabla 9 Métricas utilizadas para evaluar los modelos predictivos	87
Tabla 10 Criterios utilizados para la comparación de los modelos predictivos.....	90
Tabla 11 Desempeño promedio de los modelos en entrenamiento, validación y prueba ..	102
Tabla 12 Comparación de los modelos horarios en el conjunto de prueba	105
Tabla 13 Número de locales con mejor desempeño por métrica y modelo	107
Tabla 14 Modelo con menor MAE por local en el conjunto de prueba.....	109
Tabla 15 Validación de los modelos con datos posteriores para Cumbayá–Enrollado.....	112

Índice de Figuras

Figura 1 Flujo metodológico de la investigación	51
Figura 2 Serie temporal diaria de ventas con medias móviles	92
Figura 3 Mapa de calor: ventas por día de semana por hora del día	93
Figura 4 Mapa de calor: ventas por local por hora del día	94
Figura 5 Perfil de demanda horaria relativo por local	95
Figura 6 Tasa de devolución por local	96
Figura 7 Relación entre ventas y devoluciones por local y mes.....	97
Figura 8 Ranking compuesto de locales	98
Figura 9 Top 15 de productos de panadería seleccionados.....	100
Figura 10 Flujo general del prototipo académico para consulta de predicciones de demanda	152
Figura 11 Interfaz del prototipo académico para consulta de predicciones de demanda .	155

Capítulo 1

1. Introducción

La transformación de los datos operativos en información útil para la toma de decisiones constituye una de las principales aplicaciones de la inteligencia artificial en las organizaciones. En sectores caracterizados por una demanda variable y por productos de vida útil limitada, la capacidad de anticipar el comportamiento de las ventas puede aportar información para mejorar la planificación y apoyar decisiones relacionadas con la disponibilidad de producto.

En una cadena de panaderías con procesos de terminación y horneado en los puntos de venta, la decisión sobre cuánto producto preparar y en qué momento hacerlo tiene una incidencia directa en la operación diaria. Una cantidad superior a las ventas observadas puede incrementar el producto no vendido al cierre de la jornada, mientras que una disponibilidad insuficiente puede limitar la atención de la demanda existente. Esta decisión se vuelve más compleja debido a que el comportamiento de las ventas varía entre locales, productos, días y horas.

En este contexto, la investigación tuvo como propósito desarrollar un modelo predictivo para estimar la demanda horaria observada de productos de panadería por local y producto, utilizando como fuente principal los registros históricos de ventas del sistema de punto de venta. Como información complementaria se analizaron las devoluciones de producto no vendido, con el propósito de contextualizar diferencias operativas entre locales, sin considerarlas una medición directa de las cantidades horneadas, de la disponibilidad o de la merma total.

El estudio utilizó información correspondiente al período 2024–2026. La extracción inicial consolidó 2.910.313 registros de ventas POS y 116.710 registros de devoluciones. Después de los procesos de limpieza, selección y transformación, se obtuvo un subconjunto de 959.692 registros transaccionales correspondiente a diez locales y quince productos de panadería. A partir de estos registros se construyó una base horaria final de 498.750 observaciones utilizada para el entrenamiento, validación y prueba de los modelos predictivos.

La metodología comprendió la preparación y exploración de los datos, la construcción de variables temporales, cíclicas, categóricas e históricas, la partición temporal de la información y la comparación entre modelos específicos por local y modelos globales. Se evaluaron Ridge Regression, Random Forest, XGBoost, LightGBM, Support Vector Regression y Holt-Winters. El desempeño se analizó utilizando el error absoluto medio (MAE) como criterio principal, la raíz del error cuadrático medio (RMSE) y el coeficiente de determinación (R^2) como métricas complementarias, y el error porcentual absoluto medio (MAPE) como referencia descriptiva.

Como desarrollo complementario se implementó un prototipo académico para consultar las estimaciones generadas por el modelo. Este componente tuvo una finalidad demostrativa y no constituye una implementación productiva ni una integración con los sistemas transaccionales de la organización. Tampoco genera órdenes automáticas de horneo ni sustituye el criterio del personal responsable de la operación.

El documento se organiza en cinco capítulos. El primero presenta la definición del proyecto, su justificación, alcance, objetivos e hipótesis. El segundo desarrolla el estado del arte, el marco teórico y la terminología utilizada. El tercero describe el diseño

metodológico, las fuentes de datos, su preparación y el procedimiento de modelado predictivo. El cuarto presenta los resultados, su interpretación y discusión. El quinto contiene las conclusiones y recomendaciones derivadas de la investigación.

1.1 Definición del Proyecto

El presente proyecto de investigación se orienta al desarrollo de una solución de inteligencia artificial aplicada para apoyar la planificación del horneo en los puntos de venta de una cadena de panaderías ubicada en Quito y sus alrededores. La propuesta parte de una necesidad operativa concreta: mejorar la capacidad de estimar la demanda horaria observada, de manera que los locales cuenten con una referencia cuantitativa para decidir cuánto producto preparar y en qué momento hacerlo.

El problema actual radica en que las decisiones de horneo se realizan en un entorno de alta variabilidad, debido a que el comportamiento de las ventas cambia según el local, el producto, la hora del día y el día de la semana. En la práctica, estas decisiones se apoyan principalmente en la experiencia del personal, la observación de las ventas recientes y criterios operativos propios de cada local, sin que exista una herramienta predictiva que utilice sistemáticamente los registros históricos disponibles.

Esta situación puede ocasionar dos efectos opuestos. Por una parte, cuando la cantidad preparada supera las ventas registradas, puede incrementarse el producto no vendido al cierre de la jornada. Por otra, una cantidad insuficiente puede afectar la disponibilidad y generar oportunidades de venta no atendidas. Sin embargo, debido a que el sistema no registra cada evento de horneo ni la disponibilidad efectiva de los productos

durante toda la jornada, estos efectos no pueden medirse de manera completa únicamente a partir de los datos transaccionales.

Como evidencia complementaria de la dimensión operativa del problema, la tasa global de devolución de producto no vendido alcanzó el 6,08 %, con diferencias entre locales que variaron entre el 2,51 % y el 16,33 %. Esta tasa se calculó como la proporción de unidades devueltas respecto del total estimado mediante la suma de las unidades vendidas y devueltas. Los valores obtenidos muestran que la proporción de producto registrado como no vendido no es uniforme entre los locales y sugieren que el fenómeno presenta características particulares según cada punto de venta. Debido a las limitaciones de los registros disponibles, este indicador no representa una medición exacta de las cantidades horneadas, la disponibilidad ni la merma total.

La empresa objeto de estudio cuenta con una planta de producción y varios puntos de venta. Para el desarrollo del modelo se seleccionaron diez locales bajo operación directa de la empresa y quince productos pertenecientes a la categoría de panadería. El modelo operativo combina la preparación centralizada con procesos de terminación y horneado en los locales, lo que requiere adaptar las decisiones a las variaciones de la demanda a lo largo del día.

Frente a esta situación, el proyecto construye una base analítica a partir de los registros históricos del sistema de punto de venta y desarrolla modelos predictivos para estimar la demanda horaria observada por local y producto. La variable objetivo corresponde a la cantidad de unidades vendidas durante una hora específica, registrada mediante las transacciones del sistema de punto de venta.

Las devoluciones de producto no vendido se mantienen como una capa analítica complementaria. Estos registros corresponden a movimientos internos efectuados al cierre de la jornada y se utilizan para contextualizar las diferencias operativas entre locales y productos. No se incorporan como variable objetivo ni como predictor del modelo, y tampoco se consideran una medición directa del volumen horneado.

Por tanto, el problema de investigación se concreta en la ausencia de una herramienta basada en datos que permita transformar los registros históricos de ventas en estimaciones horarias diferenciadas por local y producto. La solución propuesta busca reducir la incertidumbre asociada con la planificación del horneado mediante una referencia cuantitativa, sin sustituir la experiencia del personal ni generar órdenes automáticas de producción.

1.1.1 Perfil de la organización

La empresa objeto de estudio es una cadena de panaderías ecuatoriana dedicada a la producción, distribución y comercialización de productos de panadería. Para preservar la confidencialidad institucional, no se menciona su razón social ni su nombre comercial.

Su modelo operativo combina la preparación centralizada de determinados productos con procesos de terminación y horneado en los locales. Esta dinámica permite ofrecer productos frescos durante la jornada y requiere una planificación continua de las cantidades que deben prepararse.

Los locales presentan diferencias en ubicación, volumen de clientes, horarios y patrones de consumo. Estas condiciones generan comportamientos de venta distintos entre puntos de venta y aumentan la complejidad de la planificación operativa.

La empresa utiliza un sistema de planificación de recursos empresariales y un sistema de punto de venta para registrar sus operaciones comerciales y logísticas. Las transacciones del sistema de punto de venta contienen información sobre el producto, la cantidad vendida, el local y la fecha y hora de la operación. Estos registros constituyeron la fuente principal de información de la investigación. De manera complementaria, se analizaron los movimientos internos de producto no vendido al cierre de la jornada.

En este contexto, el estudio utiliza la información histórica disponible para desarrollar un modelo predictivo orientado a apoyar la planificación del horneo.

1.2 Justificación e Importancia del Trabajo de Investigación

La investigación se justifica por la necesidad de fortalecer la planificación del horneo en los puntos de venta de una cadena de panaderías, en un contexto donde la demanda presenta diferencias según el local, el producto, el día y la hora. Estas variaciones dificultan la definición de las cantidades que deben prepararse durante la jornada y aumentan la incertidumbre asociada con la disponibilidad de producto y con las unidades que permanecen sin vender al cierre del día.

Una estimación más sistemática de la demanda horaria puede aportar una referencia cuantitativa para complementar la experiencia del personal. Su aplicación podría contribuir a mejorar la planificación de las cantidades preparadas y a reducir la exposición a dos situaciones operativas: la disponibilidad insuficiente de determinados productos y la acumulación de unidades no vendidas. Sin embargo, el modelo se plantea como una herramienta de apoyo y no como un mecanismo de automatización de la producción ni como una garantía directa de reducción de merma o incremento de ventas.

Los principales beneficiarios potenciales son la empresa, el personal de los locales y los clientes. La empresa podría disponer de información adicional para analizar la demanda y orientar sus decisiones operativas. El personal contaría con una referencia basada en los registros históricos para complementar su conocimiento de cada local. Para los clientes, una planificación mejor informada podría favorecer la disponibilidad de productos en los momentos de mayor demanda.

Desde el punto de vista académico, el estudio aplica técnicas de inteligencia artificial y aprendizaje automático a un problema real del sector de alimentos perecibles. La investigación integra el tratamiento de datos transaccionales, la representación de patrones temporales, la comparación de modelos predictivos y la interpretación de resultados dentro de un contexto operativo específico. Esta combinación permite vincular el desarrollo técnico con una necesidad concreta de la organización y resulta coherente con el enfoque aplicado de la maestría.

El aporte metodológico específico se encuentra en la construcción y evaluación de una estrategia de predicción con granularidad horaria, diferenciada por local y producto. El estudio compara modelos específicos por local con modelos globales construidos a partir de la información consolidada de los puntos de venta. Esta comparación permite analizar si los patrones particulares de cada local pueden representarse mejor mediante modelos especializados o si existen comportamientos compartidos que pueden ser aprendidos mediante una aproximación conjunta.

Otro elemento diferenciador es la separación metodológica entre la demanda horaria observada y las devoluciones de producto no vendido. Las ventas registradas en el sistema de punto de venta constituyen la fuente principal para construir la variable objetivo,

mientras que las devoluciones se mantienen como una capa analítica complementaria. Este tratamiento evita utilizarlas como una medición directa del horneo o como una variable disponible al momento de generar la predicción y permite emplearlas únicamente para contextualizar diferencias operativas entre locales y productos.

La contribución del estudio también se relaciona con el contexto analizado. Entre los antecedentes revisados sobre predicción de demanda en alimentos, retail y panaderías, no se identificó de manera equivalente la combinación de granularidad horaria, diferenciación por local y producto, comparación entre modelos específicos por local y modelos globales y una operación con terminación y horneo en el punto de venta. En consecuencia, el trabajo aporta evidencia aplicada sobre el comportamiento de distintos modelos en una cadena de panaderías ecuatoriana, sin pretender generalizar automáticamente sus resultados a otras organizaciones.

Desde el punto de vista tecnológico, la investigación evidencia la posibilidad de transformar registros históricos de ventas en una base analítica estructurada y en estimaciones de demanda horaria mediante un proceso reproducible de preparación de datos, entrenamiento y evaluación de modelos predictivos. Este enfoque permite aprovechar información transaccional disponible en la organización y convertirla en resultados cuantitativos que pueden servir como apoyo para la planificación operativa.

El proyecto presenta, además, condiciones favorables de viabilidad. La extracción inicial consolidó 2.910.313 registros de ventas POS y 116.710 registros de devoluciones correspondientes al período 2024–2026. La disponibilidad de esta información, junto con los atributos descriptivos de productos, locales y bodegas, permitió desarrollar las etapas de

preparación, análisis exploratorio, selección de datos, entrenamiento y evaluación de los modelos.

La investigación también fue viable dentro del tiempo establecido para el trabajo de titulación. El problema se delimitó a diez locales bajo operación directa, quince productos de la categoría de panadería y una variable objetivo definida como la cantidad de unidades vendidas por hora, local y producto. Además, se excluyeron la implementación productiva, la automatización del horneo y la optimización integral de los procesos logísticos y comerciales. Esta delimitación permitió mantener un alcance compatible con los recursos técnicos, la disponibilidad de información y los objetivos académicos del proyecto.

La investigación es pertinente porque aborda una necesidad operativa real mediante un enfoque cuantitativo y reproducible. Su importancia radica en aprovechar los registros históricos disponibles para generar información que pueda apoyar la planificación del horneo. El aporte del estudio se concreta en la predicción horaria por local y producto, la comparación entre modelos específicos por local y modelos globales, y el tratamiento de las devoluciones como información analítica complementaria, dentro de un alcance claramente delimitado.

1.3 Alcance del Proyecto

El proyecto se enmarcó en el desarrollo de una solución de inteligencia artificial aplicada para estimar la demanda horaria observada de productos de panadería en los puntos de venta seleccionados de una cadena ecuatoriana. Su propósito fue generar una referencia cuantitativa que apoyara la planificación del horneo, sin abarcar de manera integral los procesos productivos, logísticos y comerciales de la empresa.

El alcance principal comprendió la construcción de una base analítica a partir de los registros históricos de ventas del sistema de punto de venta. Estas transacciones se utilizaron para representar la variable objetivo, definida como la cantidad de unidades vendidas por producto, local y hora. De manera complementaria, se analizaron los movimientos internos de producto no vendido registrados al cierre de la jornada.

Las ventas y las devoluciones se mantuvieron como capas de información diferenciadas debido a su distinta granularidad y función metodológica. Las ventas constituyeron la fuente principal para el modelado predictivo, mientras que las devoluciones se emplearon únicamente para contextualizar el comportamiento del producto registrado como no vendido. Por tanto, no se incorporaron como variable objetivo ni como predictor de los modelos y tampoco se consideraron una medición directa de las cantidades horneadas.

El proyecto comprendió cuatro componentes principales. El primero correspondió a la identificación, extracción, depuración y consolidación de los datos históricos disponibles. El segundo consistió en la preparación de una base analítica estructurada y en la selección de los locales y productos incluidos en el estudio. El tercero comprendió el análisis exploratorio de los patrones de venta y devolución. El cuarto correspondió al desarrollo de modelos predictivos de demanda horaria y a la comparación de su desempeño para identificar las alternativas con mejores resultados dentro del contexto analizado.

El estudio utilizó información estructurada de naturaleza transaccional y operativa. El modelado se concentró en diez locales bajo operación directa de la empresa y quince productos pertenecientes a la categoría de panadería. Los registros analizados

correspondieron al período disponible entre 2024 y 2026, respetando el orden cronológico de la información durante las etapas de entrenamiento, validación y prueba.

El componente predictivo se orientó a estimar la demanda horaria observada para cada combinación de local y producto. Los resultados se interpretaron como una referencia para apoyar decisiones relacionadas con cuánto y cuándo hornear. En consecuencia, las predicciones no se consideraron órdenes automáticas de producción ni sustituyeron el criterio operativo del personal.

Como parte del análisis complementario, se examinó la relación entre las unidades vendidas y las unidades devueltas al cierre de la jornada. Esta aproximación permitió contextualizar diferencias entre locales y productos, pero no tuvo como finalidad reconstruir con precisión los eventos de horneo, la disponibilidad horaria, la merma total ni cada lote preparado en los puntos de venta.

No formó parte del alcance la implementación del modelo dentro del sistema de planificación de recursos empresariales o del sistema de punto de venta. Tampoco se contempló la automatización del horneo, la optimización integral de la producción ni el despliegue de una aplicación corporativa en un ambiente productivo.

Como desarrollo complementario al alcance aprobado, se implementó un prototipo académico de consulta para visualizar las predicciones y mostrar de manera funcional la relación entre los datos de entrada, el modelo entrenado y los resultados generados. Este componente fue concebido como apoyo para la presentación y defensa del trabajo y no constituye una implementación integrada con los sistemas productivos de la empresa.

Desde el punto de vista geográfico y organizacional, la investigación se limitó a la cadena de panaderías analizada y a los locales seleccionados. En consecuencia, los resultados no pueden generalizarse directamente a otras panaderías, cadenas comerciales o contextos de productos perecibles.

En síntesis, el alcance incluyó la preparación de los datos, el análisis exploratorio, el desarrollo de modelos predictivos de demanda horaria, la comparación de su desempeño como parte del proceso de selección y la formulación de recomendaciones para una posible aplicación futura. Se excluyeron la reconstrucción exacta del proceso de horneo, la medición completa de la disponibilidad y la merma, la automatización productiva y la implementación del modelo en los sistemas empresariales.

1.4 Objetivos

1.4.1 Objetivo General

Desarrollar un modelo predictivo para estimar la demanda horaria observada de productos de panadería por local y producto, como apoyo a la planificación del horneo en los puntos de venta seleccionados de una cadena de panaderías ecuatoriana.

1.4.2 Hipótesis

La aplicación de modelos predictivos basados en registros históricos de ventas POS permitirá estimar la demanda horaria observada de productos de panadería por local y producto, como apoyo a la planificación del horneo en los puntos de venta seleccionados.

1.4.3 Objetivos Específicos

Construir una base analítica horaria a partir de los registros históricos de ventas POS y de los atributos descriptivos de productos y locales, mediante procesos de depuración, consolidación y transformación de la información.

Analizar los patrones de la demanda horaria observada por local, producto y período temporal, considerando las devoluciones de producto no vendido como información complementaria para contextualizar las diferencias identificadas.

Desarrollar modelos predictivos de demanda horaria mediante enfoques de aprendizaje automático y series temporales, considerando modelos específicos por local y modelos globales aplicados a los locales y productos seleccionados.

Determinar los modelos con mejor desempeño predictivo mediante la comparación de sus resultados, utilizando el MAE como criterio principal y el RMSE y el R^2 como métricas complementarias sobre el conjunto de prueba definido.

Interpretar los resultados del modelado predictivo y del análisis complementario de las devoluciones para establecer su utilidad como apoyo a la planificación del horneo.

Capítulo 2

2. Revisión de Literatura

2.1 Estado del Arte

La predicción de demanda constituye una aplicación relevante de la analítica predictiva y el aprendizaje automático en contextos comerciales donde las decisiones de producción, abastecimiento e inventario dependen de comportamientos variables a lo largo del tiempo. En el caso de productos perecibles, esta problemática adquiere especial importancia debido a que las diferencias entre la cantidad disponible y la demanda pueden relacionarse con producto no vendido o con insuficiencia de disponibilidad. La literatura reciente muestra la utilización de diferentes enfoques estadísticos y de aprendizaje automático para abordar este tipo de problemas, aunque sus resultados dependen del contexto, la granularidad de los datos, las variables utilizadas y el procedimiento de evaluación aplicado (Muth et al., 2025).

En el ámbito del retail, la literatura también evidencia que la predicción puede plantearse en diferentes niveles de agregación, desde estimaciones globales hasta pronósticos por tienda y producto. Fildes et al. (2022) señalan que los problemas de forecasting en retail varían según el nivel de decisión y que las predicciones a nivel de producto y tienda se relacionan especialmente con decisiones operativas. Esta diversidad de niveles de análisis incrementa la complejidad del problema, debido a que los patrones de venta pueden diferir entre productos, locales y períodos temporales.

Para analizar los antecedentes de manera crítica, en esta investigación se consideran cuatro dimensiones principales: la granularidad temporal del pronóstico, el contexto

operativo, los enfoques y variables utilizados y las métricas o resultados reportados. Esta estructura permite identificar semejanzas y diferencias entre los estudios y evita realizar comparaciones directas entre resultados obtenidos sobre conjuntos de datos y condiciones experimentales diferentes.

Uno de los antecedentes más próximos al contexto de esta investigación corresponde a Huber y Stuckenschmidt (2020), quienes estudiaron la predicción de demanda diaria en una cadena de panaderías, con especial atención a días calendáricos particulares. El problema fue formulado como una tarea de aprendizaje supervisado a nivel de tienda y categoría de producto, y se compararon diferentes métodos, entre ellos redes neuronales artificiales y modelos basados en *gradient boosting*. Los autores encontraron que los enfoques de aprendizaje automático presentaron ventajas frente a métodos tradicionales dentro del contexto evaluado y analizaron, además, formulaciones de clasificación y regresión. Este antecedente es particularmente relevante porque demuestra la aplicabilidad del aprendizaje automático en una operación de panadería; sin embargo, trabaja con demanda diaria por categoría de producto, mientras que la presente investigación analiza demanda horaria por producto y local (Huber & Stuckenschmidt, 2020).

Desde una perspectiva más amplia del pronóstico temporal, Lim et al. (2021) desarrollaron el modelo Temporal Fusion Transformer (TFT), una arquitectura orientada a pronósticos multihorizonte e interpretables. El modelo integra procesamiento recurrente, mecanismos de atención, selección de variables y capas de control para representar relaciones entre covariables estáticas, entradas conocidas a futuro y series históricas. Los autores reportaron mejoras frente a diferentes modelos de referencia en varios conjuntos de

datos y destacaron la capacidad del enfoque para aportar elementos de interpretabilidad. Aunque este trabajo no se desarrolla específicamente en una cadena de panaderías, constituye un antecedente metodológico relevante porque evidencia la importancia de representar relaciones temporales complejas y combinar diferentes tipos de variables dentro de los modelos predictivos (Lim et al., 2021).

En el ámbito de los servicios de alimentación, Rodrigues et al. (2024) analizaron modelos de aprendizaje automático para la predicción de demanda de corto plazo con el propósito de contribuir a la reducción del desperdicio de alimentos. El estudio utilizó datos diarios reales provenientes de tres servicios de catering y comparó diferentes enfoques predictivos con la demanda efectivamente registrada. Los resultados mostraron desempeños favorables para determinados modelos y estimaron reducciones de desperdicio entre el 14 % y el 52 % bajo los escenarios evaluados. Este antecedente resulta importante porque relaciona la predicción de demanda con consecuencias operativas sobre productos perecibles; sin embargo, su unidad de análisis, contexto y granularidad son diferentes a los del presente estudio, por lo que dichos porcentajes no pueden extrapolarse directamente a una operación de panadería con planificación horaria del horneado (Rodrigues et al., 2024).

Seyam et al. (2025) propusieron un modelo de *stacking ensemble* para la predicción de demanda de alimentos en retail, combinando Random Forest, Support Vector Regression, XGBoost y LSTM como modelos base y Ridge Regression como metamodelo. La propuesta fue comparada con modelos individuales y alcanzó, en el conjunto de datos analizado, un coeficiente de determinación R^2 de 0,99, un MAE de 0,63 y un MAPE de 1,8 %. Los autores también evaluaron la utilización de las predicciones para apoyar decisiones de reposición orientadas a reducir desperdicio. Este estudio evidencia el potencial de

combinar algoritmos con características diferentes; no obstante, sus métricas corresponden a un conjunto de datos, escala y protocolo experimental propios, por lo que no deben utilizarse como referencia numérica directa para juzgar el desempeño obtenido en una predicción horaria de productos de panadería (Seyam et al., 2025).

Otra línea relevante corresponde al reconocimiento de la heterogeneidad existente dentro de los datos. Jauhar et al. (2025) propusieron un enfoque en dos etapas que combina segmentación de clientes y predicción de demanda en cadenas de suministro de productos perecibles. Los autores utilizaron técnicas de *clustering*, inteligencia artificial explicable y diferentes métodos de pronóstico. Entre sus resultados, reportaron que una estrategia basada en segmentación mediante agrupamiento jerárquico y modelos LSTM presentó un RMSE ponderado inferior al obtenido por el enfoque no segmentado. El estudio muestra que diferenciar grupos con características distintas puede mejorar la representación de la demanda, aspecto conceptualmente relacionado con la decisión de esta investigación de analizar las particularidades existentes entre locales, aunque ambas investigaciones utilizan unidades de segmentación y conjuntos de datos diferentes (Jauhar et al., 2025).

La aplicación práctica de modelos predictivos en panaderías también ha sido estudiada desde una perspectiva sociotécnica. Fries y Ludwig (2024) analizaron los desafíos relacionados con la utilización de aprendizaje automático para pronosticar ventas en una panadería. Su trabajo muestra que los sistemas de predicción pueden considerar variables como precios, promociones y condiciones climáticas, pero también destaca que su utilización efectiva depende de factores asociados con los usuarios, los procesos organizacionales y la interpretación de las predicciones. Este antecedente es relevante para el presente proyecto porque refuerza la idea de que el valor de una estimación no depende

únicamente de sus métricas estadísticas, sino también de su capacidad para incorporarse de manera comprensible dentro de las decisiones operativas (Fries & Ludwig, 2024).

Desde una perspectiva más transversal, Muth et al. (2025) desarrollaron una revisión sistemática sobre la aplicación del aprendizaje automático a la predicción de demanda bajo condiciones de volatilidad macroeconómica. Los autores identificaron tendencias relacionadas con la incorporación de información prospectiva, indicadores económicos, combinaciones de algoritmos y redes neuronales para abordar entornos dinámicos. Aunque el ámbito de la revisión es más amplio que el sector de panadería, sus resultados muestran que no existe una única estrategia aplicable de manera uniforme a todos los problemas de demanda y que la selección metodológica depende del contexto, de la información disponible y de la naturaleza del fenómeno analizado (Muth et al., 2025).

A gran escala, la competencia M5 constituye otro antecedente relevante para comprender la predicción de ventas en retail. Makridakis et al. (2022) evaluaron métodos de pronóstico sobre 42.840 series temporales jerárquicas de ventas unitarias de Walmart, incluyendo niveles de agregación que permiten analizar el comportamiento desde estructuras consolidadas hasta series detalladas. Sus resultados mostraron diferencias de desempeño entre los métodos evaluados y entre distintos niveles de agregación de las series. Aunque este contexto es diferente al de una cadena de panaderías, el estudio aporta un antecedente relevante sobre la evaluación de métodos predictivos en estructuras con múltiples productos y unidades comerciales.

La revisión conjunta de estos antecedentes permite identificar varios consensos. En primer lugar, los modelos de aprendizaje automático constituyen alternativas viables para representar comportamientos de demanda que presentan relaciones temporales y no

lineales, aunque ningún algoritmo puede considerarse universalmente superior independientemente del conjunto de datos. En segundo lugar, diferentes estudios muestran la importancia de incorporar información temporal, histórica o contextual para representar mejor el fenómeno analizado. Finalmente, los antecedentes orientados a aplicaciones reales evidencian que el desempeño estadístico constituye únicamente una parte del problema y que la utilidad práctica de las predicciones depende también de su relación con las decisiones operativas (Huber & Stuckenschmidt, 2020; Lim et al., 2021; Fries & Ludwig, 2024; Muth et al., 2025).

La comparación también evidencia diferencias metodológicas relevantes. Los antecedentes revisados corresponden a contextos diversos, entre ellos panaderías, servicios de alimentación, retail y cadenas de suministro de productos perecibles. Asimismo, utilizan diferentes granularidades, unidades de análisis, variables y métricas. Por ejemplo, Huber y Stuckenschmidt (2020) estudian demanda diaria a nivel de tienda y categoría; Rodrigues et al. (2024) trabajan con información diaria de servicios de catering; Seyam et al. (2025) evalúan un modelo de ensamblaje en retail alimentario; mientras que Makridakis et al. (2022) trabajan con decenas de miles de series jerárquicas de ventas minoristas. Estas diferencias impiden comparar directamente los valores de MAE, RMSE, MAPE o R^2 reportados por cada trabajo sin considerar previamente la escala de la variable objetivo, la granularidad temporal, el conjunto de datos y el protocolo experimental utilizado.

En relación con la granularidad temporal, los antecedentes más próximos al sector de panadería revisados no presentan la misma combinación utilizada en esta investigación de predicción horaria, producto individual y local. Huber y Stuckenschmidt (2020) trabajan con demanda diaria por categoría y tienda, mientras que Fries y Ludwig (2024) se

concentran principalmente en los desafíos sociotécnicos de la utilización de modelos de pronóstico en una panadería. Esta diferencia resulta relevante para el problema abordado, debido a que la planificación del horneado se realiza durante distintos momentos de la jornada y requiere una referencia temporal más próxima a la frecuencia con la que se toman dichas decisiones.

Otro aspecto de interés es la heterogeneidad entre unidades de análisis. Jauhar et al. (2025) muestran que segmentar grupos con características diferentes puede modificar el desempeño del pronóstico. De manera complementaria, la literatura de retail muestra que los problemas de forecasting pueden organizarse en diferentes niveles, desde estimaciones consolidadas hasta pronósticos específicos por producto y tienda (Fildes et al., 2022). En el contexto de la presente investigación, esta idea se relaciona con la existencia de comportamientos diferenciados entre los puntos de venta y con la necesidad de analizar tanto patrones particulares como compartidos.

Las devoluciones de producto no vendido constituyen otra diferencia respecto de los antecedentes considerados. Los estudios revisados se concentran principalmente en la predicción de demanda, el desperdicio, los niveles de inventario o la disponibilidad de productos. Dentro de este conjunto de antecedentes no se identificó la utilización de registros internos de devolución de producto no vendido con la misma función metodológica. En la presente investigación, esta información se delimita como una capa analítica complementaria y no como parte de la variable objetivo del modelo. Esta decisión permite contextualizar el producto registrado como no vendido sin interpretar las devoluciones como una medición directa del volumen horneado, la disponibilidad o la merma total.

A partir de la revisión realizada, el vacío abordado por la investigación no se plantea como la ausencia absoluta de estudios sobre predicción de demanda en panaderías o productos perecibles. Por el contrario, existen antecedentes que respaldan ampliamente la utilización de aprendizaje automático para estos problemas. El espacio específico de estudio se encuentra en la combinación de granularidad horaria, diferenciación por local y producto, comparación entre modelos específicos por local y modelos globales, y utilización de devoluciones de producto no vendido como información analítica complementaria, dentro de una operación de panadería con procesos de terminación y horneado en los puntos de venta. Esta combinación no fue identificada de manera equivalente en los antecedentes revisados.

El aporte metodológico del proyecto no consiste en desarrollar un algoritmo nuevo, sino en aplicar y comparar de manera estructurada diferentes enfoques predictivos sobre datos reales de una cadena de panaderías y analizar su comportamiento dentro de una problemática operativa específica. La predicción horaria permite aproximar el análisis a la frecuencia con la que se toman las decisiones de horneado; la comparación entre modelos específicos y globales permite estudiar la heterogeneidad entre locales; y la capa complementaria de devoluciones aporta contexto sobre el producto registrado como no vendido sin modificar la variable objetivo del modelo.

El estado del arte muestra que la predicción de demanda en alimentos, retail y productos perecibles constituye un campo con múltiples alternativas metodológicas y evidencia favorable respecto de la utilización de técnicas de aprendizaje automático. Sin embargo, también confirma que los resultados dependen del contexto, la granularidad, la estructura del conjunto de datos y las métricas empleadas. Frente a estas condiciones, la

presente investigación se posiciona como un estudio aplicado orientado a estimar la demanda horaria observada por local y producto, comparar modelos específicos por local y modelos globales y complementar la interpretación mediante información de devoluciones de producto no vendido, con el propósito de generar una referencia cuantitativa para apoyar la planificación del horneo.

2.2 Marco Teórico

2.2.1 Inteligencia artificial aplicada a la predicción de demanda

La inteligencia artificial reúne distintos enfoques computacionales orientados al desarrollo de sistemas capaces de aprender a partir de información, identificar patrones y generar resultados para tareas como clasificación o predicción. Dentro de este campo, el aprendizaje automático permite construir modelos a partir de datos históricos para identificar relaciones y patrones sin requerir que todas las reglas de decisión sean definidas explícitamente de forma previa (Russell & Norvig, 2021; Géron, 2022).

En problemas empresariales, estas técnicas se utilizan para analizar grandes volúmenes de información y generar estimaciones sobre comportamientos futuros. La predicción de demanda constituye una de sus aplicaciones, debido a que permite utilizar registros históricos y variables asociadas con el entorno para estimar comportamientos futuros de ventas o demanda. La literatura muestra que los métodos de aprendizaje automático han adquirido relevancia en este ámbito debido a su capacidad para representar relaciones no lineales y patrones complejos presentes en los datos (Muth et al., 2025).

La utilización de aprendizaje automático en predicción de demanda no implica la existencia de un único algoritmo adecuado para todos los contextos. El desempeño de los

modelos depende de factores como la calidad y cantidad de información disponible, la estructura temporal de los datos, las variables incorporadas, el nivel de agregación de la demanda y el procedimiento de validación utilizado. Por ello, la selección de un modelo predictivo requiere analizar su comportamiento dentro del problema específico y no únicamente considerar el desempeño reportado en otros conjuntos de datos (Muth et al., 2025).

Desde una perspectiva organizacional, el valor de una solución predictiva tampoco se limita a la exactitud estadística de sus resultados. La incorporación de modelos de aprendizaje automático en procesos reales requiere considerar la forma en que las predicciones son interpretadas y utilizadas por las personas responsables de tomar decisiones. En el contexto específico de una panadería, Fries y Ludwig (2024) evidencian que la adopción de sistemas de pronóstico basados en aprendizaje automático involucra factores tecnológicos y organizacionales, así como la interacción entre los modelos, los usuarios y las prácticas existentes.

La inteligencia artificial aplicada a la predicción de demanda puede entenderse como un mecanismo de apoyo a la toma de decisiones, en el que los datos históricos se transforman en estimaciones cuantitativas que deben ser interpretadas dentro de su contexto operativo. Su utilidad depende tanto de la capacidad predictiva del modelo como de la calidad de los datos, la comprensión de sus limitaciones y la forma en que los resultados se incorporan al proceso de decisión (Fries & Ludwig, 2024; Muth et al., 2025).

2.2.2 Predicción de demanda en productos perecibles y en el contexto de la panadería

La predicción de demanda en productos perecibles presenta particularidades relacionadas con su vida útil limitada y con la necesidad de coordinar las cantidades disponibles con el comportamiento esperado de las ventas. A diferencia de productos que pueden conservarse durante períodos prolongados, los alimentos perecibles requieren decisiones operativas sensibles al tiempo, ya que una cantidad disponible superior a la demanda efectivamente atendida puede incrementar el producto no vendido, mientras que una disponibilidad insuficiente puede limitar la atención de la demanda existente. Estas características condicionan las decisiones de producción, inventario y reposición en este tipo de operaciones (van Donselaar et al., 2006).

En este contexto, el pronóstico de demanda constituye una herramienta para anticipar las cantidades requeridas y aportar información a decisiones relacionadas con producción, abastecimiento y disponibilidad. Rodrigues et al. (2024), en un estudio aplicado a servicios de alimentación, analizaron el uso de modelos de aprendizaje automático para estimar demanda de corto plazo y evaluaron su relación con el desperdicio de alimentos y la demanda insatisfecha. Sus resultados muestran la relevancia de contar con estimaciones de demanda que apoyen las decisiones operativas cuando se trabaja con productos de vida útil limitada.

El sector de panadería comparte varias de estas características y presenta además una dinámica de producción y comercialización sensible al momento de la jornada. Los patrones observados pueden variar entre locales, productos y períodos temporales, además de estar influenciados por factores de calendario. Esto hace necesario considerar que la demanda registrada no presenta necesariamente un comportamiento uniforme entre productos ni entre puntos de venta.

Huber y Stuckenschmidt (2020) analizaron la predicción de demanda en una cadena de panaderías mediante técnicas de aprendizaje automático, utilizando información diaria a nivel de tienda y categoría de producto. Los autores evaluaron diferentes métodos y mostraron que determinados enfoques de aprendizaje automático pueden presentar ventajas frente a métodos tradicionales dentro del contexto estudiado. Asimismo, incorporaron información relacionada con días calendáricos especiales, lo que evidencia la importancia de considerar factores temporales en la construcción de los pronósticos.

Un aspecto adicional es la granularidad temporal con la que se representa la demanda. Una estimación diaria permite conocer el volumen esperado para una jornada, pero no describe necesariamente cómo se distribuyen las ventas dentro del día. Cuando las decisiones operativas se toman varias veces durante una misma jornada, una resolución temporal más detallada puede aportar información sobre las variaciones existentes entre diferentes períodos. Por ello, la granularidad seleccionada debe guardar relación tanto con la frecuencia de las decisiones que se pretende apoyar como con la disponibilidad y calidad de los registros históricos.

La literatura revisada muestra que la predicción de productos perecibles no depende únicamente del algoritmo utilizado. También intervienen la escala temporal del pronóstico, las diferencias entre productos y puntos de venta, la estructura de los datos y las condiciones particulares de la operación. En una panadería, estos elementos adquieren especial relevancia debido a la necesidad de coordinar la disponibilidad de productos con patrones de venta que pueden cambiar a lo largo de la jornada (Huber & Stuckenschmidt, 2020; Rodrigues et al., 2024).

2.2.3 Series temporales y representación de las variables temporales

Una serie temporal está constituida por observaciones registradas de manera ordenada a lo largo del tiempo. Su análisis permite identificar estructuras como tendencia, estacionalidad, ciclos y variaciones irregulares, elementos relevantes para comprender el comportamiento histórico de una variable antes de desarrollar un modelo de pronóstico. Hyndman y Athanasopoulos (2021) señalan que el análisis previo de los patrones de una serie constituye una etapa fundamental para seleccionar y aplicar adecuadamente los métodos de predicción.

La estacionalidad se presenta cuando determinados patrones tienden a repetirse con una periodicidad relativamente estable. En problemas de demanda pueden existir comportamientos asociados con la hora del día, el día de la semana, el mes, las temporadas o determinados eventos del calendario. La identificación de estos patrones permite representar la estructura temporal de la información y aporta elementos que pueden ser utilizados por los modelos para estimar observaciones futuras (Hyndman & Athanasopoulos, 2021).

Cuando la información contiene variables temporales de naturaleza periódica, su representación debe conservar la proximidad existente entre posiciones consecutivas dentro de un ciclo. Por ejemplo, las últimas y primeras horas de un día se encuentran temporalmente próximas, aunque una codificación numérica directa pueda ubicarlas en extremos opuestos de una escala. Una situación similar ocurre entre el último y el primer día de la semana o entre el cierre y el inicio de un ciclo anual. En problemas de aprendizaje automático, estas relaciones pueden representarse mediante transformaciones cíclicas basadas en funciones seno y coseno, evitando introducir discontinuidades artificiales en variables periódicas (Kuhn & Johnson, 2019).

Además de las variables de calendario, los modelos de series temporales y aprendizaje automático pueden incorporar información procedente de observaciones anteriores. Los rezagos permiten utilizar valores históricos de la variable como predictores para representar dependencias temporales, mientras que las estadísticas móviles, como medias o desviaciones estándar calculadas sobre ventanas anteriores, permiten resumir el nivel y la variabilidad reciente de la serie. La construcción de este tipo de características forma parte del proceso de ingeniería de variables y puede contribuir a que los modelos representen patrones temporales que no se encuentran expresados directamente en las variables originales (Kuhn & Johnson, 2019; Hyndman & Athanasopoulos, 2021).

La representación de dependencias temporales constituye también un elemento central en modelos de mayor complejidad. Lim et al. (2021), por ejemplo, proponen el Temporal Fusion Transformer para pronósticos multihorizonte y muestran la relevancia de combinar información histórica, variables conocidas a futuro y covariables estáticas mediante mecanismos capaces de capturar relaciones temporales de distintas escalas. Este enfoque evidencia que la selección y representación de las variables temporales constituye una parte fundamental del diseño de un problema predictivo.

La granularidad temporal también influye en la forma en que se construyen y se interpretan las variables derivadas. Athanasopoulos et al. (2017) muestran que una misma serie puede analizarse mediante distintos niveles de agregación temporal y que estas escalas aportan información diferente sobre el comportamiento del fenómeno. Por ello, la frecuencia de los datos debe guardar correspondencia con el objetivo del pronóstico y con la unidad temporal en la que se toman las decisiones.

Cuando las observaciones no conforman una secuencia cronológica continua, la interpretación de los rezagos y de las ventanas móviles requiere especial precaución. En esas condiciones, un determinado número de observaciones anteriores no representa necesariamente un intervalo calendario exacto. Resulta necesario distinguir entre una posición anterior dentro de la secuencia disponible y un desplazamiento temporal estricto, especialmente cuando existen intervalos sin registros.

La representación temporal en modelos de demanda puede combinar variables de calendario, transformaciones cíclicas, rezagos y estadísticas móviles para capturar diferentes dimensiones del comportamiento histórico. La selección de estas representaciones depende de la frecuencia y continuidad de los datos, así como de la estructura temporal del fenómeno analizado, por lo que su interpretación debe mantenerse coherente con las características reales del conjunto de información (Hyndman & Athanasopoulos, 2021; Kuhn & Johnson, 2019; Lim et al., 2021).

2.2.4 Modelos predictivos aplicados a la demanda

La predicción de demanda puede abordarse mediante diferentes familias de modelos estadísticos y de aprendizaje automático. La elección entre estos enfoques depende de las características de los datos, la granularidad temporal, el volumen de observaciones, las relaciones entre las variables y el horizonte de predicción. La literatura muestra que no existe un algoritmo que presente el mejor desempeño de manera universal, por lo que la comparación empírica entre alternativas constituye una práctica relevante para seleccionar modelos adecuados al contexto analizado (Muth et al., 2025; Seyam et al., 2025).

Los modelos lineales constituyen una aproximación básica para representar relaciones entre una variable objetivo y un conjunto de características predictoras. Dentro de esta familia, la regresión Ridge incorpora una penalización L2 sobre los coeficientes, lo que permite controlar su magnitud y reducir problemas asociados con la multicolinealidad y la complejidad del modelo. Esta regularización puede resultar útil cuando existen múltiples variables relacionadas entre sí y se busca mantener una estructura lineal relativamente interpretable (Hoerl & Kennard, 1970).

Los métodos basados en árboles de decisión permiten representar relaciones no lineales sin requerir una especificación previa de la forma funcional entre las variables. Random Forest combina múltiples árboles contruidos sobre diferentes muestras y subconjuntos de características, agregando sus resultados para generar una estimación final. Este procedimiento permite reducir la variabilidad asociada con un árbol individual y representar interacciones complejas entre las variables predictoras (Breiman, 2001).

Los algoritmos de *gradient boosting* construyen modelos de manera secuencial, de modo que cada nueva etapa busca corregir parte de los errores producidos por las anteriores. Dentro de esta familia, XGBoost incorpora mecanismos de regularización y optimizaciones computacionales orientadas a mejorar la eficiencia y controlar la complejidad del modelo (Chen & Guestrin, 2016). LightGBM utiliza una implementación basada en histogramas y estrategias optimizadas de crecimiento de árboles, diseñadas para trabajar eficientemente con conjuntos de datos de gran volumen y múltiples características (Ke et al., 2017).

Otra alternativa corresponde a Support Vector Regression (SVR), extensión de las máquinas de vectores de soporte para problemas de regresión. Este enfoque busca encontrar

una función que mantenga los errores dentro de un margen de tolerancia definido y puede utilizar funciones *kernel* para representar relaciones no lineales entre las variables de entrada y la variable objetivo. Su desempeño puede resultar competitivo en determinados problemas, aunque su costo computacional puede incrementarse cuando aumenta considerablemente el número de observaciones (Drucker et al., 1997; Smola & Schölkopf, 2004).

Junto con los enfoques de aprendizaje automático, los métodos estadísticos de series temporales continúan siendo utilizados como alternativas de pronóstico. Holt-Winters, basado en suavizamiento exponencial, permite representar componentes como nivel, tendencia y estacionalidad mediante actualizaciones sucesivas de la serie. Este método constituye una referencia clásica dentro del pronóstico temporal y resulta especialmente apropiado cuando el comportamiento histórico presenta patrones relativamente estructurados y recurrentes (Winters, 1960; Hyndman & Athanasopoulos, 2021).

La literatura reciente también muestra el uso de arquitecturas de aprendizaje profundo y modelos ensamblados. Lim et al. (2021), mediante Temporal Fusion Transformer, evidencian la capacidad de arquitecturas profundas para representar dependencias temporales complejas y diferentes tipos de covariables. Por su parte, Seyam et al. (2025) muestran que la combinación de varios algoritmos mediante un esquema de *stacking ensemble* puede mejorar el desempeño predictivo en determinados conjuntos de datos de demanda alimentaria. Estos resultados amplían las alternativas disponibles, aunque su conveniencia depende de factores como el volumen y continuidad de los datos, los recursos computacionales y la complejidad del problema.

Las diferencias entre estas familias de modelos responden a distintos mecanismos de aprendizaje. Los modelos lineales priorizan una estructura relativamente simple; los métodos basados en árboles y potenciación por gradiente permiten representar relaciones no lineales e interacciones; SVR utiliza una formulación basada en márgenes y funciones *kernel*; mientras que los métodos estadísticos de series temporales modelan explícitamente componentes temporales. Por ello, el desempeño de cada alternativa debe evaluarse bajo un mismo conjunto de datos y condiciones de validación antes de establecer su pertinencia para un problema específico (Muth et al., 2025).

La diversidad de enfoques disponibles hace que la selección de un modelo deba responder a las características del problema analizado. No debería fundamentarse únicamente en la complejidad del algoritmo o en resultados reportados en otros estudios, sino también en su desempeño bajo condiciones comparables, su estabilidad, su capacidad de generalización y la interpretación de sus resultados dentro del contexto donde serán utilizados.

2.2.5 Heterogeneidad de la demanda por local y producto

La demanda dentro de una cadena comercial puede presentar comportamientos diferentes entre sus puntos de venta y productos. En el ámbito del retail, los problemas de pronóstico pueden plantearse en distintos niveles de agregación, desde estimaciones generales hasta predicciones específicas por tienda y producto, dependiendo de la decisión que se desea apoyar (Fildes et al., 2022). Esta diversidad implica que una representación agregada de la demanda puede no reflejar completamente los patrones particulares presentes en unidades de análisis más específicas.

La heterogeneidad constituye un aspecto relevante en los problemas de predicción. Jauhar et al. (2025) muestran que la incorporación de estructuras diferenciadas mediante segmentación puede modificar el desempeño predictivo frente a enfoques que tratan conjuntamente observaciones con características distintas. Aunque su unidad de análisis corresponde a segmentos de clientes en cadenas de suministro de productos perecibles, el estudio aporta evidencia sobre la conveniencia de reconocer estructuras diferenciadas dentro de conjuntos de datos heterogéneos.

Cuando existen múltiples series relacionadas, la literatura distingue entre modelos locales o específicos y modelos globales. En el enfoque local se ajusta un modelo separado para cada serie, mientras que en el enfoque global un único modelo aprende a partir del conjunto de series disponibles. Montero-Manso y Hyndman (2021) formalizan esta distinción y muestran que los modelos globales pueden aprovechar información conjunta sin requerir necesariamente que todas las series presenten un comportamiento idéntico.

Los modelos globales pueden beneficiarse de un mayor volumen de información y de patrones compartidos entre distintas series, mientras que los modelos específicos permiten concentrar el aprendizaje en las características particulares de una unidad determinada. La conveniencia de uno u otro enfoque depende de factores como la cantidad de series, su longitud, el grado de heterogeneidad y la complejidad del método utilizado. Hewamalage et al. (2022) analizaron estas condiciones mediante experimentos con diferentes escenarios de series temporales y encontraron que los modelos globales pueden mantener un desempeño competitivo incluso ante conjuntos heterogéneos, aunque sus resultados están condicionados por la disponibilidad y complejidad de los datos y por el tipo de modelo empleado.

En el contexto específico de la predicción de demanda, la diferenciación entre unidades comerciales también ha sido considerada en estudios previos. Huber y Stuckenschmidt (2020), al analizar una cadena de panaderías, desarrollaron pronósticos diferenciados por tienda y categoría de producto. De forma más amplia, Fildes et al. (2022) señalan que los pronósticos a nivel de producto y tienda tienen especial relevancia para decisiones operativas en retail. Estos antecedentes respaldan la importancia de definir adecuadamente la unidad sobre la cual se realizará la predicción.

La diferenciación por producto agrega otra dimensión al problema. Artículos comercializados en un mismo local pueden presentar niveles y patrones de venta distintos, de modo que una representación conjunta debe ser capaz de conservar la identificación de cada unidad y, al mismo tiempo, aprovechar posibles regularidades compartidas. Esta consideración resulta relevante cuando se trabaja con múltiples combinaciones de producto y punto de venta.

A partir de estos antecedentes, los modelos específicos y globales pueden entenderse como estrategias complementarias para representar la heterogeneidad presente en grupos de series temporales. Los primeros priorizan el comportamiento particular de cada unidad, mientras que los segundos buscan aprender relaciones compartidas entre varias series. La elección entre ambos enfoques debe sustentarse en su desempeño sobre el conjunto de datos analizado y no en la presunción de que una estrategia será superior en todos los contextos (Montero-Manso & Hyndman, 2021; Hewamalage et al., 2022).

2.2.6 Devoluciones de producto no vendido como capa analítica complementaria

En operaciones relacionadas con productos perecibles, el análisis de la demanda puede complementarse con información sobre las unidades que permanecen sin vender después del período previsto para su comercialización. Este tipo de información aporta contexto adicional al comportamiento observado de las ventas, aunque su interpretación depende de la naturaleza del registro y del procedimiento mediante el cual se genera.

La literatura sobre productos perecibles ha relacionado la calidad de los pronósticos con decisiones de producción, abastecimiento y disponibilidad. Rodrigues et al. (2024), en el contexto de servicios de alimentación, analizaron el efecto potencial de diferentes modelos de predicción sobre el desperdicio y la demanda insatisfecha. Los autores reportaron posibles reducciones de desperdicio bajo los escenarios evaluados, lo que evidencia la relevancia de coordinar las cantidades disponibles con la demanda estimada cuando se trabaja con productos de vida útil limitada (Rodrigues et al., 2024).

Es necesario distinguir entre ventas, demanda, producto no vendido, devoluciones y desperdicio, ya que estos conceptos representan aspectos diferentes de la operación. Las ventas corresponden a las transacciones efectivamente registradas, mientras que la demanda puede incluir requerimientos de clientes que no se concretan en una venta cuando existe insuficiencia de disponibilidad. Por su parte, una devolución constituye un registro operativo asociado con el retorno o movimiento de un producto, pero no equivale necesariamente a desperdicio, merma ni volumen producido.

Esta diferencia resulta especialmente importante cuando los registros de devoluciones se generan al final de la jornada. Un movimiento de producto no vendido puede aportar información sobre las unidades registradas para devolución, pero por sí solo no permite reconstruir todas las decisiones realizadas durante el día. Para ello serían

necesarios datos adicionales, como los eventos de preparación u horneado, los niveles de inventario en distintos momentos, las reposiciones efectuadas y la disponibilidad real de cada producto.

Bajo estas condiciones, las devoluciones pueden utilizarse como una capa analítica complementaria a la información de ventas, siempre que su función se mantenga claramente delimitada. Su análisis puede contribuir a identificar diferencias entre productos, períodos o puntos de venta y aportar contexto para interpretar las ventas observadas, pero no permite establecer causalidad directa ni determinar por sí solo si existió sobreproducción, una planificación inadecuada o una oportunidad de venta perdida.

La utilización conjunta de fuentes con distinta granularidad temporal también requiere precaución. Cuando las ventas se registran a nivel transaccional u horario y las devoluciones únicamente al cierre de la jornada, ambas fuentes describen aspectos relacionados, pero no equivalentes, del proceso. En este escenario, la información de devoluciones puede resultar útil para análisis descriptivos agregados, pero su incorporación directa como predictor de demanda horaria podría introducir información que no estaría disponible en el momento en que debe generarse la predicción.

Este último aspecto se relaciona con el concepto de fuga de información (*data leakage*), entendido como la incorporación al proceso de modelado de información que no estaría legítimamente disponible al momento de realizar una predicción. Kaufman et al. (2012) señalan que este problema puede producir estimaciones artificialmente optimistas del desempeño de un modelo, debido a que parte de la información utilizada durante el aprendizaje contiene conocimiento que solo estaría disponible posteriormente. Por ello, en problemas temporales debe mantenerse una separación clara entre la información

disponible antes de la predicción y aquella conocida después del período que se desea estimar.

Los registros de producto no vendido pueden complementar el análisis de demanda en operaciones con productos perecibles, siempre que su significado se delimite de acuerdo con el proceso que los genera. Las devoluciones permiten contextualizar el comportamiento de las ventas y analizar diferencias entre unidades operativas, pero no deben interpretarse automáticamente como demanda, producción, horneo, merma o desperdicio. Su utilidad depende de mantener explícitas estas limitaciones y de preservar su separación respecto de las variables utilizadas para generar la predicción.

2.2.7 Evaluación, selección e interpretación operativa de los modelos

La evaluación de un modelo predictivo permite determinar qué tan próximas se encuentran sus estimaciones respecto de los valores observados y constituye una etapa fundamental para comparar alternativas de modelado. Las medidas de exactitud presentan propiedades e interpretaciones diferentes, por lo que la evaluación no debería basarse exclusivamente en un único indicador, sino considerar de manera conjunta distintas métricas y las características del problema analizado (Makridakis, 1993; Hyndman & Athanasopoulos, 2021).

El Error Absoluto Medio (MAE) representa el promedio de las diferencias absolutas entre los valores observados y las predicciones. Su interpretación se mantiene en las mismas unidades de la variable objetivo, lo que facilita comprender la magnitud promedio del error. Una de sus ventajas es que los errores contribuyen de manera proporcional a su

magnitud, por lo que resulta menos sensible a valores extremos que las métricas basadas en errores cuadrados (Hyndman & Athanasopoulos, 2021).

La raíz del error cuadrático medio (RMSE) se obtiene a partir de los errores elevados al cuadrado antes de calcular su promedio y expresar nuevamente el resultado en las unidades de la variable objetivo. Debido a esta construcción, penaliza con mayor intensidad los errores de gran magnitud. Puede complementar al MAE cuando interesa identificar modelos que producen desviaciones importantes, aunque su interpretación debe realizarse considerando la escala de la variable analizada (Hyndman & Athanasopoulos, 2021).

El Error Porcentual Absoluto Medio (MAPE) expresa el error en términos porcentuales respecto del valor observado. Esta característica facilita su comunicación en determinados contextos, pero presenta limitaciones cuando la variable contiene valores iguales o próximos a cero, debido a que el denominador puede generar valores indefinidos o porcentajes desproporcionadamente elevados. Hyndman y Koehler (2006) advierten sobre estas limitaciones de las medidas porcentuales de error, especialmente en series que contienen valores pequeños. En problemas de demanda con este tipo de observaciones, el MAPE debe interpretarse con precaución y no utilizarse de manera aislada como criterio de selección de un modelo.

Otra medida utilizada en problemas de regresión es el coeficiente de determinación (R^2), que expresa la proporción de variabilidad de la variable objetivo explicada por el modelo respecto de una referencia basada en la media. Valores más altos indican una mayor proporción de variabilidad explicada dentro del conjunto evaluado; sin embargo, un R^2 moderado o bajo no implica por sí solo que el modelo carezca de utilidad predictiva. Su

interpretación debe realizarse conjuntamente con las métricas de error y considerando la variabilidad propia del fenómeno y la granularidad de los datos.

Las métricas MAE, RMSE, MAPE y R^2 aportan, por tanto, perspectivas complementarias sobre el comportamiento de un modelo. Dos alternativas pueden presentar valores similares en una medida y diferencias mayores en otra, debido a que cada indicador responde de forma distinta ante errores extremos, valores pequeños o variabilidad de los datos. Para que estas diferencias sean comparables, los modelos deben evaluarse sobre el mismo conjunto de datos y bajo condiciones equivalentes.

En problemas de series temporales, la forma en que se construye el conjunto de evaluación también influye en la estimación del desempeño. Cerqueira et al. (2020) señalan que las dependencias temporales requieren precaución al aplicar procedimientos de validación y muestran que los métodos de evaluación que preservan el orden temporal de las observaciones resultan especialmente relevantes cuando las series presentan comportamientos no estacionarios. Esto refuerza la necesidad de evaluar los modelos sobre información posterior a la utilizada durante el entrenamiento y evitar esquemas que introduzcan conocimiento del futuro en etapas anteriores del modelado.

La selección de un modelo tampoco debe fundamentarse únicamente en la diferencia numérica mínima entre métricas. Cuando dos alternativas presentan resultados próximos, es necesario valorar si esa diferencia tiene relevancia práctica para el contexto en el que se utilizará la predicción. En aplicaciones operativas, una reducción pequeña del error estadístico puede no traducirse necesariamente en una diferencia apreciable en la toma de decisiones. También pueden intervenir aspectos como estabilidad, capacidad de generalización, costo computacional, interpretabilidad y facilidad de implementación.

La comparación amplia entre métodos de pronóstico confirma la importancia de la evaluación empírica. La competencia M4 analizó 100.000 series temporales y 61 métodos de forecasting, evidenciando diferencias de desempeño entre enfoques y tipos de series y mostrando que la selección de un método no puede establecerse únicamente a partir de su familia o nivel de complejidad (Makridakis et al., 2020).

Esta consideración adquiere especial importancia cuando el desempeño de los modelos varía entre unidades de análisis. Un algoritmo puede presentar mejores resultados para determinados productos o puntos de venta y un desempeño inferior en otros. Por ello, la identificación de las alternativas más adecuadas puede requerir analizar tanto resultados agregados como desagregados, evitando asumir la existencia de un modelo universalmente superior para todos los contextos.

También es importante diferenciar entre el desempeño estadístico del modelo y su utilidad para la operación. Un modelo con menor error proporciona una estimación más próxima a las observaciones dentro del conjunto evaluado, pero ello no demuestra automáticamente que producirá una reducción de producto no vendido, un incremento de ventas o una mejora en la eficiencia del proceso. Para establecer efectos de este tipo serían necesarias evaluaciones adicionales que relacionen las predicciones con decisiones y resultados observados en condiciones reales.

La interpretación operativa de un modelo predictivo implica analizar cómo sus estimaciones pueden apoyar una decisión, considerando su nivel de error, sus limitaciones y las condiciones en las que serán utilizadas. En el caso de la demanda de productos perecibles, las predicciones pueden servir como una referencia cuantitativa para la

planificación, pero no sustituyen otras fuentes de información ni el conocimiento de las personas responsables de la operación.

La evaluación de los modelos predictivos debe considerar tanto las métricas de error y ajuste como las características del problema analizado. MAE, RMSE, MAPE y R^2 permiten observar diferentes aspectos del desempeño, por lo que la selección del modelo no debería depender de una sola métrica. También es importante considerar la magnitud y relevancia de las diferencias entre los modelos, la estabilidad de su comportamiento y la finalidad de las predicciones. De esta manera, las métricas proporcionan una base para comparar los modelos, pero sus resultados deben interpretarse teniendo en cuenta las condiciones reales de la operación.

2.3 Terminología

POS (Point of Sale o Punto de venta): Sistema utilizado para registrar las transacciones comerciales realizadas en los puntos de venta. Contiene información asociada con cada operación, como producto vendido, cantidad de unidades, fecha, hora y local donde se efectuó la venta. Los registros generados por este sistema constituyen una fuente de información para el análisis de la demanda observada.

Demanda horaria observada: Cantidad de unidades de un producto vendidas durante un intervalo específico de una hora y registradas mediante las transacciones del sistema POS. Representa las ventas observadas dentro del período analizado y puede utilizarse como variable objetivo en un problema de predicción de demanda. No constituye necesariamente una medición completa de la demanda real, debido a que no incorpora

posibles requerimientos de clientes que no se concretaron en ventas por falta de disponibilidad del producto.

Producto perecible: Producto cuya calidad o aptitud comercial disminuye con el paso del tiempo y que requiere condiciones adecuadas de producción, manejo, almacenamiento o comercialización. En el contexto de los productos de panadería, esta característica incrementa la importancia de coordinar las cantidades disponibles con el comportamiento esperado de las ventas.

Merma: Pérdida de producto ocasionada por deterioro, vencimiento, manipulación, daño u otras causas operativas. En esta investigación, las devoluciones por no venta no se consideran equivalentes a la merma total; se utilizan únicamente como una referencia complementaria del producto registrado como no vendido al cierre de la jornada.

Devolución por no venta: Movimiento interno mediante el cual un producto registrado como no vendido al cierre de la jornada es trasladado desde un punto de venta hacia una bodega destinada al manejo de devoluciones. Este registro puede utilizarse como información analítica complementaria del producto no vendido, pero no constituye una medición directa de la cantidad horneada, la disponibilidad, la producción ni la merma total.

Horneo en local: Proceso mediante el cual determinados productos de panadería completan una etapa de preparación o cocción directamente en el punto de venta para atender la demanda esperada. La planificación de este proceso debe considerar el comportamiento esperado de las ventas, debido a que una cantidad insuficiente puede

afectar la disponibilidad, mientras que una cantidad superior a las ventas puede incrementar el producto no vendido.

Local (sucursal): Punto de venta físico perteneciente a la cadena de panaderías donde se realizan actividades de venta al consumidor final. Cada local puede presentar patrones particulares de ventas y demanda observada, asociados con las condiciones propias de su operación y su entorno comercial.

Rezago temporal (lag): Variable construida a partir de observaciones anteriores de una serie y utilizada para representar dependencias temporales. Cuando la secuencia no contiene observaciones para todos los intervalos, un número determinado de posiciones anteriores no necesariamente equivale al mismo intervalo de tiempo calendario.

lag_1d: Variable de rezago construida mediante el desplazamiento de 24 observaciones anteriores dentro de la serie de cada local-producto. Aporta información sobre el comportamiento histórico reciente de la demanda. Debido a que las horas sin venta no generan necesariamente un registro, las 24 posiciones anteriores no equivalen de forma estricta a un día calendario.

lag_7d: Variable de rezago construida mediante el desplazamiento de 168 observaciones anteriores dentro de la serie de cada local-producto. Permite incorporar información histórica de mayor profundidad; sin embargo, debido a la ausencia de registros en determinados intervalos sin ventas, las 168 posiciones no corresponden necesariamente a siete días calendario.

Media móvil (rolling mean): Estadística calculada sobre una ventana de observaciones anteriores que permite resumir el nivel reciente de una serie y reducir el efecto de fluctuaciones puntuales.

roll_mean_7d: Media móvil calculada sobre las 168 observaciones previas de cada serie local-producto, con un mínimo de 24 observaciones para su cálculo. Permite resumir el nivel reciente de la demanda y suavizar fluctuaciones puntuales. La denominación de la variable se conserva por corresponder a la implementación técnica, aunque la ventana no representa necesariamente siete días calendario.

Desviación estándar móvil (rolling standard deviation): Estadística calculada sobre una ventana histórica que representa la dispersión o variabilidad reciente de una serie.

roll_std_7d: Desviación estándar móvil calculada sobre las 168 observaciones previas de cada serie local-producto, con un mínimo de 24 observaciones para su cálculo. Permite representar la variabilidad reciente de la demanda. La denominación de la variable se conserva por corresponder a la implementación técnica, aunque la ventana no representa necesariamente siete días calendario.

Codificación cíclica (seno-coseno): Transformación aplicada a variables temporales con comportamiento periódico, como hora del día, día de la semana o mes del año. Mediante funciones seno y coseno se conserva la relación circular del tiempo, evitando que el modelo interprete como lejanas posiciones que realmente son consecutivas dentro de un ciclo.

Codificación de variables categóricas: Proceso mediante el cual categorías expresadas originalmente mediante identificadores o etiquetas se transforman en

representaciones numéricas que pueden ser procesadas por determinados algoritmos de aprendizaje automático.

local_enc: Representación numérica del identificador del local mediante codificación categórica. Permite conservar la identificación de los distintos puntos de venta dentro de los modelos construidos con información conjunta de múltiples locales.

producto_enc: Representación numérica del identificador del producto mediante codificación categórica. Permite conservar la diferenciación entre productos dentro de modelos construidos con información consolidada.

Modelo específico por local: Modelo predictivo entrenado utilizando información correspondiente a un punto de venta determinado, con el propósito de representar patrones particulares de su comportamiento histórico.

Modelo global: Modelo predictivo entrenado mediante información procedente de múltiples locales y productos. Este enfoque busca aprovechar patrones compartidos entre las distintas unidades de análisis y requiere mecanismos que permitan conservar su identificación dentro del conjunto de datos.

Ridge Regression: Modelo de regresión lineal con regularización L_2 que incorpora una penalización sobre los coeficientes para limitar su magnitud y reducir problemas asociados con la multicolinealidad y la complejidad del modelo.

Random Forest: Algoritmo de aprendizaje automático basado en la combinación de múltiples árboles de decisión construidos sobre diferentes muestras y subconjuntos de variables. La agregación de sus resultados permite representar relaciones no lineales e interacciones entre características y reducir la variabilidad asociada con un árbol individual.

XGBoost: Algoritmo de *gradient boosting* que construye árboles de decisión de manera secuencial, de modo que cada nueva etapa busca reducir los errores generados por las anteriores. Incorpora mecanismos de regularización y optimizaciones computacionales orientadas al procesamiento eficiente de datos estructurados.

LightGBM: Algoritmo de *gradient boosting* basado en árboles de decisión que incorpora técnicas optimizadas de construcción y procesamiento orientadas a mejorar la eficiencia computacional en conjuntos de datos de gran tamaño.

SVR (Support Vector Regression): Método de regresión basado en máquinas de vectores de soporte que busca encontrar una función capaz de representar la relación entre las variables dentro de un margen de tolerancia. Puede utilizar funciones *kernel* para capturar relaciones no lineales entre las variables de entrada y la variable objetivo.

Holt-Winters: Método estadístico de pronóstico basado en suavizamiento exponencial que permite representar componentes de nivel, tendencia y estacionalidad dentro de una serie temporal.

MAE (Mean Absolute Error, error absoluto medio): Métrica que calcula el promedio de las diferencias absolutas entre los valores observados y las predicciones. Se expresa en las mismas unidades de la variable objetivo, lo que permite interpretar directamente la magnitud promedio del error.

RMSE (Root Mean Squared Error, raíz del error cuadrático medio): Métrica que calcula la raíz cuadrada del promedio de los errores elevados al cuadrado. Asigna mayor peso a las desviaciones elevadas y se expresa en las mismas unidades de la variable objetivo.

MAPE (*Mean Absolute Percentage Error*, error porcentual absoluto medio):

Métrica que expresa el error absoluto promedio en términos porcentuales respecto del valor observado. Su interpretación requiere precaución cuando existen valores iguales o cercanos a cero, debido a que puede producir valores indefinidos o porcentajes desproporcionadamente elevados.

R² (Coeficiente de determinación): Indicador estadístico que expresa la proporción de variabilidad de la variable objetivo explicada por el modelo respecto de una referencia basada en la media. Debe interpretarse de manera complementaria con las métricas de error y considerando la variabilidad propia del fenómeno analizado.

Ingeniería de características (Feature engineering): Proceso mediante el cual se construyen, transforman o seleccionan variables de entrada para representar de manera adecuada la información relevante para un modelo predictivo. Puede incluir variables temporales, rezagos históricos, estadísticas móviles, transformaciones cíclicas y codificaciones categóricas.

Partición temporal: Proceso de división de un conjunto de datos respetando el orden cronológico de las observaciones. Permite separar datos de entrenamiento, validación y prueba evitando utilizar información futura durante el aprendizaje de modelos que serán evaluados sobre períodos posteriores.

Capítulo 3

3. Desarrollo

3.1 Diseño metodológico de la investigación

La investigación se desarrolló bajo un enfoque cuantitativo, debido a que utiliza registros históricos de ventas y devoluciones expresados mediante variables numéricas y categóricas, sobre los cuales se aplican procedimientos de preparación, análisis y modelado predictivo. El estudio se orienta a identificar patrones en la información disponible y a evaluar, mediante métricas cuantitativas, el desempeño de distintos modelos para estimar la demanda horaria observada. El enfoque cuantitativo resulta pertinente cuando el análisis se sustenta en datos medibles y en procedimientos sistemáticos para contrastar los resultados obtenidos (Hernández-Sampieri & Mendoza Torres, 2018).

Por su finalidad, corresponde a una investigación aplicada, ya que parte de una problemática operativa concreta relacionada con la planificación del horneo en los puntos de venta de una cadena de panaderías. El propósito no es desarrollar un algoritmo nuevo, sino utilizar y comparar métodos existentes de aprendizaje automático y series temporales para generar estimaciones que puedan servir como referencia cuantitativa en este proceso de decisión.

El diseño es no experimental, puesto que las variables analizadas no fueron manipuladas de manera deliberada. La investigación utilizó información generada previamente por la operación habitual de los puntos de venta y observó el comportamiento registrado de las ventas y devoluciones dentro del período estudiado. Esta característica diferencia el trabajo de un diseño experimental, en el que existiría una intervención

deliberada sobre las variables o condiciones analizadas (Hernández-Sampieri & Mendoza Torres, 2018).

Por la dimensión temporal de los datos, el estudio presenta un carácter longitudinal y retrospectivo, debido a que analiza registros históricos correspondientes a distintos momentos comprendidos entre 2024 y 2026, sin realizar una intervención sobre la operación durante ese período. La secuencia temporal se conserva tanto en el análisis de los datos como en el proceso de modelado, especialmente en la construcción de variables históricas y en la separación cronológica de los conjuntos de entrenamiento, validación y prueba.

El alcance metodológico es predictivo, dado que el componente principal de la investigación consiste en construir y evaluar modelos destinados a estimar una variable cuantitativa futura a partir de información histórica. La variable objetivo corresponde a la demanda horaria observada, representada mediante la cantidad de unidades vendidas por producto, local y hora a partir de los registros del sistema POS. Esta definición se limita a las ventas efectivamente registradas y no supone que dichas ventas representen la totalidad de la demanda real, debido a que no se dispone de información horaria suficiente para identificar posibles requerimientos no atendidos por falta de disponibilidad.

La unidad de análisis para el modelado corresponde a cada observación definida por la combinación de fecha, hora, local y producto. El estudio se delimitó a diez locales de operación directa y quince productos pertenecientes a la categoría de panadería, seleccionados mediante los criterios descritos posteriormente en este capítulo. Las ventas POS constituyen la fuente principal del análisis predictivo, mientras que las devoluciones

de producto no vendido se mantienen como una fuente complementaria y no forman parte de la variable objetivo ni de las características utilizadas para generar las predicciones.

Esta clasificación metodológica establece el marco dentro del cual se desarrollaron las etapas posteriores de preparación de datos, análisis exploratorio, construcción de variables, entrenamiento y evaluación de los modelos. El proceso se realizó sobre información histórica observacional y sus resultados se interpretan dentro del período, los locales, los productos y las condiciones de información disponibles para la investigación.

3.2 Desarrollo del Trabajo

El desarrollo de la investigación se organizó en cinco fases sucesivas que abarcaron desde la revisión inicial de los datos hasta la construcción y evaluación de los modelos predictivos. Esta organización permitió mantener la trazabilidad entre las fuentes originales, las transformaciones aplicadas y los conjuntos de información utilizados en las diferentes etapas del análisis.

La primera fase comprendió la carga, consolidación y perfilado inicial de las fuentes de ventas POS y devoluciones de producto no vendido. En esta etapa se revisaron la estructura de los archivos, los tipos de variables, los períodos disponibles, la presencia de valores nulos y posibles registros duplicados. El propósito fue conocer las condiciones iniciales de la información antes de aplicar procedimientos de depuración.

La segunda fase correspondió a la limpieza y preparación de los datos. Se normalizaron los tipos de variables, se validaron los registros necesarios para el análisis y se generaron atributos temporales derivados de las fechas y horas disponibles. Las ventas y

devoluciones se mantuvieron como conjuntos independientes debido a que presentan diferente granularidad y cumplen funciones distintas dentro de la investigación.

La tercera fase estuvo orientada al análisis exploratorio. Las ventas POS se analizaron como fuente principal para estudiar la demanda observada, considerando su comportamiento temporal y las diferencias existentes entre locales y productos. Las devoluciones se examinaron de manera complementaria con el propósito de caracterizar el producto registrado como no vendido, sin incorporarlas como variable objetivo ni como característica predictora del modelo.

La cuarta fase consistió en delimitar el conjunto de información destinado al modelado. Para ello se aplicaron criterios de selección sobre los locales y productos disponibles, de acuerdo con su participación dentro de la operación y con el alcance definido para la investigación. A partir de esta selección se construyó posteriormente la base horaria utilizada para representar la demanda observada por local y producto.

La quinta fase comprendió la ingeniería de características, la partición temporal de los datos, el entrenamiento de los modelos y la evaluación de su desempeño. En esta etapa se construyeron variables temporales, cíclicas, categóricas e históricas, se separó cronológicamente la información en conjuntos de entrenamiento, validación y prueba y se compararon modelos específicos por local con modelos globales. Las métricas obtenidas permitieron analizar el comportamiento de las diferentes alternativas bajo un mismo esquema de evaluación.

El procesamiento se implementó mediante notebooks de Jupyter organizados de forma secuencial. Las distintas etapas se distribuyeron entre los notebooks 01 al 07, de

acuerdo con la función asignada a cada uno. Esta estructura permitió separar las actividades de carga, limpieza, análisis exploratorio, selección y modelado y conservar una secuencia reproducible de las transformaciones efectuadas sobre los datos.

Figura 1

Flujo metodológico de la investigación



La Figura 1 presenta el flujo metodológico seguido en la investigación, desde las fuentes de datos y su preparación hasta el entrenamiento, evaluación y comparación de los

modelos, concluyendo con la generación de predicciones como apoyo a la planificación del horneado.

3.3 Fuente y características de los datos

3.3.1 Fuente de los datos y justificación de la elección

La investigación utilizó registros históricos transaccionales y operativos generados por el sistema ERP de una cadena de panaderías ubicada en Quito, Ecuador. La selección de estas fuentes respondió a su relación directa con el problema analizado, ya que contienen información sobre las ventas realizadas en los puntos de venta y sobre los movimientos asociados al producto registrado como no vendido al cierre de la jornada.

La fuente principal estuvo conformada por los registros de ventas facturadas en los puntos de venta, obtenidos a partir de la vista `v_ventas`. Para identificar las operaciones originadas en el sistema POS se utilizó el campo `fecha_creacion_pos`, que contiene la fecha y hora en la que se registró la transacción en el punto de venta. Este criterio permitió diferenciar las ventas POS de otros documentos generados mediante procesos distintos de la organización y conservar únicamente las transacciones correspondientes al alcance de la investigación.

Los registros de ventas fueron seleccionados como fuente principal debido a que permiten identificar el producto vendido, la cantidad de unidades, el local y el momento de la transacción. A partir de esta información fue posible construir la variable objetivo utilizada en el modelado, definida como la cantidad de unidades vendidas por producto, local y hora. En este trabajo, esta variable representa la demanda horaria observada,

entendida como las ventas efectivamente registradas en el POS y no como una medición completa de toda la demanda potencial.

La segunda fuente correspondió a los movimientos de devolución de producto no vendido, obtenidos desde la vista `v_devoluciones`. Para delimitar estos registros se consideraron los movimientos con destino a la bodega DEVOLUCION, tipo documental “Devolución por NO VENTA” y estado CONTABILIZADO. Esta información se mantuvo separada de las ventas debido a que cumple una función distinta dentro del análisis y presenta una granularidad temporal diferente.

Las devoluciones no corresponden a devoluciones realizadas por clientes ni constituyen una medición directa de la cantidad producida, horneada o disponible durante la jornada. Su utilización se limitó a caracterizar el producto registrado como no vendido y a complementar la interpretación de los patrones identificados en las ventas. Por esta razón, no se incorporaron como variable objetivo ni como característica predictora de los modelos.

Ambas fuentes incluyeron atributos descriptivos relacionados con productos, locales y bodegas, como identificadores, códigos, nombres, categorías y subcategorías. Estos campos permitieron mantener la trazabilidad de los registros y desarrollar análisis diferenciados por producto y punto de venta.

La utilización de información histórica proveniente de la operación real permitió trabajar con un volumen amplio de observaciones. No obstante, estas fuentes presentan condiciones de cobertura y granularidad que deben considerarse al interpretar los resultados y que se analizan específicamente en la sección 3.3.4.

3.3.2 Descripción y estructura del conjunto de datos

El conjunto de datos utilizado en la investigación se organizó en dos capas principales: ventas POS y devoluciones de producto no vendido. Ambas fuentes presentan una estructura tabular, pero difieren en su granularidad temporal y en la función que cumplen dentro del análisis.

La capa de ventas POS constituyó la fuente principal para el modelado predictivo. Cada registro correspondió a una transacción realizada en un punto de venta y contenía información relacionada con el local, el producto, la fecha y hora de la operación, la cantidad vendida y los valores comerciales asociados. A partir de estos registros se construyó posteriormente una base agregada a nivel de fecha, hora, local y producto, utilizada para representar la demanda horaria observada.

Las variables de la capa de ventas pueden agruparse en cuatro categorías: organizacionales, temporales, de producto y comerciales. Las variables organizacionales permiten identificar la sucursal y la bodega asociadas a la operación; las temporales registran el momento en que se produjo la venta; las variables de producto contienen los identificadores y clasificaciones comerciales; y las variables comerciales incluyen la cantidad vendida y los valores económicos de la transacción.

La Tabla 1 presenta las principales variables utilizadas en la capa de ventas POS.

Tabla 1

Principales variables de la capa de ventas POS

Grupo	Variables principales
Organizacionales	Identificador de organización, sucursal, nombre del local y bodega

Temporales	Fecha del documento, fecha y hora de creación POS, hora, día de la semana, mes, año y franja horaria
Producto	Identificador, código, nombre, categoría y subcategoría
Comerciales	Cantidad vendida, precio, venta bruta, descuento y venta neta

La Tabla 1 muestra que la información disponible permite identificar el producto vendido, el local y el momento de cada transacción, además de cuantificar las unidades y los valores comerciales asociados. Las variables temporales son especialmente relevantes, ya que constituyen la base para construir la demanda horaria observada utilizada posteriormente en el modelado. La segunda capa estuvo conformada por los movimientos de devolución de producto no vendido. A diferencia de las ventas POS, estos registros corresponden a movimientos operativos asociados principalmente al cierre de la jornada y no contienen una marca temporal equivalente que permita reconstruir con precisión el comportamiento del producto a nivel horario. Por esta razón, su utilización se mantuvo separada de la construcción de la variable objetivo.

Las variables disponibles en la capa de devoluciones pueden agruparse en componentes temporales, documentales, organizacionales, de producto y operativos. La Tabla 2 resume las principales variables consideradas.

Tabla 2

Principales variables de la capa de devoluciones

Grupo	Variables principales
Temporales	Fecha, año, mes y día de la semana

Documentales	Número de documento, tipo documental, estado y usuario creador
Organizacionales	Sucursal, bodega de origen y bodega de destino
Producto	Identificador, código, nombre, categoría y subcategoría
Operativas	Cantidad devuelta y precio unitario

La Tabla 2 permite identificar la información disponible sobre el producto registrado como no vendido, el local asociado, las bodegas involucradas y la cantidad devuelta. A diferencia de las ventas POS, estos registros no permiten determinar el momento exacto en que el producto fue horneado, estuvo disponible para la venta o dejó de estarlo; por esta razón, se utilizaron únicamente como información complementaria para contextualizar los patrones observados en las ventas.

Después de las etapas de carga, perfilado y limpieza, las ventas y devoluciones se mantuvieron en conjuntos procesados independientes, debido a que cumplen funciones diferentes dentro del análisis. La capa de ventas constituyó la base para construir la demanda horaria observada y posteriormente fue delimitada a los locales y productos seleccionados para el modelado, mientras que las devoluciones se conservaron como fuente complementaria para los análisis descriptivos, sin incorporarse como variable objetivo ni como predictor de los modelos.

3.3.3 Volumen y período de información

La extracción inicial reunió información correspondiente al período comprendido entre 2024 y 2026. En total se consolidaron 2.910.313 registros de ventas POS y 116.710 registros de devoluciones de producto no vendido. La información correspondiente a 2026

comprende únicamente los datos disponibles hasta la fecha de corte utilizada en la investigación, por lo que este año no representa un período anual completo.

La Tabla 3 presenta la distribución de los registros iniciales por fuente y año, antes de aplicar los procedimientos de limpieza, exclusión y selección definidos posteriormente.

Tabla 3

Volumen inicial de registros por fuente y año

Año	Ventas POS	Devoluciones
2024	800.702	63.294
2025	1.789.787	45.847
2026 ¹	319.824	7.569
Total	2.910.313	116.710

Nota. ¹El año 2026 corresponde a un período parcial hasta la fecha de corte utilizada en el proyecto.

Los valores de la Tabla 3 muestran que el mayor volumen de registros de ventas corresponde a 2025, mientras que la menor cantidad observada en 2026 se explica por el carácter parcial de este período. En las devoluciones, 2024 concentra el mayor número de registros. Estas diferencias reflejan el volumen de observaciones disponible en cada fuente y no deben interpretarse directamente como variaciones equivalentes en la cantidad vendida o devuelta.

Los valores presentados corresponden a la extracción consolidada inicial. Los registros utilizados posteriormente en el análisis exploratorio y en el modelado predictivo se obtuvieron después de aplicar los procedimientos de limpieza, validación, exclusión de

locales fuera del alcance y selección de los 10 locales y 15 productos definidos para la investigación.

3.3.4 Consideraciones de calidad, alcance y posibles sesgos de los datos

La utilización de registros históricos provenientes de la operación real permitió trabajar con un volumen amplio de información. No obstante, al tratarse de datos generados por sistemas operativos y no específicamente para fines de investigación, su interpretación requiere considerar ciertas condiciones relacionadas con su cobertura y granularidad.

Las ventas registradas en el sistema POS representan únicamente las transacciones efectivamente realizadas. Por esta razón, la variable objetivo se definió como demanda horaria observada y no como demanda total o potencial. La ausencia de ventas en una determinada hora no implica necesariamente una demanda igual a cero, ya que no se dispone de información horaria suficiente para determinar si el producto estuvo disponible, se agotó previamente o existieron requerimientos que no pudieron ser atendidos.

De igual manera, no se cuenta con un registro horario completo de los eventos de horneo ni de la cantidad exacta de producto disponible para la venta en cada momento. Las devoluciones presentan una granularidad diferente, pues corresponden al producto registrado como no vendido y se generan principalmente al cierre de la jornada. Por ello, no permiten determinar cuándo fue horneado el producto o durante qué período permaneció disponible, y se utilizaron únicamente como información complementaria, sin incorporarse como variable objetivo ni como predictor de los modelos.

El período correspondiente a 2026 es parcial, debido a que incluye únicamente la información disponible hasta la fecha de corte definida para el estudio. Asimismo, el

alcance final se delimitó a diez locales de operación directa y quince productos de la categoría de panadería, por lo que los resultados obtenidos corresponden a las condiciones, locales, productos y período analizados y no deben generalizarse automáticamente a otros contextos.

Finalmente, las fuentes disponibles no incorporan de manera estructurada todos los factores que podrían influir en el comportamiento de las ventas, como promociones, condiciones climáticas, eventos especiales o cambios operativos. Las predicciones se encuentran limitadas a los patrones que pueden ser identificados a partir de las variables disponibles en el conjunto de datos.

3.3.5 Procedimiento de extracción y consolidación de los datos

La extracción de la información se realizó a partir de las fuentes definidas para ventas POS y devoluciones de producto no vendido. Debido al volumen de registros disponible, el proceso se efectuó por bloques temporales con el propósito de facilitar la ejecución de las consultas y reducir el riesgo de interrupciones durante la obtención de los datos.

En el caso de las ventas POS, la información se extrajo en archivos correspondientes a períodos anuales y semestrales, mientras que las devoluciones se obtuvieron mediante archivos anuales. Esta división respondió únicamente a una necesidad técnica de procesamiento y no modificó el período considerado ni la estructura de los registros analizados.

Una vez obtenidos los archivos de cada período, se verificó que conservaran una estructura compatible y posteriormente se consolidaron por fuente. De esta manera, los

archivos de ventas se integraron en un único conjunto y se aplicó el mismo procedimiento para las devoluciones. La consolidación permitió disponer de bases continuas para las etapas posteriores de perfilado, limpieza, validación y análisis.

Como parte del proceso se conservaron los campos necesarios para identificar la fecha, el local, el producto, la cantidad y los demás atributos requeridos para las etapas posteriores del análisis. Las ventas y devoluciones se consolidaron de manera independiente, conservando la estructura correspondiente a cada fuente.

El procedimiento de extracción y consolidación permitió construir una base de trabajo trazable desde los registros originales hasta los conjuntos procesados utilizados en las etapas posteriores. A partir de estas fuentes consolidadas se aplicaron los procedimientos de limpieza, transformación y selección descritos en la siguiente sección.

3.4 Preparación y procesamiento de datos

3.4.1 Limpieza y validación de los datos

La limpieza y validación de la información se realizó mediante el notebook 02_limpieza.ipynb, a partir de los conjuntos consolidados en la etapa de carga y perfilado inicial. El objetivo de este proceso fue asegurar que las variables necesarias para las etapas posteriores presentaran una estructura consistente y tipos de datos adecuados para el análisis.

Como parte de la preparación se normalizaron los tipos de las variables temporales y numéricas, con especial atención a los campos de fecha, hora, cantidad y valores comerciales. Esta transformación permitió trabajar posteriormente con la información en

formatos homogéneos y facilitó la generación de variables derivadas utilizadas en el análisis temporal.

El tratamiento de los valores faltantes se definió de acuerdo con su proporción y con la naturaleza de cada variable. El procedimiento contempló la eliminación de registros cuando la proporción de valores nulos era inferior al 1 %, mientras que para porcentajes intermedios se consideró la imputación mediante la mediana en variables numéricas y la moda en variables categóricas. Los campos con una proporción elevada de datos faltantes debían ser revisados antes de aplicar una sustitución automática. En los conjuntos utilizados durante la etapa de limpieza, los campos que presentaron valores faltantes registraron proporciones inferiores al 1 %, por lo que no fue necesario aplicar procedimientos de imputación.

También se verificó la existencia de registros completamente duplicados en las capas de ventas y devoluciones. Esta revisión no identificó duplicados exactos, por lo que no fue necesario eliminar observaciones por este criterio. De manera complementaria, se revisaron la estructura de las variables, los rangos de cantidades y valores comerciales y la integridad de los conjuntos resultantes.

En la capa de ventas se mantuvo `fecha_creacion_pos` como referencia temporal de las transacciones realizadas en los puntos de venta. En las devoluciones se conservaron los movimientos previamente delimitados como producto no vendido, de acuerdo con los criterios de negocio establecidos para esta fuente. Las transformaciones temporales derivadas se desarrollaron dentro del mismo proceso de preparación, aunque su construcción y utilización se describen de manera específica en la siguiente sección.

Como resultado de la limpieza se obtuvieron 2.905.211 registros de ventas y 111.920 registros de devoluciones. Los conjuntos procesados fueron almacenados en los archivos `ventas_limpio.parquet` y `devoluciones_limpio.parquet`, que constituyeron la base para las etapas posteriores de análisis exploratorio, selección y modelado.

3.4.2 Transformación y construcción de variables temporales

A partir de la marca temporal asociada a las transacciones de ventas POS se generaron variables derivadas destinadas a representar el comportamiento de las ventas en diferentes escalas de tiempo. Estas transformaciones permitieron descomponer la fecha y hora original en atributos que posteriormente fueron utilizados en el análisis exploratorio y en la preparación del conjunto de datos para el modelado.

Entre las variables generadas se incluyeron la hora, el minuto, el día del mes, el mes, el año, el día de la semana, la semana del año, el indicador de fin de semana, la fecha sin componente horario, el período año-mes y la franja horaria. Estas variables permitieron analizar los registros desde diferentes niveles temporales y facilitaron la identificación de patrones intradiarios, semanales y mensuales.

Para analizar el comportamiento de las ventas dentro de la jornada, las horas fueron agrupadas en cinco franjas horarias. La Tabla 4 presenta los intervalos definidos durante el procesamiento.

Tabla 4

Franjas horarias utilizadas en el análisis

Franja	Intervalo
Mañana	06:00–09:59

Mediodía	10:00–13:59
Tarde	14:00–17:59
Noche	18:00–21:59
Madrugada	22:00–05:59

La Tabla 4 resume la agrupación empleada para representar distintos momentos de la jornada. La categoría madrugada permitió clasificar los registros ubicados entre las 22:00 y las 05:59, mientras que las demás franjas organizaron la actividad registrada desde la mañana hasta la noche. Esta clasificación se utilizó principalmente con fines descriptivos para comparar el comportamiento de las ventas a lo largo del día.

Las variables temporales obtenidas en esta etapa constituyeron la base para los análisis posteriores. Algunas de ellas fueron empleadas directamente en el análisis exploratorio, mientras que otras sirvieron como insumo para construir las características predictoras utilizadas posteriormente en los modelos. Las transformaciones específicas destinadas al modelado, como las codificaciones cíclicas y las variables históricas, se describen en la sección correspondiente a la ingeniería de características.

3.4.3 Tratamiento analítico de las devoluciones

Las devoluciones de producto no vendido se analizaron como una capa complementaria a las ventas POS. Debido a que estos registros se generan principalmente al cierre de la jornada y no presentan la misma granularidad horaria que las transacciones de venta, no fueron integrados como variable predictora ni utilizados para construir la demanda horaria observada. Su finalidad fue aportar información adicional sobre el producto que permaneció sin vender dentro del período analizado.

Para facilitar la comparación entre ventas y devoluciones, ambas fuentes se agregaron a niveles compatibles de local y producto y, cuando correspondía, por fecha. Los identificadores utilizados para los locales representan las diferentes ubicaciones o sectores considerados dentro del alcance del estudio y permiten diferenciar su comportamiento operativo.

A partir de las cantidades vendidas y devueltas se construyó un total de referencia, definido mediante la siguiente expresión:

$$\text{Total de referencia} = \text{Cantidad vendida} + \text{Cantidad devuelta}$$

Sobre esta base se calculó la tasa de devolución:

$$\text{Tasa de devolución (\%)} = [\text{Cantidad devuelta} / (\text{Cantidad vendida} + \text{Cantidad devuelta})] \times 100$$

Aunque dentro del procesamiento computacional esta suma fue utilizada como aproximación para relacionar ventas y devoluciones, en la investigación se interpreta únicamente como un total de referencia vendido-devuelto. Esto se debe a que la información disponible no permite conocer de forma completa todos los eventos de horneado ni la disponibilidad exacta del producto durante la jornada. Por tanto, esta magnitud no representa una medición directa de la cantidad producida o efectivamente horneada.

La tasa de devolución permitió comparar proporcionalmente el producto registrado como no vendido entre los distintos locales y productos, evitando limitar el análisis a cantidades absolutas. Este indicador se utilizó para complementar la interpretación de los patrones de ventas, sin asumir que representa de manera directa la merma total o la eficiencia operativa de cada local.

Las devoluciones se mantuvieron fuera del conjunto de características utilizado para entrenar los modelos predictivos. Esta decisión respondió tanto a su función complementaria como a su menor resolución temporal, ya que incorporar información registrada al cierre de la jornada en una predicción horaria podría introducir datos que no estarían disponibles en el momento de realizar la estimación. Por esta razón, las devoluciones se utilizaron principalmente en los análisis descriptivos y como referencia para contextualizar el comportamiento de las ventas.

3.4.4 Selección de locales y productos

La delimitación del conjunto destinado al modelado se realizó mediante el notebook 06_analisis_top_locales_productos.ipynb. Esta etapa tuvo como propósito concentrar el análisis predictivo en locales con suficiente volumen y continuidad de información y en productos de la categoría Panadería que presentaran condiciones adecuadas para el entrenamiento y evaluación de los modelos.

La selección de locales se desarrolló en dos etapas. En la primera se aplicaron criterios de alcance y continuidad operativa para retirar del análisis aquellos puntos que no podían considerarse comparables con los locales incluidos en la investigación. Planta fue excluida por corresponder al centro de producción y no a un punto de venta. Santa Lucía y La Viña se retiraron por condiciones relacionadas con disponibilidad, continuidad de información o alcance operativo. Puembo y Bosque dejaron de considerarse debido a cambios en su esquema de operación que afectaron la continuidad requerida para el estudio.

Una vez aplicadas estas exclusiones, los locales restantes fueron incorporados al proceso de evaluación mediante una puntuación compuesta. En esta segunda etapa se

analizaron indicadores relacionados con cantidad total vendida, valor de ventas, número de transacciones, promedio diario de unidades vendidas y un indicador complementario asociado a la relación entre ventas y devoluciones. Para obtener la puntuación final se asignó una ponderación del 30 % a la cantidad vendida, 25 % al valor de ventas y 15 % a cada uno de los tres indicadores restantes.

El ranking resultante permitió ordenar los locales que cumplían las condiciones mínimas para continuar en el proceso de selección. A partir de este ordenamiento se escogieron los diez locales incluidos finalmente en el modelado: Cumbayá, Coruña, Edén, Primavera, Brasil, Marina, San Juan, Hilacril, Gaspar y Portugal. Los locales Monteserrín, Pinar y Madrid participaron en el ranking, pero no quedaron dentro de los diez seleccionados de acuerdo con la puntuación obtenida.

Esta distinción permite separar las exclusiones previas por alcance o continuidad de los locales que sí fueron evaluados en el ranking, pero no alcanzaron una posición dentro del grupo final seleccionado. Las figuras exploratorias conservan algunos de estos locales porque fueron generadas durante la fase previa a la delimitación definitiva del conjunto de modelado.

Para la selección de productos se consideraron únicamente los registros pertenecientes a la categoría Panadería. El criterio principal fue la cantidad total vendida durante el período analizado. Como información complementaria se revisaron el valor de ventas, el número de transacciones, la presencia del producto en los locales seleccionados, el precio mediano, la cantidad devuelta, la tasa de devolución, el coeficiente de variación y el número de días con ventas.

Como resultado, se seleccionaron los siguientes quince productos: Enrollado, Croissant, Cacho, Pan reventado, Rosa de agua, Croissant integral, Empanada de queso, Enrollado integral, Pan de maíz, Rosa especial, Croissant de chocolate 2022, Pan suizo, Baguette, Baguette para sanduche y MB cacho integral. La selección de diez locales y quince productos definió un universo potencial de 150 combinaciones local-producto para las etapas posteriores de preparación y modelado. El conjunto transaccional resultante quedó conformado por 959.692 registros de ventas.

La selección permitió delimitar el problema a un conjunto con suficiente volumen y continuidad de información para la comparación de modelos, manteniendo diversidad entre locales y productos sin extender el análisis a toda la operación disponible. Esta delimitación debe considerarse al interpretar los resultados, ya que las conclusiones del estudio corresponden específicamente a los locales, productos y período incluidos en el modelado.

3.4.5 Construcción del conjunto de datos horario

A partir del subconjunto de 959.692 registros transaccionales correspondiente a los diez locales y quince productos seleccionados, se construyó el conjunto destinado al modelado predictivo. Debido a que la unidad de análisis definida en la investigación corresponde a la demanda horaria observada, las transacciones se agruparon por fecha, hora, local y producto, sumando la cantidad vendida dentro de cada combinación. De esta forma, cada registro de la nueva base representa las unidades vendidas de un producto en un local durante una hora específica.

Una vez realizada la agregación, los registros se ordenaron cronológicamente dentro de cada combinación local-producto. Este ordenamiento permitió conservar la secuencia

temporal necesaria para construir posteriormente las variables históricas utilizadas por los modelos. Las series resultantes presentaron diferencias en frecuencia y continuidad, debido al comportamiento particular de las ventas de cada producto en cada local.

En la construcción de la base se conservaron únicamente las horas en las que existieron transacciones registradas. Las horas sin ventas no se incorporaron automáticamente como observaciones con cantidad igual a cero. Esta decisión responde a que la ausencia de una transacción puede representar diferentes situaciones: una demanda efectivamente nula, falta temporal de disponibilidad del producto, ausencia de horneado o condiciones particulares de operación del local. Dado que la información disponible no permite distinguir de manera horaria entre estos escenarios, la imputación generalizada de ceros podría introducir valores que no necesariamente representan el comportamiento real de la demanda observada.

Esta característica afecta directamente la interpretación de las variables históricas. Los rezagos de 24 y 168 posiciones, utilizados posteriormente en el modelado, corresponden a observaciones anteriores dentro de la secuencia disponible para cada combinación local-producto y no necesariamente a 24 horas o siete días calendario exactos. Del mismo modo, las estadísticas móviles se calcularon sobre las observaciones históricas disponibles y no sobre una cuadrícula horaria continua.

La discontinuidad de las series fue relevante dentro del conjunto analizado. Para cada combinación día-local-producto se registró una mediana de seis horas con ventas dentro de una jornada de referencia de 17 horas, con una media de 6,2 horas, un mínimo de una y un máximo de quince horas con transacciones. Estos valores muestran que la

actividad de venta de cada producto no se distribuye de forma continua durante toda la jornada.

Durante el diseño metodológico se consideraron dos posibilidades para tratar las horas sin transacción: completar una cuadrícula horaria asignando cero a las horas faltantes o conservar únicamente las horas efectivamente observadas. Se adoptó la segunda alternativa debido a que, en ausencia de información sobre disponibilidad horaria del producto, completar los registros con ceros podía generar una subestimación de la demanda en aquellos casos donde la falta de venta hubiera sido consecuencia de falta de producto y no de ausencia de demanda.

Posteriormente se generaron las variables predictoras descritas en la siguiente sección y se eliminaron las observaciones iniciales que no disponían de antecedentes suficientes para calcular los rezagos y las estadísticas móviles. Como resultado de este proceso, el conjunto final utilizado para el modelado quedó conformado por 498.750 observaciones horarias, distribuidas posteriormente entre los conjuntos de entrenamiento, validación y prueba.

Al interpretar las predicciones, es importante considerar que el modelo estima el nivel esperado de ventas a partir de los patrones históricos observados, pero no representa directamente la disponibilidad real del producto ni el proceso de horneado realizado durante cada hora.

3.4.6 Ingeniería de características para el modelado

A partir del conjunto de datos horario se generaron las variables utilizadas como entrada para los modelos predictivos. El proceso de ingeniería de características se aplicó

de manera consistente tanto en el notebook de modelado como en el módulo independiente de entrenamiento, con el propósito de mantener las mismas transformaciones durante las etapas de análisis, entrenamiento e inferencia.

Las variables predictoras se organizaron en cuatro grupos: temporales, cíclicas, categóricas e históricas. En total se utilizaron 18 características, mientras que la variable objetivo correspondió a la cantidad de unidades vendidas por hora para cada combinación de local y producto.

Las variables temporales incluyeron la hora del día, el día de la semana, el mes, el día del mes, la semana del año y un indicador para diferenciar los fines de semana. Estas características permitieron representar la posición de cada observación dentro del calendario y capturar posibles diferencias asociadas con distintos momentos de la jornada, la semana y el año.

Debido a que algunas variables temporales presentan un comportamiento periódico, la hora, el día de la semana y el mes se transformaron adicionalmente mediante funciones seno y coseno. Esta codificación cíclica permite representar la proximidad entre valores ubicados al inicio y al final de un mismo ciclo, evitando tratarlos como puntos completamente separados. Por ejemplo, las últimas y primeras horas del día mantienen una relación de proximidad dentro de la representación cíclica. La implementación genera seis variables de este tipo: `hora_sin`, `hora_cos`, `día_semana_sin`, `día_semana_cos`, `mes_sin` y `mes_cos`.

Los identificadores de local y producto se transformaron en variables numéricas mediante codificación categórica, generando `local_enc` y `producto_enc`. Esta

transformación permitió incorporar la identificación de las distintas series dentro de los algoritmos que requieren entradas numéricas. En los modelos desarrollados específicamente para cada local, el identificador del local permanece constante, mientras que en los modelos globales permite distinguir las observaciones correspondientes a las diferentes ubicaciones analizadas.

También se construyeron cuatro variables históricas destinadas a representar el comportamiento reciente de cada serie local-producto. Para ello, los registros se ordenaron previamente por local, producto, fecha y hora. Se generaron dos rezagos, correspondientes a 24 y 168 observaciones anteriores, así como una media móvil y una desviación estándar calculadas sobre una ventana de hasta 168 observaciones previas.

La Tabla 5 resume las variables utilizadas durante el proceso de modelado

Tabla 5

Variables predictoras utilizadas en los modelos

Grupo	Variables	Interpretación metodológica	Finalidad
Temporales	hora, día_semana, mes, día_del_mes, semana_anio, es_fin_semana	Posición de la observación dentro del calendario operativo	Representar patrones horarios y calendáricos
Cíclicas	hora_sin, hora_cos, día_semana_sin, día_semana_cos, mes_sin, mes_cos	Proyección circular del tiempo para evitar discontinuidades artificiales	Conservar la naturaleza periódica del tiempo

Catagóricas	local_enc, producto_enc	Identificación numérica de la serie local-producto	Incorporar heterogeneidad entre locales y productos
Históricas	lag_1d, lag_7d, roll_mean_7d, roll_std_7d	Memoria de nivel y variabilidad de la demanda reciente	Representar el nivel y la variabilidad reciente de la demanda

La Tabla 5 resume las características utilizadas en el modelado, combinando información temporal, representaciones cíclicas, identificación de las series y antecedentes históricos de la demanda. Esta combinación permitió que los modelos utilizaran tanto las condiciones temporales de cada observación como información sobre el comportamiento reciente del producto en cada local.

Las variables históricas conservaron las denominaciones utilizadas en la implementación, aunque su interpretación debe realizarse considerando la estructura discontinua del conjunto horario explicada en la sección anterior. En particular, lag_1d corresponde al valor de demanda ubicado 24 observaciones atrás y lag_7d al valor ubicado 168 observaciones atrás dentro de cada serie local-producto; por tanto, estos desplazamientos no representan necesariamente un día o siete días calendario exactos.

Por su parte, roll_mean_7d representa la media de las observaciones históricas contenidas en una ventana máxima de 168 posiciones, mientras que roll_std_7d representa su dispersión. La implementación exige un mínimo de 24 observaciones para calcular estas estadísticas. Antes de obtener ambas medidas, la serie se desplaza una posición, evitando que el valor correspondiente a la observación que se pretende predecir forme parte del cálculo de sus propias características.

Una vez generadas las características, se eliminaron las observaciones iniciales que no disponían de información histórica suficiente para calcular los rezagos y estadísticas móviles. El conjunto resultante, conformado por registros con las variables requeridas para el modelado, fue utilizado posteriormente para realizar la partición temporal en entrenamiento, validación y prueba.

3.4.7 Partición temporal de los datos

Con el propósito de preservar la secuencia cronológica de la información y evitar que observaciones futuras influyeran en el entrenamiento, el conjunto final de modelado no se dividió de forma aleatoria. Se aplicó una partición temporal en tres subconjuntos: entrenamiento, validación y prueba. Esta estrategia reproduce de manera más cercana el escenario de predicción, en el cual los modelos se construyen utilizando información histórica y posteriormente se aplican sobre períodos futuros no utilizados durante su ajuste.

La definición de los cortes temporales buscó mantener una secuencia cronológica continua y asignar a cada subconjunto una función específica dentro del proceso de modelado. El período de entrenamiento concentró la mayor parte de la información histórica disponible, de manera que los modelos dispusieran de suficiente información para aprender los patrones de demanda observados. El período comprendido entre octubre y diciembre de 2025 se reservó para validación, permitiendo revisar el comportamiento de los modelos sobre información posterior al entrenamiento y apoyar determinadas decisiones de configuración. Finalmente, el período comprendido entre enero y el 14 de marzo de 2026 se mantuvo como conjunto de prueba, con el propósito de evaluar el desempeño final sobre observaciones cronológicamente posteriores y no utilizadas durante el ajuste. Esta

organización permitió conservar la temporalidad de los datos y separar claramente las etapas de aprendizaje, revisión y evaluación final.

El conjunto de entrenamiento estuvo conformado por las observaciones comprendidas entre el 15 de julio de 2024 y el 30 de septiembre de 2025, con un total de 358.558 registros, equivalentes al 71,9 % de la base final. El conjunto de validación comprendió el período entre el 1 de octubre y el 31 de diciembre de 2025, con 78.122 observaciones, correspondientes al 15,7 %. El conjunto de prueba incluyó las observaciones registradas entre el 1 de enero y el 14 de marzo de 2026, con 62.070 registros, equivalentes al 12,4 % del conjunto de modelado.

La Tabla 6 presenta la distribución temporal del conjunto utilizado para el modelado.

Tabla 6

Distribución temporal del conjunto de modelado

Partición	Período	Observaciones	Porcentaje	Finalidad
Entrenamiento	15 de julio de 2024–30 de septiembre de 2025	358.558	71,9 %	Ajuste de los modelos
Validación	1 de octubre–31 de diciembre de 2025	78.122	15,7 %	Revisión de desempeño y apoyo al ajuste
Prueba	1 de enero–14 de marzo de 2026	62.070	12,4 %	Evaluación final

Total	15 de julio de 2024–14 de marzo de 2026	498.750	100 %	—
-------	---	---------	-------	---

La distribución presentada en la Tabla 6 refleja que la mayor proporción de observaciones se destinó al entrenamiento, mientras que los períodos posteriores se reservaron para validación y prueba. La conservación del orden cronológico permitió evaluar los modelos sobre información posterior a la utilizada durante su ajuste y evitar una mezcla aleatoria entre observaciones de diferentes momentos del período analizado.

La partición de validación no tuvo la misma función en todos los algoritmos. Ridge, Random Forest y SVR se entrenaron utilizando configuraciones previamente definidas, por lo que este subconjunto no intervino directamente en el ajuste de sus parámetros. LightGBM utilizó el conjunto de validación como referencia para revisar el desempeño durante el proceso de entrenamiento, mientras que XGBoost lo utilizó para aplicar el mecanismo de parada temprana (*early stopping*) y controlar el número efectivo de iteraciones. En todos los casos, el conjunto de prueba permaneció separado hasta la etapa de evaluación final.

Esta estrategia permitió obtener una evaluación más cercana al uso esperado de los modelos, ya que el desempeño final se midió sobre un período posterior al utilizado para el entrenamiento. Los resultados del conjunto de prueba se utilizaron posteriormente para comparar el desempeño de los modelos mediante las métricas definidas en la metodología de evaluación.

3.5 Metodología de modelado predictivo

3.5.1 Enfoque de modelado

El modelado se planteó como un problema de regresión supervisada, cuya variable objetivo correspondió a la cantidad de unidades vendidas por hora. La unidad de predicción estuvo definida por la combinación de fecha, hora, local y producto, de manera que cada observación representó la demanda horaria observada de un producto específico en uno de los locales incluidos en el estudio. Para cada registro se utilizaron las 18 variables predictoras descritas en la sección anterior.

Las devoluciones de producto no vendido no se incorporaron como variable de entrada de los modelos. Debido a que se registran principalmente al cierre de la jornada y presentan una granularidad diferente a la de las ventas POS, su utilización se mantuvo como una capa complementaria para contextualizar el comportamiento del producto no vendido, sin intervenir directamente en la estimación horaria.

Se aplicaron dos estrategias de modelado. La primera consistió en desarrollar modelos específicos para cada uno de los diez locales seleccionados. En este enfoque, cada modelo utilizó únicamente la información correspondiente al local analizado e incluyó los quince productos definidos para el estudio. El propósito fue permitir que los algoritmos aprendieran patrones propios de cada punto de venta.

La segunda estrategia correspondió a la construcción de modelos globales, utilizando de manera conjunta las observaciones de los diez locales. En este caso, las variables codificadas de local y producto permitieron diferenciar las distintas series dentro de un mismo conjunto de entrenamiento. La comparación entre ambos enfoques permitió

analizar el comportamiento de modelos especializados por local frente a modelos entrenados con información consolidada de toda la muestra.

Para el enfoque por local se consideraron seis métodos: Ridge, Random Forest, XGBoost, LightGBM, Support Vector Regression (SVR) y Holt-Winters. Los primeros cinco corresponden a modelos de aprendizaje automático y fueron aplicados también en el enfoque global. Holt-Winters se utilizó únicamente a nivel de local debido a que se implementó como una referencia estadística basada en series temporales univariadas.

La estrategia experimental comprendió, por tanto, 60 aplicaciones a nivel de local, correspondientes a seis métodos en cada uno de los diez locales, y cinco modelos globales, para un total de 65 evaluaciones. Esta cantidad representa ejecuciones comparativas bajo los dos enfoques definidos y no implica que todas sean metodológicamente equivalentes, debido a las particularidades de Holt-Winters que se explican en la siguiente sección.

La secuencia temporal de los datos se mantuvo durante todo el proceso de modelado mediante la partición cronológica descrita en la sección 3.4.7. Los modelos fueron ajustados con información histórica y evaluados sobre períodos posteriores, evitando la división aleatoria entre observaciones de diferentes momentos.

En los algoritmos capaces de producir estimaciones inferiores a cero, las predicciones negativas se ajustaron a cero antes del cálculo de las métricas, debido a que la cantidad de unidades vendidas no puede tomar valores negativos. Este procedimiento se aplicó únicamente como restricción posterior a la predicción y no modificó los valores observados de la variable objetivo.

El enfoque desarrollado tuvo como propósito comparar alternativas capaces de estimar la demanda horaria observada y analizar su utilidad como apoyo a la planificación del horneo. Las predicciones no se interpretaron como órdenes automáticas de producción ni como una reconstrucción exacta del proceso de horneo, sino como estimaciones cuantitativas que pueden complementar el criterio operativo utilizado en los locales.

3.5.2 Modelos predictivos evaluados

Con el propósito de comparar enfoques con características metodológicas diferentes, se evaluaron seis métodos predictivos: Ridge Regression, Random Forest, XGBoost, LightGBM, Support Vector Regression (SVR) y Holt-Winters. La selección permitió contrastar un modelo lineal regularizado, métodos basados en árboles, algoritmos de potenciación por gradiente, un enfoque de máquinas de vectores de soporte y un método estadístico tradicional de series temporales.

Ridge Regression se incorporó como un modelo lineal regularizado de referencia. Su estructura permite estimar relaciones lineales entre las características predictoras y la variable objetivo, incorporando una penalización que reduce la magnitud de los coeficientes y limita problemas asociados con la multicolinealidad. Su inclusión permitió disponer de un punto de comparación relativamente simple frente a modelos capaces de representar relaciones no lineales.

Random Forest se utilizó como representante de los métodos de ensamble basados en múltiples árboles de decisión. Este algoritmo combina las estimaciones generadas por distintos árboles contruidos sobre subconjuntos de observaciones y variables, lo que permite modelar relaciones no lineales e interacciones entre características sin requerir una

especificación funcional previa. Su utilización resulta pertinente para conjuntos de datos que combinan variables temporales, categóricas e históricas.

XGBoost se incorporó como un algoritmo de *gradient boosting*, en el que los árboles se construyen de manera secuencial con el propósito de corregir progresivamente los errores de las iteraciones anteriores. Su estructura permite representar relaciones complejas entre las variables predictoras e incorporar mecanismos de regularización para controlar el ajuste del modelo.

LightGBM pertenece igualmente a la familia de *gradient boosting* basada en árboles. Su inclusión permitió comparar otra implementación de esta familia de algoritmos, caracterizada por un esquema de crecimiento y procesamiento orientado a mejorar la eficiencia computacional sobre conjuntos de datos de mayor volumen. La comparación con XGBoost permitió analizar el comportamiento de dos alternativas de potenciación por gradiente aplicadas sobre el mismo conjunto de características.

Support Vector Regression (SVR) se utilizó como un enfoque de regresión basado en máquinas de vectores de soporte. Mediante el uso de un núcleo no lineal, el algoritmo puede representar relaciones complejas entre las variables de entrada y la demanda observada, buscando una función que mantenga los errores dentro de un margen definido. Debido al costo computacional asociado con este método en conjuntos grandes, su entrenamiento se realizó sobre una submuestra, cuya configuración se detalla en la sección siguiente.

Holt-Winters se incorporó como un método estadístico tradicional de series temporales, con el propósito de disponer de una referencia metodológica diferente de los

modelos de aprendizaje automático. Este enfoque modela componentes de nivel, tendencia y estacionalidad mediante suavizamiento exponencial.

La implementación de Holt-Winters presenta una diferencia importante respecto de los otros cinco métodos. Mientras los modelos de aprendizaje automático se aplicaron sobre la base horaria y utilizaron las características descritas en la sección 3.4.6, Holt-Winters se construyó sobre una serie diaria agregada y posteriormente distribuyó la estimación diaria entre las horas consideradas para la operación. Por este motivo, sus resultados se utilizaron como una referencia de series temporales y no como una alternativa estrictamente equivalente a los modelos horarios.

La Tabla 7 resume los modelos evaluados y la función que cumplieron dentro del diseño metodológico.

Tabla 7

Modelos predictivos evaluados en la investigación

Modelo	Enfoque	Característica principal	Aplicación
Ridge	Regresión lineal regularizada	Modelo lineal con penalización de coeficientes	Por local y global
Random Forest	Ensamblaje de árboles	Modela relaciones no lineales mediante múltiples árboles	Por local y global
XGBoost	Potenciación por gradiente	Construcción secuencial de árboles con regularización	Por local y global
LightGBM	Potenciación por gradiente	Potenciación por gradiente con procesamiento eficiente	Por local y global

SVR	Máquinas de vectores de soporte	Regresión no lineal mediante función kernel	Por local y global
Holt- Winters	Series temporales	Suavizamiento exponencial con tendencia y estacionalidad	Solo por local; referencia diaria

La Tabla 7 resume los seis métodos evaluados y evidencia la diversidad de enfoques considerados dentro del diseño metodológico. Esta diversidad permitió evaluar si el comportamiento de la demanda horaria podía representarse mejor mediante relaciones lineales, estructuras basadas en árboles, métodos de margen no lineal o un enfoque clásico de series temporales.

3.5.3 Configuración e hiperparámetros de los modelos

Los modelos se entrenaron a partir de las particiones temporales definidas previamente, manteniendo la misma estructura de variables predictoras para los algoritmos evaluados. Las configuraciones utilizadas se establecieron con el propósito de mantener un procedimiento reproducible y computacionalmente viable, más que de alcanzar el máximo desempeño posible de cada método. No se realizó una búsqueda exhaustiva de hiperparámetros mediante procedimientos como búsqueda en rejilla (*grid search*), búsqueda aleatoria (*random search*) u optimización bayesiana. Por esta razón, los resultados obtenidos representan el desempeño de las configuraciones implementadas y no necesariamente el mejor resultado alcanzable por cada algoritmo.

Para **Ridge Regression** se utilizó un coeficiente de regularización $\alpha = 1,0$. Debido a que este modelo es sensible a la escala de las variables, las características predictoras fueron previamente estandarizadas mediante `StandardScaler`. El escalador se

ajustó exclusivamente con los datos de entrenamiento y posteriormente se reutilizó para transformar la información sobre la cual se generaron las predicciones, evitando utilizar información del conjunto de prueba durante su ajuste.

En **Random Forest** se configuraron 100 árboles ($n_estimators = 100$), una profundidad máxima de 15 niveles ($max_depth = 15$) y un mínimo de cinco observaciones por hoja ($min_samples_leaf = 5$). Se utilizó una semilla aleatoria de 42 para mantener la reproducibilidad de las ejecuciones y $n_jobs = -1$ para aprovechar los núcleos de procesamiento disponibles durante el entrenamiento.

Para **XGBoost** se definió un máximo de 600 estimadores ($n_estimators = 600$), una profundidad máxima de cuatro niveles ($max_depth = 4$) y una tasa de aprendizaje de 0,05 ($learning_rate = 0,05$). Asimismo, se utilizaron proporciones de muestreo de 0,8 tanto para las observaciones ($subsample = 0,8$) como para las variables ($colsample_bytree = 0,8$). La configuración incorporó regularización L1 y L2 mediante $reg_alpha = 0,1$ y $reg_lambda = 2,0$, además de $min_child_weight = 5$ y el método de construcción de árboles basado en histogramas. Se fijó una semilla aleatoria de 42.

El conjunto de validación se utilizó en XGBoost para aplicar un mecanismo de parada temprana (early stopping) después de 40 iteraciones consecutivas sin mejora. Por este motivo, el número efectivo de árboles pudo ser inferior al máximo configurado, ya que el entrenamiento se detenía cuando el desempeño sobre el conjunto de validación dejaba de mejorar.

LightGBM se configuró con 200 estimadores ($n_estimators = 200$), una profundidad máxima de ocho niveles ($max_depth = 8$) y una tasa de aprendizaje de 0,1.

También se establecieron proporciones de 0,8 para el muestreo de observaciones y de variables y una semilla aleatoria de 42. El conjunto de validación se incorporó como referencia durante el entrenamiento, pero en la configuración utilizada no se implementó un mecanismo explícito de parada temprana.

Para **Support Vector Regression (SVR)** se utilizó un núcleo de base radial (RBF), con $C = 10$ y $\epsilon = 0,1$. Las variables fueron estandarizadas antes del entrenamiento utilizando la misma transformación ajustada sobre el conjunto de entrenamiento. Debido al costo computacional del algoritmo sobre conjuntos de gran volumen, se empleó una submuestra aleatoria reproducible de hasta 10.000 observaciones para los modelos por local. En el modelo global se utilizó una submuestra de hasta 15.000 observaciones. El submuestreo se aplicó únicamente durante el entrenamiento; la evaluación se realizó sobre las observaciones disponibles del conjunto de prueba.

Holt-Winters se aplicó sobre una serie diaria correspondiente al producto con mayor volumen de ventas dentro de cada local. El modelo utilizó tendencia aditiva, estacionalidad aditiva y un período estacional de siete días. Los parámetros internos fueron estimados mediante el procedimiento de optimización proporcionado por la biblioteca estadística utilizada, y el modelo se ajustó únicamente cuando la serie de entrenamiento disponía de al menos 14 observaciones.

Debido a que Holt-Winters genera predicciones a nivel diario, la implementación complementaria del proyecto permitió representar esas estimaciones en una escala horaria mediante su distribución uniforme entre las 17 horas comprendidas entre las 06:00 y las 22:00. Sin embargo, esta transformación no modifica el hecho de que el método fue entrenado y evaluado originalmente sobre una serie diaria; por esta razón, sus resultados se

consideran una referencia estadística complementaria y no una alternativa estrictamente equivalente a los modelos entrenados directamente con observaciones horarias

Tabla 8

Configuración e hiperparámetros de los modelos predictivos

Modelo	Hiperparámetros principales	Tratamiento adicional
Ridge Regression	$\alpha = 1,0$	Estandarización de variables
Random Forest	100 árboles; profundidad máxima 15; mínimo 5 observaciones por hoja	Semilla 42 y procesamiento paralelo
XGBoost	600 estimadores; profundidad 4; tasa de aprendizaje 0,05; subsample = 0,8; colsample_bytree = 0,8; reg_alpha = 0,1; reg_lambda = 2,0; min_child_weight = 5	Parada temprana de 40 iteraciones; semilla 42
LightGBM	200 estimadores; profundidad 8; tasa de aprendizaje 0,1; subsample = 0,8; colsample_bytree = 0,8	Uso del conjunto de validación; semilla 42
SVR	Kernel RBF; C = 10; epsilon = 0,1	Estandarización y submuestreo
Holt-Winters	Tendencia aditiva; estacionalidad aditiva; período estacional de 7 días	Serie diaria del producto principal por local

La Tabla 8 resume las configuraciones efectivamente utilizadas y los tratamientos adicionales aplicados a cada algoritmo. Ridge y SVR requirieron estandarización previa de las variables; Random Forest, XGBoost y LightGBM trabajaron directamente con las

características de entrada; XGBoost incorporó regularización y parada temprana; y SVR utilizó submuestreo para mantener la viabilidad computacional del entrenamiento.

Los hiperparámetros reportados corresponden a las configuraciones efectivamente utilizadas en el proyecto. En consecuencia, la comparación realizada permite analizar el comportamiento de los diferentes métodos bajo un procedimiento reproducible, pero no constituye una comparación entre versiones exhaustivamente optimizadas de cada algoritmo.

Finalmente, cuando los modelos generaron predicciones negativas, estas se ajustaron a cero antes de calcular las métricas, debido a que la variable objetivo representa una cantidad de unidades vendidas y no admite valores inferiores a cero. Esta restricción se aplicó únicamente sobre las predicciones y no modificó los valores observados utilizados durante el entrenamiento o la evaluación.

3.5.4 Métricas de evaluación

El desempeño de los modelos predictivos se evaluó mediante cuatro métricas complementarias: error absoluto medio (MAE), raíz del error cuadrático medio (RMSE), error porcentual absoluto medio (MAPE) y coeficiente de determinación (R^2). La utilización conjunta de estos indicadores permitió analizar el comportamiento de los modelos desde diferentes perspectivas, considerando tanto la magnitud absoluta de los errores como su comportamiento relativo y el nivel de ajuste alcanzado.

En esta investigación, el **MAE** se estableció como la métrica principal de comparación debido a que expresa el error promedio en la misma unidad de la variable

objetivo, es decir, unidades de producto vendidas por hora. Esta característica facilita su interpretación dentro del contexto operativo del estudio.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Donde y_i corresponde al valor observado, $\hat{y}_i$ al valor estimado por el modelo y n al número de observaciones evaluadas.

El **RMSE** se utilizó como una medida complementaria, debido a que asigna un mayor peso a los errores de mayor magnitud. Esto permite identificar modelos que, aun presentando un error absoluto promedio reducido, pueden generar desviaciones importantes en determinadas observaciones.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Al igual que el MAE, el RMSE se expresa en unidades de producto por hora, lo que permite relacionar su resultado con la escala de la variable objetivo.

El **MAPE** se incorporó como una medida relativa del error expresada en términos porcentuales:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Para su cálculo se excluyeron las observaciones en las que el valor observado fuera igual a cero. No obstante, incluso bajo esta condición, el MAPE puede presentar valores

elevados cuando la demanda observada es reducida, debido a que diferencias pequeñas en términos absolutos pueden representar errores porcentuales considerables. Por este motivo, en la investigación se utilizó como una referencia complementaria y no como criterio principal para seleccionar los modelos.

El **coeficiente de determinación (R^2)** se utilizó para complementar el análisis del ajuste de los modelos respecto de la variabilidad observada en la variable objetivo:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Un valor de R^2 más cercano a uno indica una mayor capacidad del modelo para representar la variabilidad observada respecto de la referencia basada en la media. Los valores negativos indican que, para el conjunto evaluado, el modelo presenta un desempeño inferior al obtenido mediante esa referencia. Debido a que R^2 no expresa directamente la magnitud del error en unidades de producto, se consideró como una métrica complementaria.

La comparación de los modelos se realizó principalmente a partir del MAE. El RMSE permitió complementar este análisis al otorgar mayor importancia a errores de gran magnitud, mientras que R^2 aportó información sobre el ajuste general de cada modelo. El MAPE se utilizó con carácter descriptivo y se interpretó con cautela debido a su sensibilidad frente a valores bajos de demanda.

La Tabla 9 resume la función de cada métrica dentro del proceso de evaluación.

Tabla 9

Métricas utilizadas para evaluar los modelos predictivos

Métrica	Unidad	Qué mide	Pertinencia para el estudio	Rol en la evaluación
MAE	Unidades por hora	Magnitud promedio del error absoluto	Permite interpretar directamente el error en la escala de la demanda horaria observada	Principal
RMSE	Unidades por hora	Magnitud del error con mayor penalización de desviaciones grandes	Permite identificar la presencia de errores de mayor magnitud	Complementaria
MAPE	Porcentaje	Error relativo respecto del valor observado	Aporta una referencia porcentual, pero es sensible a valores bajos	Referencia complementaria
R ²	Adimensional	Ajuste del modelo respecto de la variabilidad observada	Complementa el análisis del desempeño general, pero no expresa error operativo	Complementaria

Como se aprecia en la Tabla 9, cada métrica cumple una función distinta: el MAE aporta la referencia principal, mientras que RMSE, MAPE y R² complementan la interpretación.

Las métricas se calcularon para los modelos desarrollados por local y para los modelos globales, considerando las observaciones correspondientes al conjunto de prueba. Antes del cálculo, las predicciones con valores negativos se ajustaron a cero, debido a que

la variable objetivo representa cantidades vendidas y no admite valores inferiores a ese límite. Los resultados obtenidos fueron almacenados en los artefactos de evaluación del proyecto para su posterior análisis y comparación.

3.5.5 Criterio de comparación y selección de modelos

La comparación de los modelos se realizó sobre las métricas obtenidas en el conjunto de prueba, manteniendo como criterio principal el error absoluto medio (MAE). Esta elección respondió a que el MAE expresa la diferencia promedio entre los valores observados y estimados en unidades de producto por hora, lo que facilita su interpretación dentro del problema analizado.

El criterio principal estableció que, dentro de cada conjunto comparable, un menor valor de MAE representa un mejor desempeño predictivo. El RMSE y el coeficiente de determinación (R^2) se utilizaron como métricas complementarias para analizar, respectivamente, la presencia de errores de mayor magnitud y el ajuste general del modelo. El MAPE se mantuvo como referencia descriptiva y no se utilizó como criterio principal de selección debido a su sensibilidad frente a valores bajos de demanda, como se explicó en la sección anterior.

La comparación se efectuó en dos niveles. En primer lugar, se analizaron los resultados de los modelos desarrollados para cada uno de los diez locales, lo que permitió identificar si el algoritmo con menor error era consistente entre los distintos puntos de venta o si el desempeño variaba según el local. En segundo lugar, se evaluaron los modelos globales entrenados con la información conjunta de los locales seleccionados. Esta

separación permitió analizar de manera diferenciada el comportamiento de los enfoques específicos por local y del enfoque global.

Cuando las diferencias de MAE entre dos o más modelos fueron reducidas, la interpretación no se limitó únicamente a determinar cuál presentó el menor valor numérico. En estos casos se consideraron conjuntamente el RMSE y el R^2 para valorar si las diferencias observadas mostraban un comportamiento consistente en las métricas complementarias. Por esta razón, pequeñas diferencias decimales entre algoritmos no se interpretaron automáticamente como una superioridad sustancial de un modelo sobre otro.

Holt-Winters se analizó de manera separada dentro de la comparación. Como se explicó en las secciones anteriores, este método fue ajustado sobre series diarias y sus estimaciones fueron posteriormente representadas en escala horaria. En consecuencia, sus métricas se utilizaron como referencia estadística complementaria y no para establecer una comparación estrictamente equivalente con los cinco algoritmos entrenados directamente sobre observaciones horarias.

La Tabla 10 resume los criterios utilizados en el proceso de comparación.

Tabla 10

Criterios utilizados para la comparación de los modelos predictivos

Criterio	Aplicación en la comparación	Prioridad
MAE	Identificar el menor error promedio en unidades por hora	Principal
RMSE	Examinar la presencia de errores de mayor magnitud	Complementaria

R^2	Analizar el ajuste general respecto de la variabilidad observada	Complementaria
MAPE	Disponer de una referencia porcentual del error	Descriptiva
Comportamiento por local	Identificar diferencias de desempeño entre los diez locales	Complementaria
Modelo global	Analizar el desempeño utilizando información conjunta de los locales	Complementaria

La selección no se sustentó en un único indicador, como se resume en la Tabla 10. El MAE constituyó el criterio principal, mientras que las demás métricas permitieron complementar la interpretación de las diferencias observadas entre los algoritmos.

El modelo con mejor desempeño se determinó a partir de los resultados obtenidos bajo este esquema de comparación, considerando tanto el comportamiento agregado como las diferencias observadas entre locales. En consecuencia, la metodología no presupuso la existencia de un algoritmo universalmente superior, sino que permitió identificar qué alternativas presentaron un mejor comportamiento dentro de las condiciones específicas del conjunto de datos analizado.

Capítulo 4

4. Resultados y Discusión

4.1 Análisis exploratorio de los datos

El análisis exploratorio tuvo como propósito caracterizar el comportamiento temporal de las ventas, identificar diferencias entre locales y productos y examinar, de

manera complementaria, los registros de producto no vendido. Esta etapa permitió reconocer patrones relevantes para la construcción del problema predictivo y aportó elementos para sustentar decisiones posteriores relacionadas con la selección de variables, la delimitación de locales y productos y la definición de las estrategias de modelado.

Figura 2

Serie temporal diaria de ventas con medias móviles



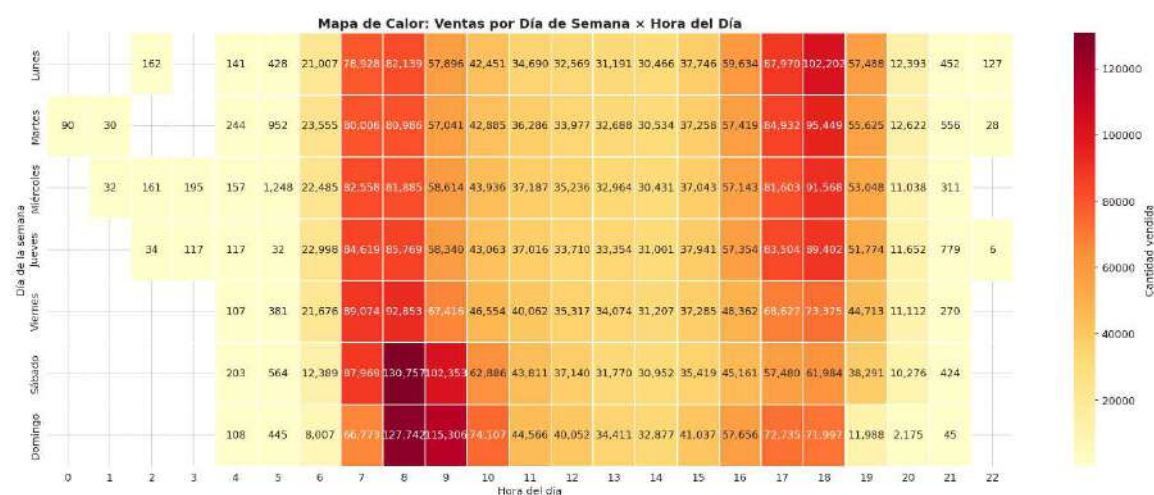
La Figura 2 muestra que el nivel diario de ventas presenta fluctuaciones a lo largo del período analizado y que estas variaciones no se reducen únicamente a cambios aislados entre días. Las medias móviles permiten observar con mayor claridad períodos en los que el nivel de ventas aumenta, disminuye o mantiene comportamientos relativamente diferenciados, lo que evidencia la existencia de una estructura temporal en los registros.

Este comportamiento resulta relevante para el modelado porque muestra que la demanda observada presenta una dimensión temporal que debe considerarse en el análisis. La evolución reciente de las ventas aporta información sobre el comportamiento posterior, lo que sustentó la incorporación de variables históricas, como rezagos y estadísticas

móviles, además de variables de calendario destinadas a representar esta dimensión temporal.

Figura 3

Mapa de calor: ventas por día de semana por hora del día



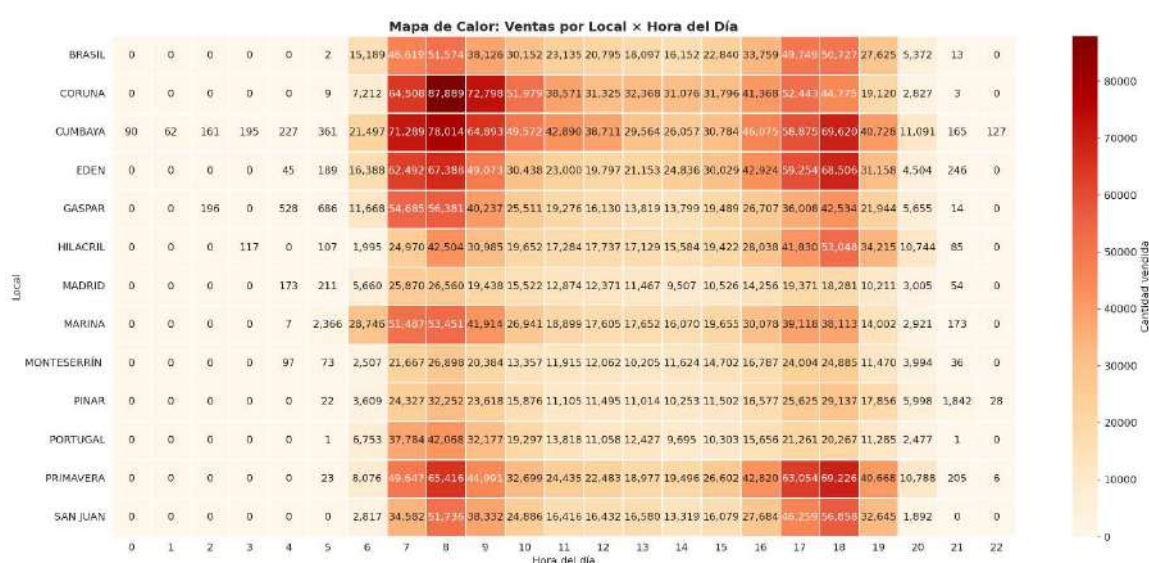
La Figura 3 evidencia que las ventas no se distribuyen uniformemente entre las horas del día ni entre los diferentes días de la semana. La presencia de zonas con distinta intensidad muestra que el nivel de demanda observada depende de la combinación entre el momento de la jornada y el día en que se produce la venta.

Este patrón justificó incorporar la hora, el día de la semana y el indicador de fin de semana entre las variables predictoras. Asimismo, debido al carácter periódico de estas variables, posteriormente se generaron representaciones cíclicas mediante funciones seno y coseno. La distribución observada muestra, por tanto, que la dimensión temporal constituye una fuente de información relevante para diferenciar el comportamiento de la demanda horaria.

Desde una perspectiva operativa, esta distribución resulta relevante porque la planificación del horneado requiere considerar no solo la cantidad total esperada durante la jornada, sino también los períodos en los que se concentra la demanda observada.

Figura 4

Mapa de calor: ventas por local por hora del día



Nota. La figura presenta los locales disponibles durante la fase exploratoria previa a la selección final del conjunto de modelado.

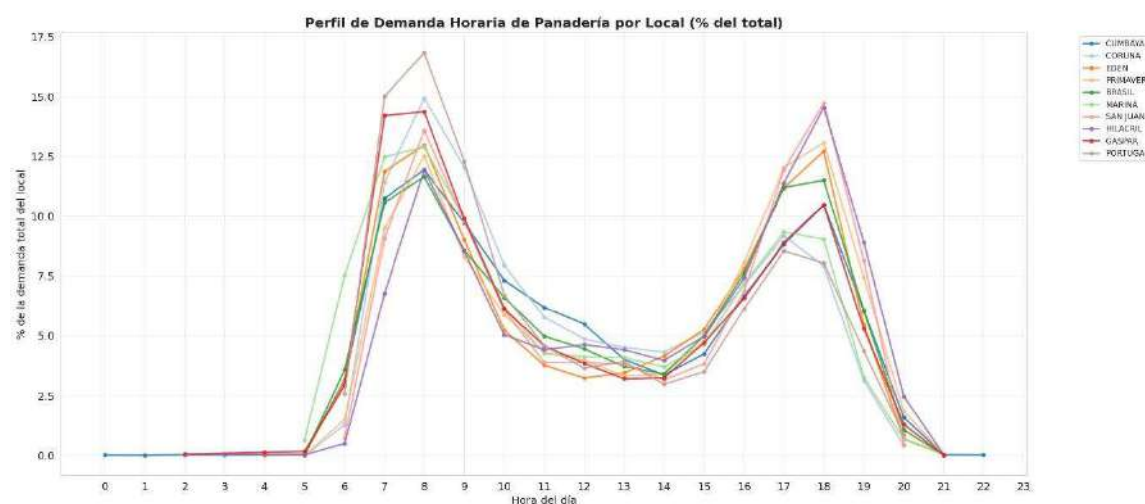
La Figura 4 muestra que los locales presentan diferencias tanto en el nivel de ventas como en su distribución a lo largo de la jornada. La intensidad de la actividad no se concentra necesariamente en las mismas horas para todos los puntos de venta, lo que evidencia que el comportamiento horario no puede considerarse homogéneo para toda la cadena.

Esta heterogeneidad fue relevante para la estrategia de modelado. Si todos los locales presentaran perfiles horarios equivalentes, un único comportamiento agregado

podría resultar suficiente para representar la demanda; sin embargo, las diferencias observadas justificaron evaluar, además del enfoque global, modelos específicos por local. De esta manera fue posible analizar posteriormente si un mismo algoritmo mantenía un desempeño similar en establecimientos con patrones de venta diferentes.

Figura 5

Perfil de demanda horaria relativo por local



La Figura 5 complementa el análisis anterior al expresar las ventas de cada hora como proporción de la actividad diaria del local. Esta transformación permite comparar la forma del perfil horario sin que los establecimientos con mayor volumen absoluto de ventas dominen visualmente el análisis.

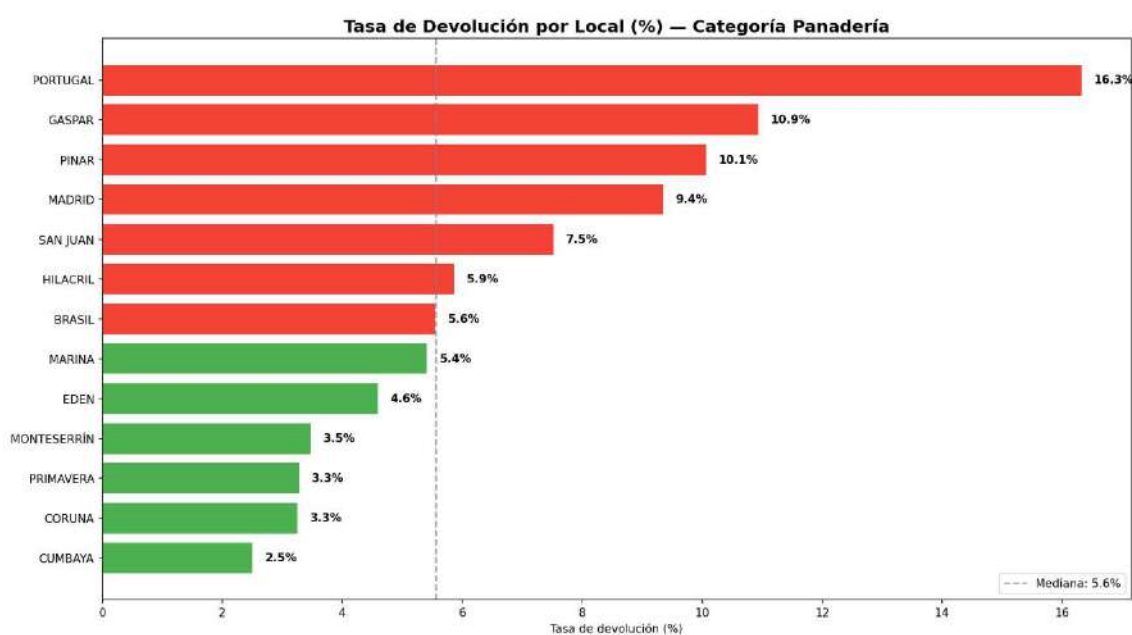
Las diferencias observadas en estos perfiles indican que la distribución intradiaria de la demanda presenta particularidades entre locales. En consecuencia, dos puntos de venta con volúmenes diarios diferentes pueden también requerir distribuciones horarias distintas. Este resultado resulta particularmente relevante para el objetivo de la

investigación, debido a que la planificación del horneado requiere información sobre cuándo se concentra la demanda, y no únicamente sobre el volumen total esperado durante el día.

Por tanto, el análisis relativo permite distinguir entre el volumen total de venta y la forma en que esa venta se distribuye durante el día, aspecto especialmente relevante cuando la decisión operativa se realiza por horas o franjas y no únicamente a nivel diario.

Figura 6

Tasa de devolución por local

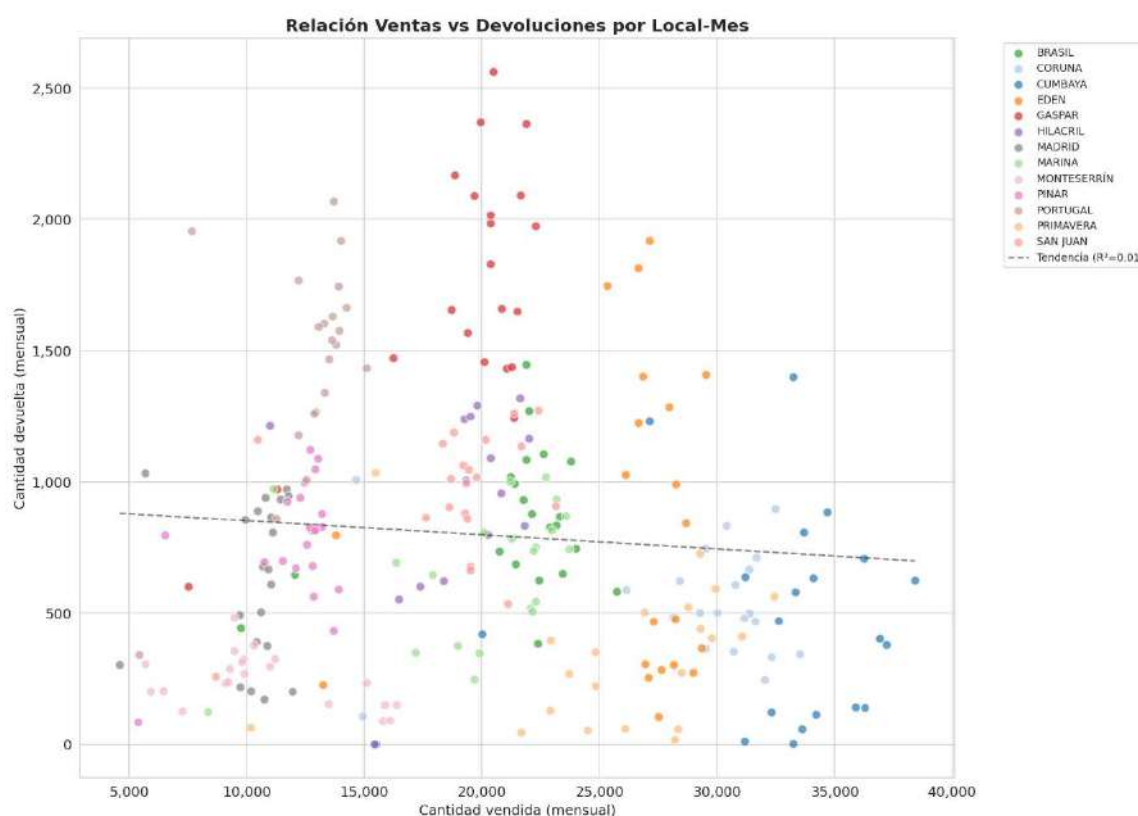


La Figura 6 muestra que la proporción de producto registrado como no vendido difiere entre los locales analizados. La existencia de estas diferencias indica que el comportamiento de las devoluciones no es uniforme y que una misma cantidad absoluta devuelta puede tener una importancia distinta dependiendo del volumen de producto vendido en cada local.

No obstante, la tasa de devolución debe interpretarse como un indicador complementario y no como una medición directa de producción, merma o eficiencia del local. Como se estableció en el Capítulo 3, la información disponible no permite conocer de manera completa cuánto producto fue efectivamente horneado ni su disponibilidad durante cada hora de la jornada. Por esta razón, este resultado se utilizó para contextualizar el comportamiento del producto no vendido y no como variable objetivo o predictora del modelo.

Figura 7

Relación entre ventas y devoluciones por local y mes



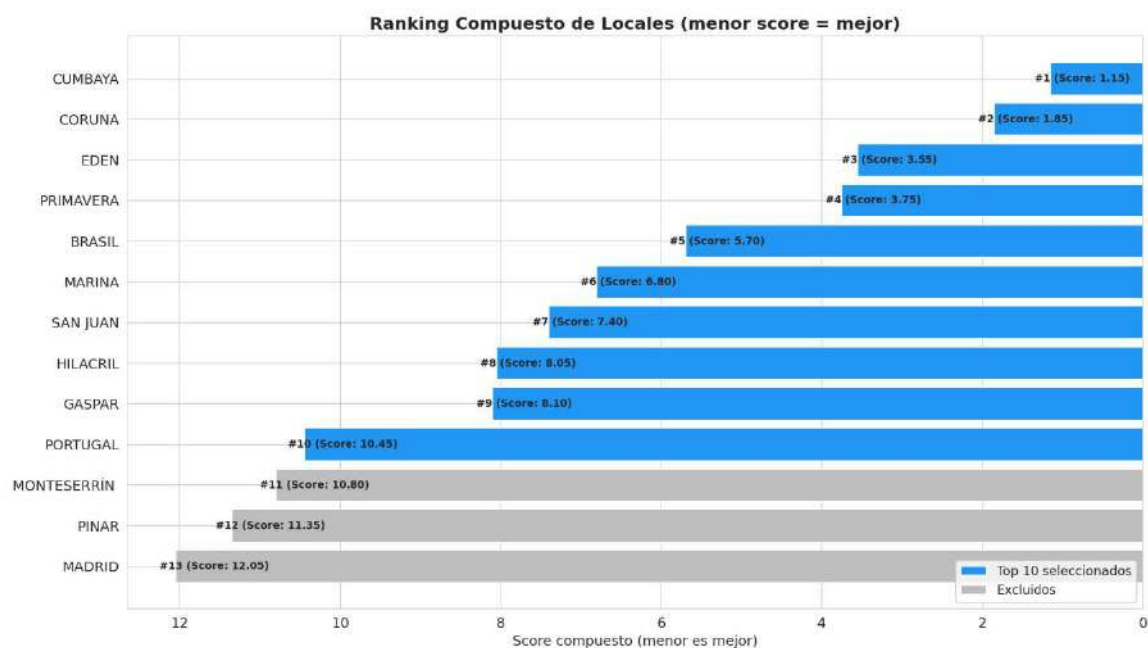
La Figura 7 permite analizar conjuntamente el volumen de ventas y las devoluciones registradas por local y mes. La dispersión observada indica que los cambios

en el volumen vendido no se corresponden de manera uniforme con cambios proporcionales en las devoluciones. En otras palabras, un mayor nivel de ventas no implica necesariamente una mayor proporción de producto registrado como no vendido.

Este resultado contribuyó a diferenciar metodológicamente ambas fuentes. Mientras las ventas POS proporcionan la información utilizada para construir la demanda horaria observada, las devoluciones describen una condición registrada principalmente al cierre de la jornada y con una granularidad diferente. Su comportamiento exploratorio respaldó, por tanto, la decisión de mantenerlas como una capa analítica complementaria, sin integrarlas directamente en la construcción de la variable objetivo.

Figura 8

Ranking compuesto de locales



Nota. Un menor valor del puntaje compuesto corresponde a una mejor posición dentro del criterio de ordenamiento utilizado.

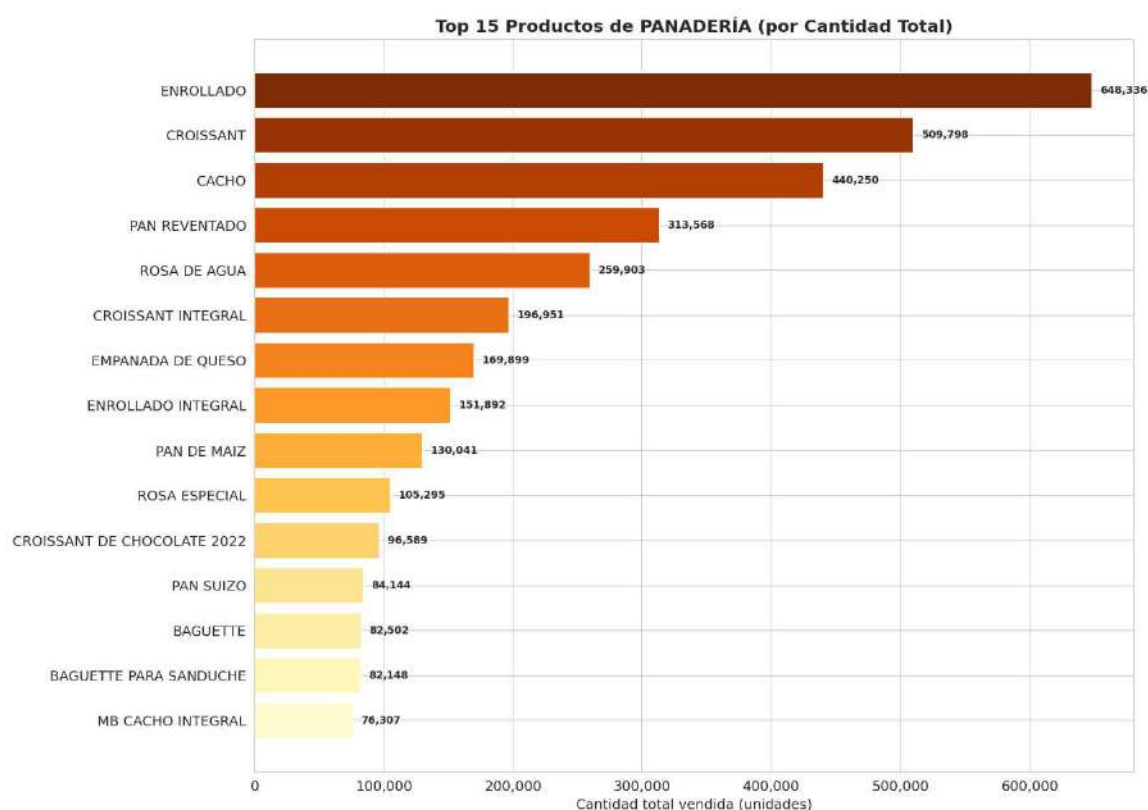
La Figura 8 presenta el ordenamiento de los locales obtenido a partir de los criterios definidos para delimitar el conjunto destinado al modelado. El ranking combinó información sobre cantidad vendida, valor de ventas, número de transacciones, promedio diario de unidades y el indicador complementario asociado a la relación entre ventas y devoluciones. Los locales que posteriormente fueron excluidos se mantienen en la figura con el propósito de mostrar el conjunto analizado antes de aplicar la delimitación final.

Más que establecer una clasificación de desempeño operativo, el ranking tuvo una función metodológica de selección, orientada a identificar locales con suficiente volumen y continuidad de información para desarrollar y comparar los modelos predictivos. A partir de este procedimiento, junto con las exclusiones relacionadas con el alcance y continuidad

operativa descritas en el Capítulo 3, se determinaron los diez locales utilizados en las etapas posteriores.

Figura 9

Top 15 de productos de panadería seleccionados



La Figura 9 presenta los quince productos seleccionados dentro de la categoría Panadería, considerando principalmente la cantidad total vendida durante el período analizado. La concentración del análisis en estos productos permitió disponer de un volumen suficiente de observaciones para construir las series local-producto y desarrollar posteriormente el modelado horario.

La selección no pretende afirmar que estos productos representen estadísticamente la totalidad del portafolio de la cadena, sino delimitar la investigación a un conjunto con

suficiente información para la comparación de los modelos. La combinación de los quince productos con los diez locales seleccionados definió un universo potencial de 150 combinaciones local-producto para las etapas posteriores del modelado.

El análisis exploratorio mostró que la demanda observada presenta una estructura temporal, diferencias en su distribución horaria entre locales y variaciones en el comportamiento de los productos analizados. Estos hallazgos sustentaron la incorporación de variables temporales e históricas y la evaluación diferenciada de modelos por local y modelos globales. El análisis de las devoluciones evidenció un comportamiento distinto del registrado por las ventas y respaldó su utilización como información complementaria, en lugar de incorporarlas directamente en la estimación de la demanda horaria. Los procedimientos de selección permitieron delimitar el modelado a diez locales y quince productos con suficiente volumen y continuidad de información para desarrollar la comparación predictiva.

4.2 Resultados del modelo predictivo

El conjunto utilizado para el modelado estuvo conformado por 498.750 observaciones horarias, obtenidas después de la selección de los diez locales y quince productos, la agregación de las transacciones por fecha, hora, local y producto y la construcción de las variables predictoras descritas en el Capítulo 3. La información se distribuyó cronológicamente en los conjuntos de entrenamiento, validación y prueba con el propósito de evaluar el comportamiento de los modelos sobre períodos posteriores.

La evaluación consideró dos enfoques: modelos desarrollados de manera específica para cada local y modelos globales entrenados con la información conjunta de los diez

locales. En total se realizaron 60 evaluaciones a nivel de local y cinco evaluaciones globales, correspondientes a los métodos definidos en la metodología.

Los resultados se presentan a continuación considerando el comportamiento de los modelos entre las particiones temporales, su desempeño en el conjunto de prueba, las diferencias observadas entre locales, la validación adicional con información posterior y la síntesis comparativa de las alternativas evaluadas.

4.2.1. Desempeño en entrenamiento, validación y prueba

La comparación entre las particiones de entrenamiento, validación y prueba permitió analizar cómo se comportaron los modelos al pasar de la información utilizada durante su ajuste a períodos cronológicamente posteriores. Para este análisis se consideraron principalmente el MAE y el coeficiente de determinación (R^2), debido a que permiten observar tanto la magnitud promedio del error como los cambios en la capacidad de representar la variabilidad de la demanda.

La Tabla 11 presenta el desempeño promedio obtenido por los cinco modelos de aprendizaje automático en los diez locales para las tres particiones temporales.

Tabla 11

Desempeño promedio de los modelos en entrenamiento, validación y prueba

Modelo	Train MAE	Train R^2	Val MAE	Val R^2	Test MAE	Test R^2
XGBoost	3,14	0,362	3,06	0,303	3,00	0,307
LightGBM	2,86	0,501	3,14	0,268	3,00	0,305
SVR	3,07	0,250	3,05	0,215	3,01	0,219
Random Forest	2,43	0,589	3,21	0,257	3,08	0,288

Ridge	3,46	0,211	3,24	0,212	3,23	0,203
-------	------	-------	------	-------	------	-------

Los valores de la Tabla 11 muestran comportamientos distintos entre los algoritmos al pasar del entrenamiento a períodos posteriores. XGBoost presentó una variación reducida entre las tres particiones: su MAE pasó de 3,14 en entrenamiento a 3,06 en validación y 3,00 en prueba, mientras que el R^2 disminuyó de 0,362 a valores cercanos a 0,30 en los dos conjuntos posteriores. Este comportamiento indica que el nivel de error se mantuvo relativamente estable entre los períodos analizados, aunque la proporción de variabilidad representada fue menor fuera del conjunto de entrenamiento.

LightGBM obtuvo un menor error y un mayor R^2 durante el entrenamiento, con un MAE de 2,86 y un R^2 de 0,501. En validación, el MAE aumentó a 3,14 y el R^2 disminuyó a 0,268; posteriormente, en prueba, el MAE se redujo a 3,00 y el R^2 alcanzó 0,305. La diferencia entre entrenamiento y los períodos posteriores evidencia una mayor adaptación a la información utilizada durante el ajuste que la observada en XGBoost, aunque el desempeño final en prueba se mantuvo entre los mejores resultados obtenidos.

Random Forest presentó la diferencia más marcada entre entrenamiento y los conjuntos posteriores. Su MAE aumentó de 2,43 en entrenamiento a 3,21 en validación y 3,08 en prueba, mientras que el R^2 disminuyó de 0,589 a 0,257 y 0,288, respectivamente. Esta brecha sugiere una mayor sensibilidad a las características del conjunto de entrenamiento. No obstante, los resultados de prueba muestran que el modelo conservó capacidad predictiva sobre información posterior, aunque con un desempeño inferior al alcanzado por XGBoost y LightGBM en las métricas principales.

Ridge mostró resultados relativamente próximos entre las tres particiones. El R^2 se mantuvo alrededor de 0,20 y el MAE pasó de 3,46 en entrenamiento a 3,24 en validación y 3,23 en prueba. Sin embargo, la estabilidad entre particiones no implicó un mejor desempeño absoluto, ya que presentó el mayor MAE de los cinco modelos de aprendizaje automático en el conjunto de prueba. Este resultado muestra que una menor diferencia entre entrenamiento y prueba debe analizarse conjuntamente con el nivel de error alcanzado.

SVR presentó igualmente variaciones reducidas entre las particiones, con valores de MAE de 3,07, 3,05 y 3,01 para entrenamiento, validación y prueba, respectivamente. El R^2 disminuyó de 0,250 en entrenamiento a 0,215 en validación y posteriormente alcanzó 0,219 en prueba. Este comportamiento resulta relevante considerando que el modelo se entrenó mediante una submuestra de los datos, como se explicó en el Capítulo 3, y aun así obtuvo niveles de error próximos a los algoritmos con mejor desempeño.

En algunos modelos, el MAE obtenido en validación o prueba fue incluso inferior al registrado durante el entrenamiento. Este comportamiento no debe interpretarse automáticamente como una mejora del modelo sobre datos no observados, ya que el nivel y la dispersión de la demanda pueden variar entre los períodos temporales. Un conjunto posterior con valores de demanda diferentes puede producir un menor error absoluto y, al mismo tiempo, un R^2 inferior, debido a que ambas métricas evalúan aspectos distintos del desempeño. Esta situación se analiza con mayor profundidad en la sección de discusión.

La comparación entre particiones muestra que la estabilidad entre particiones no constituye por sí sola un criterio suficiente para seleccionar un modelo. XGBoost, LightGBM y SVR alcanzaron valores de MAE muy próximos en el conjunto de prueba, aunque presentaron comportamientos diferentes entre entrenamiento y validación. Por esta

razón, la comparación final se complementó con RMSE, MAPE, R^2 y el análisis del desempeño por local, cuyos resultados se presentan en las siguientes secciones.

4.2.2. Comparación de modelos en el conjunto de prueba

La evaluación final de los modelos se realizó sobre el conjunto de prueba correspondiente al período comprendido entre el 1 de enero y el 14 de marzo de 2026. Al tratarse de información cronológicamente posterior a la utilizada durante el entrenamiento y la validación, estos resultados permitieron comparar el comportamiento de los algoritmos frente a observaciones no utilizadas para su ajuste.

La comparación se realizó principalmente mediante el error absoluto medio (MAE), definido en la metodología como el criterio principal de evaluación. De manera complementaria, se analizaron la raíz del error cuadrático medio (RMSE), el coeficiente de determinación (R^2) y el error porcentual absoluto medio (MAPE). Los resultados correspondientes a los modelos horarios se presentan como el promedio obtenido en los diez locales incluidos en el estudio.

Tabla 12

Comparación de los modelos horarios en el conjunto de prueba

Modelo	MAE ↓	RMSE ↓	MAPE (%)	R^2 ↑
XGBoost	3,00	4,97	98,6	0,307
LightGBM	3,00	4,98	97,7	0,305
SVR	3,01	5,28	81,1	0,219
Random Forest	3,08	5,04	103,3	0,288
Ridge	3,23	5,35	104,3	0,203

Holt-Winters*	22,48	27,82	29,6	-0,545
---------------	-------	-------	------	--------

Nota. Los resultados corresponden al promedio obtenido en los diez locales. Las flechas indican el sentido favorable de cada métrica: menor para MAE, RMSE y MAPE, y mayor para R^2 . Holt-Winters se ajustó sobre series de granularidad diaria y su estimación fue posteriormente representada en escala horaria; por esta razón, sus métricas se presentan únicamente como referencia estadística complementaria y no son directamente comparables con las de los cinco modelos entrenados sobre observaciones horarias.

En la Tabla 12 se observa que XGBoost y LightGBM alcanzaron resultados prácticamente equivalentes en las métricas principales. Ambos registraron un MAE promedio de 3,00 unidades por hora; XGBoost obtuvo un RMSE de 4,97 frente a 4,98 de LightGBM y un R^2 de 0,307 frente a 0,305. Estas diferencias son numéricamente muy reducidas y, por sí solas, no permiten sostener una superioridad sustancial de uno de los dos algoritmos.

SVR presentó igualmente un comportamiento competitivo en términos de MAE, con 3,01 unidades por hora, apenas 0,01 unidades por encima de XGBoost y LightGBM. Sin embargo, su RMSE fue mayor, con 5,28, y su R^2 alcanzó 0,219. El modelo registró el menor MAPE, con 81,1 %, pero este resultado debe interpretarse considerando la sensibilidad de esta métrica frente a valores bajos de demanda. Por esta razón, como se estableció en la metodología, el MAPE se utilizó como referencia complementaria y no como criterio principal de selección.

Random Forest obtuvo un MAE de 3,08 unidades por hora y un RMSE de 5,04, mientras que Ridge presentó los valores menos favorables entre los cinco modelos de aprendizaje automático, con un MAE de 3,23 y un RMSE de 5,35. Los resultados muestran,

en términos generales, un mejor comportamiento de los modelos no lineales frente al modelo lineal Ridge; sin embargo, las diferencias entre los tres algoritmos con menor MAE (XGBoost, LightGBM y SVR) fueron considerablemente menores que las observadas respecto de Ridge.

Holt-Winters presentó valores de error considerablemente superiores a los obtenidos por los modelos horarios. No obstante, este resultado no debe interpretarse como una comparación directa de capacidad predictiva entre métodos equivalentes, debido a que Holt-Winters fue entrenado sobre una serie diaria del producto principal de cada local y posteriormente su estimación fue representada en escala horaria. Por ello, su resultado se mantuvo únicamente como una referencia estadística de series temporales con una granularidad y un procedimiento diferentes.

Además de los valores promedio, se analizó la frecuencia con la que cada modelo alcanzó el resultado más favorable en los diferentes locales. Este análisis permitió determinar si un algoritmo mantenía una ventaja consistente en los diez puntos de venta o si el mejor desempeño variaba según el contexto analizado.

Tabla 13

Número de locales con mejor desempeño por métrica y modelo

Modelo	Victorias MAE	Victorias RMSE	Victorias R ²	Victorias MAPE
LightGBM	4 / 10	5 / 10	5 / 10	0 / 10
XGBoost	3 / 10	5 / 10	5 / 10	0 / 10
SVR	3 / 10	0 / 10	0 / 10	10 / 10
Random Forest	0 / 10	0 / 10	0 / 10	0 / 10

Ridge	0 / 10	0 / 10	0 / 10	0 / 10
-------	--------	--------	--------	--------

Nota. Cada valor representa el número de locales, de un total de diez, en los que el modelo obtuvo el resultado más favorable para la métrica correspondiente. Holt-Winters no se incluye debido a que su granularidad de evaluación no es directamente equivalente a la de los modelos horarios.

La Tabla 13 confirma que el resultado promedio no refleja por sí solo el comportamiento de los algoritmos en los distintos locales. En MAE, LightGBM obtuvo el mejor valor en cuatro locales, mientras que XGBoost y SVR lo hicieron en tres cada uno. En RMSE y R^2 , las mejores posiciones se distribuyeron por igual entre LightGBM y XGBoost, con cinco locales para cada modelo. Por otra parte, SVR obtuvo el menor MAPE en los diez locales.

La distribución de estas posiciones evidencia que no existió un único algoritmo dominante en todas las métricas y locales. Particularmente, la igualdad del MAE promedio entre XGBoost y LightGBM no implica que ambos se comportaran de manera idéntica en cada punto de venta, ya que el modelo con menor error varió entre los establecimientos.

Desde una perspectiva práctica, las diferencias promedio entre los modelos líderes también deben interpretarse en la escala de la variable objetivo. Una diferencia de 0,01 unidades por hora entre un MAE de 3,00 y 3,01 es muy reducida en términos absolutos y no constituye, por sí sola, evidencia suficiente para afirmar una ventaja operativamente relevante. Por este motivo, la comparación no se limitó al ordenamiento de los promedios, sino que consideró conjuntamente las métricas complementarias y la consistencia del desempeño entre los diferentes locales.

XGBoost y LightGBM presentaron los resultados promedio más equilibrados entre MAE, RMSE y R^2 , mientras que SVR mantuvo un MAE muy próximo y destacó en la medida porcentual. Estos resultados muestran que la elección entre los modelos con mejor desempeño requiere considerar el contexto de cada local y el criterio de evaluación priorizado, en lugar de asumir la existencia de un algoritmo universalmente superior.

4.2.3. Resultados por local en conjunto de prueba

El análisis por local permitió identificar diferencias en el desempeño de los modelos que no son visibles cuando únicamente se consideran los valores promedio. Para cada uno de los diez locales se determinó el algoritmo con menor MAE en el conjunto de prueba y se revisó, de manera complementaria, el R^2 correspondiente al modelo seleccionado.

Tabla 14

Modelo con menor MAE por local en el conjunto de prueba

Local	Mejor modelo (MAE)	MAE	R^2 del mejor
CUMBAYA	LightGBM	3,81	0,419
CORUNA	XGBoost	3,27	0,232
EDEN	LightGBM	3,69	0,346
PRIMAVERA	LightGBM	3,16	0,374
BRASIL	XGBoost	2,76	0,367
MARINA	SVR	2,87	0,023
SAN JUAN	LightGBM	2,77	0,388
HILACRIL	XGBoost	2,89	0,386
GASPAR	SVR	2,48	0,180

PORTUGAL	SVR	2,13	0,118
----------	-----	------	-------

Nota. El mejor modelo de cada local se determinó mediante el menor MAE registrado en el conjunto de prueba.

Los resultados de la Tabla 14 muestran que el algoritmo con menor MAE no fue el mismo en todos los locales. LightGBM obtuvo el menor MAE en cuatro puntos de venta, XGBoost en tres y SVR en tres. Esta distribución confirma que las diferencias observadas en los patrones de demanda entre locales también se reflejan en el desempeño de los modelos, por lo que los resultados promedio no deben interpretarse como evidencia de que un único algoritmo sea necesariamente el más adecuado para todos los casos.

Los valores de MAE de los modelos seleccionados oscilaron entre 2,13 unidades por hora en Portugal y 3,81 unidades por hora en Cumbayá. Esta variación indica que la dificultad de estimar la demanda no fue equivalente en todos los locales. Sin embargo, la comparación directa de estos valores debe realizarse con cautela, debido a que los puntos de venta presentan diferentes niveles de demanda, frecuencias de transacción y variabilidad en sus series. Un MAE mayor en un local no implica necesariamente un peor funcionamiento operativo del modelo en términos relativos.

El R^2 también presentó diferencias importantes entre los locales. Cumbayá alcanzó el mayor valor entre los modelos seleccionados, con 0,419, mientras que Marina registró 0,023. Los valores obtenidos muestran que, aun cuando un modelo pueda presentar el menor MAE disponible para un determinado local, su capacidad para representar la variabilidad de la demanda puede ser limitada.

El caso de Marina permite observar particularmente esta diferencia. SVR obtuvo el menor MAE del local, con 2,87 unidades por hora, pero un R^2 de 0,023. Esto indica que el modelo consiguió mantener un error absoluto relativamente reducido, aunque reprodujo de manera limitada las variaciones observadas en la demanda durante el período de prueba. Ambos resultados no son contradictorios, ya que MAE y R^2 evalúan dimensiones diferentes del desempeño.

Un comportamiento similar, aunque menos pronunciado, se observa en Portugal y Gaspar, donde SVR presentó los menores valores de MAE —2,13 y 2,48 unidades por hora, respectivamente— pero R^2 de 0,118 y 0,180. Estos resultados refuerzan la necesidad de evitar que la selección de un modelo se interprete exclusivamente a partir de una única métrica.

En contraste, los locales donde LightGBM y XGBoost obtuvieron el menor MAE presentaron, en general, valores de R^2 más altos dentro de los resultados seleccionados. Por ejemplo, Cumbayá, Primavera, Brasil, San Juan e Hilacril registraron valores comprendidos aproximadamente entre 0,37 y 0,42. No obstante, estos valores muestran también que una proporción importante de la variabilidad horaria permaneció sin representar, aspecto que se analiza con mayor profundidad en la sección de interpretación y discusión.

Los resultados por local muestran que el desempeño predictivo depende tanto del algoritmo como de las características particulares de cada serie de demanda. La ausencia de un modelo ganador en los diez locales respalda la decisión de evaluar los resultados de manera diferenciada y evita asumir que una misma alternativa debe utilizarse necesariamente en todos los puntos de venta.

4.2.4 Validación con datos posteriores

Como análisis complementario, los modelos fueron aplicados sobre información posterior al período utilizado para entrenamiento, validación y prueba. Esta evaluación consideró registros comprendidos entre el 15 de marzo y el 10 de julio de 2026 y se realizó específicamente para la combinación Cumbayá–Enrollado.

El propósito de esta validación adicional fue observar el comportamiento de los modelos frente a información temporalmente posterior a la utilizada durante su desarrollo. A diferencia de la evaluación principal, realizada con granularidad horaria, en este análisis las estimaciones y ventas observadas fueron comparadas a nivel diario, por lo que los valores obtenidos deben interpretarse en unidades por día y no son directamente equivalentes a las métricas horarias presentadas en las secciones anteriores.

La Tabla 15 presenta los resultados obtenidos para los cinco modelos de aprendizaje automático.

Tabla 15

Validación de los modelos con datos posteriores para Cumbayá–Enrollado

Modelo	MAE diario (unidades/día)	RMSE (unidades/día)	MAPE (%)
SVR	30,9	38,6	21,7
LightGBM	36,0	43,6	29,6
Random Forest	38,5	46,9	31,7
XGBoost	39,2	46,6	32,5

Ridge	55,4	62,2	45,4
-------	------	------	------

Nota. Los resultados corresponden exclusivamente a la combinación Cumbayá–Enrollado y fueron calculados con granularidad diaria sobre información posterior al conjunto de prueba.

Para la combinación Cumbayá–Enrollado, la Tabla 15 muestra que SVR obtuvo el menor error durante el período posterior analizado, con un MAE de 30,9 unidades por día, un RMSE de 38,6 unidades por día y un MAPE de 21,7 %. LightGBM presentó el segundo menor MAE, con 36,0 unidades por día, seguido de Random Forest con 38,5 y XGBoost con 39,2. Ridge registró los mayores errores de la comparación.

Este resultado resulta relevante porque el orden de desempeño observado en esta validación posterior no coincide completamente con el obtenido a partir de los valores promedio del conjunto de prueba. Mientras XGBoost y LightGBM presentaron los menores MAE promedio en la evaluación horaria, SVR obtuvo el menor error en esta combinación y período específicos. Esta diferencia refuerza la evidencia de que el comportamiento relativo de los algoritmos puede variar según el local, el producto, el período analizado y la granularidad utilizada para la evaluación.

Estos resultados deben interpretarse considerando el alcance de la comprobación realizada. La validación se efectuó sobre una única combinación local-producto, por lo que no permite concluir que SVR mantendría el mismo comportamiento en las demás combinaciones incluidas en el estudio. Además, al utilizarse una granularidad diaria, la magnitud de estos errores no puede compararse directamente con el MAE horario reportado en las secciones anteriores.

Esta evaluación constituye una validación temporal complementaria del caso Cumbayá–Enrollado. Su principal aporte consiste en mostrar cómo se comportaron los modelos frente a información posterior a la utilizada durante su desarrollo, sin extender este resultado al conjunto completo de locales y productos. Una evaluación posterior sobre un número mayor de combinaciones permitiría analizar con mayor amplitud la estabilidad temporal de los modelos.

4.2.5 Síntesis comparativa de los modelos

Los resultados obtenidos en las diferentes etapas de evaluación muestran que no existió un único algoritmo con el mejor desempeño en todos los locales, métricas y períodos analizados. La comparación promedio en el conjunto de prueba ubicó a XGBoost y LightGBM con los menores valores de MAE, ambos alrededor de 3,00 unidades por hora, mientras que SVR alcanzó un resultado muy próximo, con 3,01 unidades por hora. Las diferencias entre estos tres modelos fueron reducidas en términos absolutos, por lo que el orden obtenido a partir de los promedios no debe interpretarse como una superioridad amplia de un algoritmo sobre los demás.

El análisis por local permitió observar con mayor detalle esta variabilidad. LightGBM obtuvo el menor MAE en cuatro de los diez locales, XGBoost en tres y SVR en los tres restantes. Esto muestra que el modelo con mejor desempeño dependió del punto de venta analizado y refuerza la necesidad de considerar las características particulares de cada serie al interpretar los resultados predictivos.

Las métricas complementarias también mostraron diferencias entre los algoritmos. XGBoost y LightGBM presentaron los mejores resultados promedio en RMSE y R^2 ,

mientras que SVR obtuvo los menores valores de MAPE. Sin embargo, esta última métrica se interpretó con cautela por su sensibilidad cuando la demanda observada toma valores bajos. Por esta razón, el MAE se mantuvo como criterio principal y las demás métricas se utilizaron para ampliar la interpretación del desempeño.

La evaluación realizada con datos posteriores para la combinación Cumbayá–Enrollado aportó otro elemento a esta comparación. En ese caso específico, SVR presentó el menor error, a pesar de que XGBoost y LightGBM habían obtenido los mejores resultados promedio en el conjunto de prueba. Este comportamiento muestra que el orden relativo de los modelos puede modificarse cuando cambian el local, el producto, el período o la granularidad de evaluación. Debido a que esta comprobación se realizó únicamente para una combinación local-producto, el resultado se interpreta como evidencia complementaria del caso analizado y no como un criterio para seleccionar SVR de manera general.

LightGBM, XGBoost y SVR conformaron las alternativas con mejor desempeño dentro del estudio, aunque con fortalezas diferentes. LightGBM obtuvo el menor MAE en un mayor número de locales; XGBoost presentó resultados promedio muy próximos y un comportamiento relativamente estable entre las particiones temporales; y SVR alcanzó errores competitivos, además de obtener el mejor desempeño en tres locales y en la validación temporal complementaria realizada.

Por esta razón, los resultados no sustentan la selección de un modelo universal para los diez locales, sino una estrategia de evaluación en la que la elección del algoritmo puede realizarse de acuerdo con el comportamiento observado en cada local. Esta conclusión es

consistente con la heterogeneidad identificada durante el análisis exploratorio y con las diferencias encontradas en el desempeño de los modelos por punto de venta.

La comparación desarrollada permite establecer que los modelos evaluados generan estimaciones cuantificables de la demanda horaria observada, aunque con niveles de desempeño diferentes según el contexto analizado. La interpretación de su utilidad para apoyar la planificación del horneado debe considerar tanto el error de predicción como la variabilidad que permanece sin ser explicada por los modelos, aspectos que se profundizan en las siguientes secciones.

4.3 Interpretación de resultados

Los resultados obtenidos muestran que los modelos evaluados lograron generar estimaciones de la demanda horaria observada con niveles de error cuantificables y relativamente próximos entre las alternativas con mejor desempeño. En el conjunto de prueba, XGBoost y LightGBM alcanzaron un MAE promedio de aproximadamente 3 unidades por hora, mientras que SVR presentó un valor muy cercano. Esto significa que, en promedio, la diferencia absoluta entre la demanda estimada y la cantidad efectivamente vendida fue cercana a tres unidades por observación horaria.

Este valor debe interpretarse dentro de la escala y características del problema analizado. El MAE proporciona una medida directa y comprensible del error, pero no indica por sí solo si una estimación es suficiente para una decisión operativa determinada. Una diferencia de tres unidades puede tener distinta importancia dependiendo del producto, el volumen habitual de ventas del local y el momento del día. Las estimaciones deben

utilizarse como una referencia cuantitativa para apoyar la planificación del horneado y no como una instrucción exacta sobre la cantidad de producto que debe hornearse.

El coeficiente de determinación presentó valores más moderados. XGBoost y LightGBM alcanzaron R^2 promedio cercanos a 0,31, lo que indica que los modelos lograron representar una parte de la variabilidad observada en la demanda, mientras que una proporción importante permaneció sin ser explicada. Este resultado no invalida las estimaciones obtenidas, pero sí establece un límite claro sobre su capacidad para reproducir todos los cambios que se presentan en las ventas horarias.

La variabilidad no explicada puede estar asociada a distintos factores. El modelado se construyó principalmente a partir del comportamiento histórico de las ventas y de variables temporales derivadas de los registros disponibles. No se incorporaron de manera sistemática variables externas como promociones, precios, condiciones climáticas, eventos específicos, disponibilidad de producto o situaciones operativas particulares de cada local. Además, la variable objetivo representa ventas observadas y no demanda potencial, por lo que una ausencia de venta puede obedecer tanto a una menor intención de compra como a restricciones de disponibilidad que no pueden distinguirse con la información utilizada.

Esta limitación es relevante porque muestra que la mejora del desempeño predictivo no depende únicamente de probar algoritmos adicionales o de realizar ajustes más complejos. Parte de la capacidad futura de los modelos podría estar relacionada con la incorporación de información que permita explicar cambios de demanda que actualmente no se encuentran representados en los datos históricos y de calendario.

El comportamiento de las métricas también confirma la necesidad de analizarlas de forma conjunta. El MAE permitió comparar los modelos en la misma unidad de la variable objetivo y se mantuvo como el criterio principal de evaluación. El RMSE complementó esta lectura al asignar un mayor peso a errores de mayor magnitud. Por su parte, el R^2 permitió observar cuánto de la variación de la demanda fue representada por los modelos.

El MAPE presentó valores considerablemente más altos que las demás métricas relativas al error. Este comportamiento se relaciona con la presencia de observaciones de baja demanda, en las que pequeñas diferencias absolutas pueden producir errores porcentuales elevados. El MAPE se mantuvo como indicador complementario y no como criterio principal de selección.

Los resultados por local aportaron otra dimensión importante para la interpretación. La ausencia de un único algoritmo ganador en los diez puntos de venta muestra que el desempeño predictivo varía según las características de cada serie. LightGBM, XGBoost y SVR obtuvieron los mejores resultados en diferentes locales, lo que coincide con la heterogeneidad identificada previamente en los perfiles horarios de demanda. Esto sugiere que la utilización de un mismo algoritmo para toda la operación no necesariamente representa la mejor alternativa para todos los puntos de venta.

La validación temporal adicional realizada para Cumbayá–Enrollado reforzó esta lectura. SVR obtuvo el menor error en ese caso específico, a pesar de que XGBoost y LightGBM presentaron los mejores resultados promedio en el conjunto de prueba. Este resultado evidencia que el comportamiento relativo de los modelos puede modificarse al evaluar otros períodos, productos o niveles de agregación. Debido al alcance limitado de esta comprobación, el resultado no se extiende a las demás combinaciones local-producto.

En relación con las devoluciones, su análisis permitió complementar la interpretación de las ventas sin incorporarlas directamente como medida de la demanda horaria. Las diferencias observadas entre locales muestran que el producto registrado como no vendido presenta un comportamiento propio y no mantiene una relación uniforme con el volumen de ventas. Dado que los registros disponibles no permiten reconstruir de manera completa cuánto producto fue horneado ni su disponibilidad por hora, las devoluciones no se utilizaron para afirmar que los modelos reducirán directamente el producto no vendido.

En términos generales, los resultados muestran que el enfoque desarrollado permite estimar la demanda horaria observada por local y producto y proporciona una referencia cuantitativa que puede apoyar la planificación del horneo. Sin embargo, las predicciones deben utilizarse considerando el margen de error, las diferencias entre locales y la proporción de variabilidad que permanece sin explicar. La investigación demuestra el desempeño predictivo del enfoque con datos históricos, pero no evalúa todavía el efecto que su utilización tendría sobre indicadores operativos como disponibilidad de producto, reducción de devoluciones o cumplimiento de una planificación de horneo.

4.4 Discusión

La discusión se orienta a interpretar los resultados obtenidos a la luz de los estudios revisados y de las características propias del problema analizado. Además del contraste con la literatura, se revisan los hallazgos que requieren una explicación adicional, las implicaciones prácticas y metodológicas de los resultados, las principales limitaciones del estudio y el alcance de la hipótesis planteada.

4.4.1 Contraste con la literatura

Las comparaciones con estudios previos se interpretan de manera contextual y no como un benchmark directo, debido a las diferencias existentes en la granularidad temporal, el dominio de aplicación, las variables utilizadas, la escala de la variable objetivo y los protocolos de evaluación. Por esta razón, los valores de MAE, RMSE, MAPE o R^2 reportados en otros trabajos no pueden utilizarse de manera aislada para determinar si los resultados obtenidos en esta investigación son mejores o peores.

Uno de los antecedentes más cercanos al contexto analizado es el trabajo de Huber y Stuckenschmidt (2020), quienes estudiaron la predicción de demanda diaria en una cadena de panaderías y encontraron ventajas de los enfoques de aprendizaje automático frente a determinados métodos tradicionales dentro de su contexto de evaluación. En la presente investigación también se observó un mejor comportamiento general de los modelos no lineales frente a Ridge. Sin embargo, existe una diferencia metodológica importante: Huber y Stuckenschmidt trabajan con demanda diaria a nivel de tienda y categoría, mientras que este estudio estima demanda horaria por producto y local. Esta mayor desagregación responde a la necesidad de aproximar la predicción a los diferentes momentos de la jornada en los que se toman decisiones de horneado.

Los resultados también guardan relación con lo planteado por Rodrigues et al. (2024), quienes evaluaron modelos de aprendizaje automático para predicción de demanda de corto plazo en servicios de alimentación y analizaron su posible relación con la reducción del desperdicio. En ambos estudios, la predicción se plantea como una herramienta que puede aportar información para decisiones relacionadas con productos perecibles. Sin embargo, Rodrigues et al. trabajan con datos diarios y un contexto de servicios de catering, por lo que sus resultados y los porcentajes de reducción de

desperdicio reportados no pueden trasladarse directamente a una operación de panadería con demanda horaria y horneado en los puntos de venta. En esta investigación tampoco se midió el efecto operativo de las predicciones sobre las devoluciones, por lo que ese resultado debe mantenerse como una posibilidad de aplicación futura y no como un efecto demostrado.

Por su parte, Seyam et al. (2025) reportaron un alto desempeño mediante un modelo de *stacking ensemble* que combinó distintos algoritmos de aprendizaje automático. Los valores obtenidos en ese estudio son numéricamente superiores a los alcanzados en la presente investigación, pero corresponden a un conjunto de datos, una escala y un protocolo experimental diferentes. Esta diferencia refuerza la necesidad de evitar comparaciones directas entre métricas obtenidas sobre problemas con distinta granularidad y estructura. En este estudio, además, los resultados mostraron que XGBoost, LightGBM y SVR presentaron desempeños muy próximos y que el modelo con menor error varió entre locales, lo que no sustenta la existencia de un algoritmo universalmente superior para el problema analizado.

La heterogeneidad encontrada entre los locales también puede relacionarse con los planteamientos de Jauhar et al. (2025), quienes muestran que diferenciar grupos con características distintas puede modificar el desempeño del pronóstico. Aunque su estudio utiliza una estrategia de segmentación y unidades de análisis diferentes, el principio resulta pertinente para interpretar los resultados obtenidos en esta investigación. Los perfiles horarios mostraron diferencias entre locales y, posteriormente, LightGBM, XGBoost y SVR obtuvieron el menor MAE en distintos puntos de venta. Esto respalda la decisión

metodológica de evaluar el comportamiento de los modelos de forma diferenciada, en lugar de asumir que todos los locales presentan una misma dinámica de demanda.

Desde una perspectiva temporal, Lim et al. (2021) destacan la importancia de representar relaciones históricas y diferentes tipos de covariables en problemas de pronóstico multihorizonte. Aunque el modelo Temporal Fusion Transformer desarrollado por los autores no fue incluido entre los algoritmos evaluados en esta investigación, su planteamiento es consistente con la utilización de variables históricas y temporales para representar la evolución de la demanda. Al mismo tiempo, las diferencias de desempeño encontradas y los valores moderados de R^2 obtenidos en este estudio sugieren que la incorporación futura de información contextual adicional podría aportar nueva capacidad explicativa, especialmente en eventos que no pueden ser representados únicamente mediante el comportamiento histórico de las ventas.

Los resultados también pueden interpretarse a la luz del trabajo de Fries y Ludwig (2024), quienes analizaron los desafíos sociotécnicos asociados al uso de pronósticos de ventas en una panadería. Su estudio destaca que el valor de una predicción no depende exclusivamente de la precisión estadística, sino también de la manera en que la información puede incorporarse a las decisiones operativas. Esta perspectiva resulta especialmente pertinente para la presente investigación, ya que un MAE cercano a tres unidades por hora constituye una medida del desempeño predictivo, pero no determina por sí mismo cómo debería ejecutarse el horneado en cada local. La utilización de las estimaciones requiere considerar el producto, la hora, el volumen habitual de demanda y las condiciones operativas del punto de venta.

La revisión sistemática desarrollada por Muth et al. (2025) señala que la selección de los métodos de predicción depende del contexto, la información disponible y las características del fenómeno analizado. Los resultados de esta investigación coinciden con esa perspectiva, debido a que no se identificó un único modelo con el mejor desempeño en todas las situaciones. LightGBM obtuvo el menor MAE en cuatro locales, XGBoost en tres y SVR en tres, mientras que este último también presentó el menor error en la validación temporal complementaria realizada para Cumbayá–Enrollado. Estos resultados refuerzan la importancia de analizar el desempeño de manera contextual y no únicamente a partir de un promedio general.

El contraste con la literatura permite identificar tres coincidencias principales: la utilidad del aprendizaje automático para representar patrones de demanda, la importancia de incorporar información temporal e histórica y la necesidad de considerar las particularidades del contexto de aplicación. Al mismo tiempo, esta investigación se diferencia por trabajar con demanda horaria observada por producto y local, comparar el desempeño de varios algoritmos entre distintos puntos de venta y mantener las devoluciones de producto no vendido como una fuente analítica complementaria, sin incorporarlas como variable objetivo ni como medida directa del horneado. Estas características delimitan el aporte del estudio dentro de los antecedentes revisados, sin plantear que los resultados sean directamente superiores o inferiores a los reportados en otros contextos.

4.4.2 Explicación de los resultados obtenidos

El comportamiento de los modelos puede explicarse a partir de las características de la información utilizada y de la variabilidad propia de la demanda observada. Los

algoritmos trabajaron principalmente con registros históricos de ventas, variables de calendario y características derivadas del comportamiento previo de las series. Esta información permitió identificar patrones recurrentes, aunque no recoge todos los factores que pueden modificar las ventas en una hora determinada.

Las diferencias encontradas entre locales son consistentes con los patrones identificados durante el análisis exploratorio. Los perfiles horarios mostraron que la actividad comercial no se distribuye de la misma manera en todos los puntos de venta. Por ello, resulta razonable que un algoritmo que funciona mejor en un local no necesariamente mantenga la misma posición en otro. LightGBM, XGBoost y SVR obtuvieron el menor MAE en distintos locales, lo que refleja que las características de cada serie influyen en el desempeño relativo de los modelos.

El comportamiento de XGBoost y LightGBM puede relacionarse con su capacidad para representar relaciones no lineales entre las variables históricas y temporales incorporadas en el modelado. Ambos algoritmos alcanzaron los menores valores promedio de MAE en el conjunto de prueba y presentaron resultados cercanos en RMSE y R^2 . Sin embargo, la diferencia entre ellos fue reducida, por lo que estos resultados no permiten establecer una ventaja amplia de uno sobre el otro. Más que identificar un ganador único, la evidencia muestra que ambos métodos fueron competitivos dentro del conjunto analizado.

SVR presentó un comportamiento diferente. Aunque su R^2 promedio fue inferior al obtenido por XGBoost y LightGBM, su MAE se mantuvo muy próximo y fue el modelo con menor error en tres locales. También obtuvo el mejor resultado en la validación temporal complementaria de Cumbayá–Enrollado. Esto muestra que un modelo puede mantener errores absolutos relativamente bajos sin representar en la misma medida las

variaciones de la serie. La lectura conjunta de MAE y R^2 resulta necesaria para evitar que una única métrica determine la interpretación del desempeño.

Random Forest presentó una diferencia más marcada entre entrenamiento y los períodos posteriores. Durante el entrenamiento alcanzó valores de error menores y un R^2 mayor que en validación y prueba. Este comportamiento sugiere una mayor adaptación a las características de los datos utilizados durante el ajuste. Aun así, sus resultados en prueba se mantuvieron próximos a los de los modelos con mejor desempeño, aunque sin alcanzar el menor MAE en ninguno de los diez locales.

Ridge mostró un comportamiento relativamente estable entre las particiones, pero con mayores errores en el conjunto de prueba. Esta diferencia frente a los modelos no lineales puede relacionarse con la naturaleza del problema, donde la demanda horaria depende de interacciones entre variables temporales, comportamiento reciente y características de cada local que no necesariamente siguen relaciones lineales simples. El resultado no implica que los modelos lineales carezcan de utilidad, sino que, bajo la configuración evaluada, Ridge presentó una capacidad predictiva menor que las alternativas no lineales.

Holt-Winters debe interpretarse de manera distinta. El modelo se trabajó con series de granularidad diaria y posteriormente sus estimaciones se representaron en escala horaria. Por esta razón, no es directamente comparable en granularidad y objetivo con los cinco modelos entrenados sobre observaciones horarias. Su inclusión permitió mantener una referencia basada en un método clásico de series temporales, pero las diferencias en MAE y RMSE no deben utilizarse como evidencia de inferioridad frente a los modelos de aprendizaje automático.

La información disponible para el modelado también ayuda a explicar los resultados obtenidos. Las ventas pueden verse afectadas por promociones, cambios de precio, eventos, clima, disponibilidad de producto y circunstancias particulares de cada local. Al no disponer de estas variables de manera estructurada para todo el período, los modelos tuvieron que estimar la demanda principalmente a partir de su comportamiento histórico y temporal. Esto puede explicar parte de la variabilidad que permaneció sin representar en los valores de R^2 .

La disponibilidad de producto merece una consideración particular. La variable objetivo se construyó a partir de unidades efectivamente vendidas, por lo que representa demanda horaria observada. Cuando una venta no se registra, los datos no permiten distinguir si existió una menor intención de compra o si el producto no estuvo disponible para ser adquirido. Esta característica de la fuente limita la posibilidad de interpretar las predicciones como una medición completa de la demanda potencial.

Las devoluciones tampoco modificaron esta definición. Su análisis permitió observar diferencias entre locales y aportar información sobre producto registrado como no vendido, pero su granularidad y momento de registro son distintos a los de las ventas horarias. Por esta razón, fueron tratadas como una fuente complementaria y no como una medida directa de la cantidad horneada ni como un predictor incorporado de manera permanente en los modelos finales.

Los resultados pueden interpretarse a partir de tres elementos principales: la existencia de patrones históricos y temporales que los modelos pudieron aprovechar, la heterogeneidad entre locales y productos y la ausencia de variables externas que podrían explicar parte de las variaciones de corto plazo. La interacción de estos factores ayuda a

comprender por qué los modelos alcanzaron errores absolutos relativamente próximos, pero mostraron diferencias en R^2 y en el desempeño obtenido para cada local.

4.4.3 Hallazgos inesperados

Durante la evaluación se identificaron algunos resultados que se apartaron del comportamiento observado en los promedios generales y que aportan elementos adicionales para interpretar el problema predictivo.

Uno de estos hallazgos corresponde al desempeño de SVR en la validación temporal complementaria de Cumbayá–Enrollado. En el conjunto de prueba general, XGBoost y LightGBM presentaron los menores valores promedio de MAE, mientras que SVR se ubicó muy cerca de estos modelos. Sin embargo, al utilizar información posterior para la combinación Cumbayá–Enrollado, SVR obtuvo el menor MAE, RMSE y MAPE entre los cinco algoritmos evaluados. Este resultado muestra que el orden relativo de los modelos puede modificarse cuando cambian el período, el producto, el local o la granularidad de evaluación. Debido a que la comprobación se realizó para una única combinación local-producto, el hallazgo no permite concluir que SVR mantendría esa ventaja en los demás casos.

También resultó relevante la diferencia observada entre MAE y R^2 en algunos locales. Marina constituye el caso más evidente: SVR presentó el menor MAE del local, con 2,87 unidades por hora, pero un R^2 de 0,023. Portugal y Gaspar mostraron un comportamiento similar, aunque menos marcado. Esto evidencia que un error absoluto reducido no implica necesariamente que el modelo reproduzca adecuadamente la variación de la serie. El hallazgo refuerza la decisión de interpretar conjuntamente las métricas y

evita considerar el MAE como una medida suficiente para describir todo el desempeño predictivo.

Otro resultado llamativo fue que en algunos modelos el MAE de validación o prueba fue menor que el registrado durante el entrenamiento, mientras que el R^2 disminuyó. En XGBoost, por ejemplo, el MAE pasó de 3,14 en entrenamiento a 3,00 en prueba, mientras que el R^2 se redujo de 0,362 a 0,307. Este comportamiento no implica que el modelo haya funcionado mejor sobre información no observada. Las particiones corresponden a períodos cronológicos distintos y pueden presentar diferencias en nivel y dispersión de la demanda. Una serie con menor variabilidad puede generar un MAE más bajo y, al mismo tiempo, un R^2 menor, debido a que ambas métricas representan dimensiones distintas del error.

La comparación entre algoritmos mostró además que las diferencias promedio entre los modelos con mejor desempeño fueron menores de lo que podría esperarse al considerar métodos con estructuras distintas. XGBoost y LightGBM registraron un MAE promedio de 3,00 unidades por hora y SVR de 3,01. A pesar de esta proximidad, el modelo con menor MAE cambió entre los diez locales. Esto sugiere que, para el problema analizado, las características propias de cada serie pueden tener tanta relevancia como la diferencia global entre los algoritmos evaluados.

Los hallazgos identificados muestran que el resultado predictivo no depende únicamente del algoritmo utilizado. Las diferencias entre locales, períodos y métricas modifican la forma en que deben interpretarse los resultados. La evaluación realizada favorece así una lectura contextual, en la que el MAE constituye el criterio principal, pero debe analizarse junto con las métricas complementarias y las características de cada serie.

4.4.4 Implicaciones

Los resultados obtenidos muestran que la estimación de la demanda horaria puede aportar información adicional para apoyar la planificación del horneo en los locales analizados. Al disponer de una predicción diferenciada por local, producto y hora, la decisión deja de depender exclusivamente de promedios diarios o de criterios generales y puede incorporar una referencia cuantitativa sobre el comportamiento esperado de las ventas.

La utilidad de estas estimaciones debe entenderse como apoyo a la decisión y no como una instrucción automática de producción. Un MAE cercano a tres unidades por hora indica que existe un margen de error que debe considerarse antes de trasladar una predicción directamente a una cantidad de horneo. La magnitud de ese error puede tener diferente importancia según el producto, el volumen habitual de ventas y el momento de la jornada. Por ello, una aplicación operativa requeriría definir reglas que traduzcan las estimaciones en cantidades o rangos de horneo adecuados para cada contexto.

Los resultados por local también tienen una implicación relevante para el diseño de una eventual solución operativa. La ausencia de un algoritmo ganador en los diez puntos de venta indica que no necesariamente sería conveniente adoptar un único modelo para toda la cadena. Una alternativa sería mantener un criterio de selección por local, utilizando el algoritmo que presente el comportamiento más adecuado para cada caso y revisando periódicamente su desempeño cuando se incorpore nueva información.

Desde el punto de vista metodológico, el estudio muestra la importancia de analizar las métricas de manera conjunta. El MAE permitió establecer un criterio principal de

comparación en unidades directamente interpretables, mientras que RMSE y R^2 aportaron información adicional sobre los errores de mayor magnitud y la capacidad de los modelos para representar la variabilidad de la demanda. Esta combinación evita seleccionar un modelo únicamente porque obtiene una ventaja mínima en una sola métrica.

La investigación también evidencia la relevancia de trabajar con información desagregada. Los perfiles exploratorios y los resultados predictivos mostraron diferencias entre locales y horarios que podrían quedar ocultas en un análisis exclusivamente diario o agregado. Para una operación en la que el producto puede terminarse u hornearse varias veces durante la jornada, esta granularidad permite acercar la estimación al momento en que se toman las decisiones.

Las devoluciones aportan una perspectiva complementaria para interpretar la operación, pero los datos disponibles no permiten establecer que el uso de los modelos reduzca directamente el producto no vendido. Tampoco se evaluó mediante un piloto si las predicciones mejoran la disponibilidad de producto o modifican otros indicadores operativos. Estos efectos requieren una fase posterior de aplicación en condiciones reales, con indicadores definidos antes y después de utilizar las estimaciones.

Como complemento al trabajo predictivo, puede presentarse en el Anexo 1 el prototipo académico desarrollado para consultar estimaciones por local, producto y período. Su función es mostrar una forma posible de visualizar los resultados generados por los modelos. Este componente tiene un carácter demostrativo y no constituye evidencia adicional sobre la precisión predictiva ni sobre el impacto operativo de la propuesta.

La aplicación futura del enfoque podría plantearse de manera gradual. Una primera etapa podría utilizar las predicciones como referencia paralela a la planificación actual y registrar las decisiones tomadas, la disponibilidad de producto y las devoluciones. Esta información permitiría posteriormente evaluar si el uso de las estimaciones produce cambios medibles en la operación y ajustar las reglas necesarias antes de considerar una integración más amplia.

Las implicaciones del estudio se concentran, por tanto, en demostrar que es posible generar estimaciones horarias diferenciadas a partir de los registros históricos disponibles y utilizarlas como una fuente adicional de información para la planificación. El paso desde la predicción hacia una mejora operacional comprobada requiere una evaluación posterior en condiciones reales, aspecto que queda fuera del alcance de esta investigación.

4.4.5 Discusión de las limitaciones

La interpretación de los resultados debe considerar varias limitaciones relacionadas con la información disponible, la construcción del conjunto de modelado y el alcance de las evaluaciones realizadas. Estas limitaciones no invalidan los resultados obtenidos, pero sí delimitan las conclusiones que pueden derivarse de ellos.

Una primera limitación corresponde a la naturaleza de la variable objetivo. El estudio utiliza las unidades efectivamente vendidas como representación de la demanda horaria observada. Los registros disponibles no permiten conocer si, en una hora determinada, la ausencia o disminución de ventas se debió a una menor intención de compra o a la falta de disponibilidad del producto. Por esta razón, las predicciones

representan el comportamiento observado en las ventas y no la totalidad de la demanda potencial.

La información utilizada para el modelado se concentró principalmente en registros históricos de ventas y variables temporales derivadas de estos datos. No se dispuso de manera estructurada y continua de información sobre promociones, variaciones de precio, condiciones climáticas, eventos, disponibilidad horaria de producto u otras situaciones operativas que podrían afectar las ventas. La literatura revisada muestra que este tipo de información contextual puede aportar capacidad explicativa adicional en problemas de pronóstico, y que su utilidad depende del contexto y de las variables disponibles para cada aplicación (Lim et al., 2021; Fries & Ludwig, 2024; Muth et al., 2025).

Otra limitación se relaciona con la construcción de la serie horaria. Las horas sin registros de venta no se incorporaron como observaciones con valor cero, debido a que los datos disponibles no permiten diferenciar con certeza entre una hora con demanda nula y una hora en la que el producto no estuvo disponible. Esta decisión evita introducir ceros cuya interpretación sería incierta, pero también implica que los rezagos utilizados por los modelos representan observaciones previas dentro de la secuencia disponible y no necesariamente intervalos exactos de 24 horas o siete días calendario. La literatura sobre representación temporal señala precisamente que, cuando las observaciones no conforman una secuencia cronológica continua, los rezagos deben interpretarse como posiciones anteriores dentro de la secuencia y no como desplazamientos temporales estrictos (Hyndman & Athanasopoulos, 2021; Lim et al., 2021).

El entrenamiento de SVR se realizó mediante una submuestra del conjunto disponible debido al costo computacional asociado al algoritmo. Aunque el modelo alcanzó

resultados competitivos y obtuvo el menor MAE en tres locales, esta estrategia implica que su entrenamiento no utilizó exactamente el mismo volumen de información que los demás métodos. La comparación permite valorar su desempeño bajo la configuración aplicada en el estudio, pero no determinar cómo habría respondido si hubiera sido entrenado con la totalidad de las observaciones.

La validación temporal adicional constituye otra limitación del alcance. Esta evaluación se realizó exclusivamente para la combinación Cumbayá–Enrollado con datos posteriores al conjunto de prueba. El resultado permitió observar el comportamiento de los modelos en un período posterior para ese caso específico, pero no permite extender la misma conclusión a los demás locales y productos incluidos en la investigación.

Las devoluciones también presentan restricciones para su interpretación. La información corresponde a registros de producto no vendido efectuados principalmente al cierre de la jornada y no permite reconstruir de manera horaria la cantidad horneada, la disponibilidad de producto ni las decisiones de reposición realizadas durante el día. Por este motivo, las devoluciones fueron utilizadas como información complementaria y no permiten establecer de forma directa la relación entre las predicciones y una eventual reducción del producto no vendido.

El estudio tampoco incluyó una aplicación piloto de las predicciones dentro de la operación diaria. Los resultados permiten evaluar el desempeño predictivo, pero no medir todavía efectos sobre variables operativas como disponibilidad de producto, cumplimiento de la planificación del horneo o cambios en las devoluciones. Esta distinción es relevante porque la literatura sobre aplicaciones de pronóstico en panaderías señala que la utilidad de

una predicción no depende únicamente de sus métricas estadísticas, sino también de la forma en que se incorpora a los procesos y decisiones operativas (Fries & Ludwig, 2024).

A partir de estas limitaciones se identifican varias posibilidades para continuar el trabajo. La incorporación de variables externas, especialmente promociones, precios, condiciones climáticas, eventos y registros de disponibilidad de producto, permitiría evaluar si aumenta la capacidad explicativa de los modelos. Este tipo de ampliación es consistente con trabajos que destacan la importancia de combinar información histórica con covariables conocidas o contextuales y con revisiones que señalan la integración de información adicional como una línea relevante para mejorar los modelos de demanda (Lim et al., 2021; Muth et al., 2025).

También sería conveniente ampliar la validación temporal a un mayor número de combinaciones local-producto y analizar esquemas de actualización periódica conforme se incorporen nuevos datos. Esta ampliación permitiría comprobar si los patrones observados en el período de estudio se mantienen ante cambios en la demanda y si la selección del modelo por local conserva su desempeño con información posterior.

Una futura disponibilidad de información horaria sobre inventario, horneado y disponibilidad efectiva del producto permitiría diferenciar con mayor precisión entre ausencia de venta y ausencia de demanda. Bajo esas condiciones podría evaluarse la construcción de una serie temporal continua e investigar nuevas variables históricas calculadas sobre intervalos cronológicos exactos, manteniendo coherencia entre la granularidad de los datos y la unidad temporal de decisión. La importancia de esta correspondencia entre frecuencia, representación temporal y objetivo del pronóstico ha sido

señalada en la literatura de series temporales y aprendizaje automático (Hyndman & Athanasopoulos, 2021; Lim et al., 2021).

El desarrollo posterior también podría considerar otros enfoques de modelado, incluidos métodos diseñados para series temporales multivariadas y arquitecturas de mayor complejidad. Lim et al. (2021), por ejemplo, proponen modelos capaces de combinar información histórica, covariables conocidas a futuro y variables estáticas para problemas de pronóstico multihorizonte. Su incorporación en futuras investigaciones debería realizarse bajo el mismo esquema temporal y con métricas equivalentes, de modo que la comparación mantenga coherencia metodológica.

La selección de nuevas alternativas tampoco debería asumir que una arquitectura más compleja producirá necesariamente mejores resultados. La literatura revisada señala que el desempeño depende de la estructura de los datos, la granularidad, las variables disponibles y las características del problema, por lo que la comparación empírica entre modelos continúa siendo necesaria (Muth et al., 2025).

Una fase piloto permitiría determinar si las estimaciones generan mejoras medibles cuando se incorporan al proceso de planificación. Para ello sería necesario definir previamente indicadores operativos y comparar su comportamiento antes y después de utilizar las predicciones. Esta fase permitiría pasar de la evaluación exclusivamente predictiva a una evaluación de utilidad operativa, manteniendo separadas ambas dimensiones y evitando atribuir al modelo efectos que todavía no han sido comprobados.

4.4.6 Grado de validación de la hipótesis

La hipótesis planteada estableció que la aplicación de modelos predictivos basados en registros históricos de ventas POS permitiría estimar la demanda horaria observada de productos de panadería por local y producto, como apoyo a la planificación del horneado en los puntos de venta seleccionados. Los resultados obtenidos permiten valorar esta hipótesis a partir del desempeño alcanzado por los modelos y de las condiciones bajo las cuales fueron evaluados.

La evaluación realizada sobre información cronológicamente posterior al entrenamiento mostró que los modelos generaron estimaciones cuantificables de la demanda horaria observada. En el conjunto de prueba, XGBoost y LightGBM alcanzaron un MAE promedio cercano a 3 unidades por hora, mientras que SVR presentó un resultado muy próximo. Estos valores evidencian que los registros históricos de ventas, combinados con variables temporales y características derivadas de la evolución previa de las series, contienen información útil para estimar el comportamiento posterior de las ventas.

Los resultados también mostraron que la capacidad predictiva no fue uniforme entre locales ni entre algoritmos. LightGBM obtuvo el menor MAE en cuatro locales, XGBoost en tres y SVR en tres, sin que se identificara un único modelo superior para todos los puntos de venta. Los valores de R^2 , cercanos a 0,31 en los modelos con mejor desempeño promedio, indican además que una parte de la variabilidad de la demanda permanece sin ser explicada. Este resultado delimita el alcance de la hipótesis, ya que demuestra capacidad de estimación, pero no implica una representación completa de todos los factores que modifican la demanda horaria.

La validación temporal complementaria realizada con información posterior para Cumbayá–Enrollado aportó evidencia adicional para esa combinación específica. En este

caso, SVR obtuvo el menor error entre los modelos evaluados. Debido a que esta comprobación se realizó sobre un único local y producto, su resultado no puede extenderse al conjunto completo de combinaciones analizadas, pero sí muestra que los modelos mantienen capacidad de generar estimaciones frente a información posterior en el caso evaluado.

La segunda parte de la hipótesis se relaciona con el uso de las estimaciones como apoyo a la planificación del horneo. La granularidad por local, producto y hora permite disponer de una referencia cuantitativa alineada con el momento en que pueden tomarse decisiones de horneo. Sin embargo, la investigación no incluyó un piloto operativo que permitiera medir el efecto de las predicciones sobre la ejecución de la planificación, la disponibilidad de producto o las devoluciones. Por esta razón, el estudio demuestra la viabilidad predictiva del enfoque y su potencial como fuente de información para la planificación, pero no permite afirmar todavía que su aplicación genere mejoras operativas medibles.

A partir de la evidencia obtenida, la hipótesis se considera respaldada dentro del alcance predictivo definido para la investigación. Los modelos desarrollados permitieron estimar la demanda horaria observada utilizando registros históricos de ventas y mostraron resultados diferenciados por local y producto. Este respaldo debe interpretarse considerando las limitaciones identificadas, especialmente la ausencia de variables externas y de información horaria sobre disponibilidad de producto, la variabilidad no explicada por los modelos y el alcance restringido de la validación temporal adicional.

La investigación permite, por tanto, sostener que el enfoque predictivo desarrollado constituye una base cuantitativa para apoyar la planificación del horneo, sin extender esta

conclusión a efectos operativos que no fueron evaluados. La comprobación de mejoras sobre la operación requeriría una etapa posterior de aplicación y seguimiento en condiciones reales.

Capítulo 5

5. Conclusiones y Recomendaciones

5.1 Conclusiones

La investigación permitió desarrollar y evaluar un enfoque predictivo para estimar la demanda horaria observada de productos de panadería por local y producto, utilizando como fuente principal los registros históricos de ventas del sistema POS. Los resultados obtenidos muestran que esta información puede transformarse en estimaciones cuantitativas diferenciadas temporalmente y constituir una base para complementar la planificación del horneado en los puntos de venta seleccionados. Este resultado responde al objetivo general del estudio, dentro de un alcance orientado al apoyo a la decisión y no a la automatización de la producción.

En relación con el primer objetivo específico, fue posible construir una base analítica horaria mediante la depuración, consolidación y transformación de la información disponible. A partir de los registros transaccionales seleccionados se obtuvo un conjunto final de 498.750 observaciones para el modelado, correspondiente a diez locales y quince productos de panadería. La organización cronológica de la información, la construcción de variables temporales e históricas y la separación de los conjuntos de entrenamiento, validación y prueba permitieron disponer de una estructura reproducible para el desarrollo y evaluación de los modelos.

Respecto del segundo objetivo específico, el análisis exploratorio evidenció que la demanda horaria observada presenta diferencias según el momento de la jornada, el local y el producto. Los perfiles horarios mostraron que la distribución de las ventas no es

homogénea entre los puntos de venta, lo que confirmó la necesidad de conservar la identificación de cada local y considerar tanto patrones particulares como compartidos durante el modelado. Las devoluciones de producto no vendido aportaron información complementaria para contextualizar estas diferencias, pero su distinta granularidad y el momento en que se registran impiden interpretarlas como una medición directa de la demanda, del volumen horneado, de la disponibilidad o de la merma. Por esta razón, se mantuvieron fuera de la variable objetivo y de las características predictoras.

En cuanto al tercer objetivo específico, se desarrollaron modelos predictivos utilizando cinco algoritmos de aprendizaje automático (Ridge Regression, Random Forest, XGBoost, LightGBM y Support Vector Regression) y un método estadístico de series temporales, Holt-Winters. La estrategia consideró modelos específicos para cada uno de los diez locales y modelos globales entrenados con información consolidada, alcanzando 65 evaluaciones dentro del diseño experimental. Esta comparación permitió analizar el problema mediante enfoques con diferentes mecanismos de aprendizaje y evaluar su comportamiento bajo las condiciones del conjunto de datos utilizado.

En relación con el cuarto objetivo específico, los resultados del conjunto de prueba mostraron que XGBoost y LightGBM alcanzaron el mejor desempeño agregado entre los modelos horarios, con valores de MAE cercanos a 3 unidades por hora y resultados prácticamente equivalentes en las métricas principales. Sin embargo, las diferencias entre algoritmos fueron reducidas y el análisis desagregado confirmó que no existió un modelo universalmente superior para todos los locales. LightGBM obtuvo el menor MAE en cuatro locales, XGBoost en tres y SVR en tres. Este resultado evidencia que la selección de una

alternativa predictiva no debe sustentarse únicamente en el promedio general, sino también en el comportamiento específico observado en cada local y en la métrica priorizada.

El valor de MAE cercano a tres unidades por hora proporciona una referencia directa sobre la magnitud promedio del error, pero debe interpretarse dentro de las características de cada producto y local. Una diferencia de tres unidades puede tener distinta relevancia dependiendo del volumen habitual de ventas y del momento de la jornada. De igual manera, los valores moderados de R^2 muestran que los modelos representan una parte de la variabilidad presente en la demanda observada. Esto no invalida las estimaciones obtenidas, pero establece un límite sobre la capacidad de los modelos para representar toda la variabilidad observada y deja espacio para evaluar información adicional, como cambios operativos, promociones, condiciones climáticas, eventos especiales y situaciones relacionadas con la disponibilidad del producto.

La validación realizada con información posterior para la combinación Cumbayá–Enrollado mostró que los modelos conservaron capacidad para generar estimaciones sobre observaciones no utilizadas en el período inicial de entrenamiento y que, para ese caso específico, SVR presentó el menor error. Este resultado constituye una evidencia complementaria sobre el comportamiento temporal del enfoque, pero no permite afirmar que SVR sea el mejor algoritmo para todos los productos y locales ni generalizar los resultados de esta validación al conjunto completo de las combinaciones analizadas.

En relación con el quinto objetivo específico, los resultados permiten establecer que las estimaciones generadas pueden utilizarse como una referencia cuantitativa para complementar la planificación del horneado. La granularidad por local, producto y hora permite aproximar la predicción al momento en que se toman decisiones operativas durante

la jornada. Sin embargo, las estimaciones no representan instrucciones exactas sobre la cantidad que debe hornearse, debido a que el modelo utiliza ventas históricas observadas y no dispone de información completa sobre disponibilidad horaria, eventos de horneado o demanda potencial no atendida. La interpretación de las predicciones debe complementarse con el conocimiento operativo del personal y con las condiciones particulares de cada local.

La hipótesis planteada se considera respaldada dentro del alcance predictivo de la investigación. Los registros históricos de ventas POS permitieron construir modelos capaces de estimar la demanda horaria observada por local y producto y evaluar su desempeño mediante métricas cuantificables. Este respaldo corresponde a la capacidad predictiva demostrada con los datos disponibles y al potencial de las estimaciones como información de apoyo para la planificación. No se evaluó mediante un piloto operativo si el uso de las predicciones produce una reducción de devoluciones, un incremento de ventas, una mejora en la disponibilidad o cambios en las cantidades efectivamente horneadas; dichos efectos no forman parte de las conclusiones demostradas por el estudio.

El aporte del trabajo se encuentra en la construcción de una estrategia aplicada de predicción con granularidad horaria y diferenciación por local y producto, en la comparación entre modelos específicos y globales y en la separación metodológica entre las ventas utilizadas para representar la demanda observada y las devoluciones empleadas únicamente como información complementaria. Los resultados muestran que la heterogeneidad entre puntos de venta constituye un elemento relevante para la selección de modelos y que una futura solución de apoyo a la planificación debería considerar tanto la capacidad predictiva como las limitaciones de la información disponible.

5.2 Recomendaciones

Se recomienda establecer un proceso periódico de actualización de la información y reentrenamiento de los modelos. Debido a que las estimaciones se apoyan en patrones históricos, rezagos y estadísticas móviles, su desempeño puede verse afectado por cambios en productos, horarios, condiciones operativas o comportamiento de los clientes. La frecuencia de actualización debería definirse a partir del seguimiento de las métricas de desempeño y no mediante un intervalo fijo establecido de manera previa.

Como mejora prioritaria de la información disponible, se recomienda incorporar registros que permitan distinguir con mayor precisión las horas sin ventas. Para ello sería necesario disponer de información sobre la apertura del local, la disponibilidad efectiva de cada producto y, de ser posible, los eventos de preparación u horneado realizados durante la jornada. Con esta información podría construirse una secuencia horaria continua por fecha, local y producto, asignando valores iguales a cero únicamente cuando se confirme que el producto estuvo disponible y no registró ventas. Esto permitiría interpretar los rezagos mediante intervalos cronológicos más precisos y reducir la incertidumbre asociada con las horas actualmente ausentes del conjunto de datos.

También se recomienda ampliar el conjunto de variables explicativas mediante información que aporte contexto adicional al comportamiento de la demanda. Entre las variables que podrían evaluarse se encuentran promociones, cambios de precio, feriados, eventos locales, condiciones climáticas, modificaciones en los horarios de atención y disponibilidad del producto. Su incorporación permitiría analizar si parte de

la variabilidad que actualmente no es representada por los modelos puede explicarse mediante factores adicionales a los registros históricos y las variables de calendario.

Se recomienda ampliar la validación temporal a diferentes combinaciones de locales y productos. La evaluación posterior realizada para Cumbayá–Enrollado corresponde a un caso específico y no permite generalizar el comportamiento observado al conjunto completo del estudio. Replicar este procedimiento en productos y locales con distintos niveles y patrones de venta permitiría evaluar con mayor profundidad la estabilidad temporal de los modelos y determinar si las diferencias entre algoritmos se mantienen en períodos posteriores.

En futuras investigaciones se recomienda aplicar procedimientos sistemáticos de optimización de hiperparámetros, manteniendo esquemas de validación que respeten la temporalidad de los datos. La investigación utilizó configuraciones previamente definidas y no realizó una búsqueda exhaustiva mediante técnicas como *grid search*, *random search* u optimización bayesiana. Este tipo de procedimientos permitiría analizar si es posible mejorar el desempeño de los algoritmos y realizar comparaciones bajo configuraciones optimizadas de manera sistemática.

También resulta pertinente profundizar la evaluación de SVR bajo condiciones de entrenamiento comparables con las utilizadas para los demás algoritmos. En este estudio, su entrenamiento se realizó mediante submuestreo debido a restricciones computacionales, pese a lo cual presentó resultados competitivos y obtuvo el menor error en la validación posterior del caso analizado. Futuras pruebas podrían considerar diferentes tamaños de muestra, mayor capacidad computacional o esquemas

equivalentes de submuestreo para analizar con mayor precisión el efecto de esta decisión sobre su desempeño.

Como línea adicional de investigación, podrían evaluarse modelos de mayor complejidad, incluidas arquitecturas de aprendizaje profundo orientadas a series temporales. Su incorporación debería realizarse cuando se disponga de series con mayor continuidad, períodos históricos suficientemente amplios y variables adicionales que justifiquen el incremento de complejidad. La finalidad sería ampliar la comparación entre familias de modelos bajo condiciones metodológicas equivalentes y no asumir de manera previa que una mayor complejidad producirá necesariamente mejores resultados.

Desde la perspectiva de aplicación, se recomienda desarrollar un piloto controlado en uno o varios locales antes de considerar una implementación productiva. Este piloto debería comparar las estimaciones del modelo con el procedimiento habitual de planificación y registrar indicadores que actualmente no forman parte del conjunto de datos, como cantidades efectivamente horneadas, disponibilidad del producto durante la jornada, devoluciones y posibles ventas no atendidas. Las estimaciones deberían mantenerse como información de apoyo para el personal y no como órdenes automáticas de horneo. Este tipo de evaluación permitiría determinar si el uso de las predicciones genera cambios observables en la operación y diferenciar el desempeño predictivo del impacto operativo real.

Cualquier ampliación del modelo a otros productos, locales o cadenas de panaderías debería realizarse mediante una nueva evaluación de la calidad y estructura de los datos y del desempeño predictivo. Los resultados de esta investigación

corresponden a los diez locales, quince productos y período analizados, por lo que su aplicación a otros contextos requiere verificar previamente que las condiciones de demanda, disponibilidad de información y operación sean comparables.

6. Referencias Bibliográficas

- Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1), 60–74. <https://doi.org/10.1016/j.ejor.2017.02.046>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, 109(11), 1997–2028. <https://doi.org/10.1007/s10994-020-05910-7>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. In M. C. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9* (pp. 155–161). MIT Press.
- Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, 38(4), 1283–1318. <https://doi.org/10.1016/j.ijforecast.2019.06.004>
- Fries, M., & Ludwig, T. (2024). ‘Why are the sales forecasts so low?’ Socio-technical challenges of using machine learning for forecasting sales in a bakery. *Computer*

Supported Cooperative Work, 33, 253–293. <https://doi.org/10.1007/s10606-022-09458-z>

Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.

Hernández-Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.

Hewamalage, H., Bergmeir, C., & Bandara, K. (2022). Global models for time series forecasting: A simulation study. *Pattern Recognition*, 124, 108441. <https://doi.org/10.1016/j.patcog.2021.108441>

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>

Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*, 36(4), 1420–1438. <https://doi.org/10.1016/j.ijforecast.2020.02.005>

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>

- Jauhar, S. K., Harinath, S., Krishnaswamy, V., & Paul, S. K. (2025). Explainable artificial intelligence to improve the resilience of perishable product supply chains by leveraging customer characteristics. *Annals of Operations Research*, 354, 103–142. <https://doi.org/10.1007/s10479-024-06348-z>
- Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data*, 6(4), Article 15. <https://doi.org/10.1145/2382577.2382579>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.
- Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press.
- Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- Makridakis, S. (1993). Accuracy measures: Theoretical and practical concerns. *International Journal of Forecasting*, 9(4), 527–529. [https://doi.org/10.1016/0169-2070\(93\)90079-3](https://doi.org/10.1016/0169-2070(93)90079-3)
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74. <https://doi.org/10.1016/j.ijforecast.2019.04.014>

- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364. <https://doi.org/10.1016/j.ijforecast.2021.11.013>
- Montero-Manso, P., & Hyndman, R. J. (2021). Principles and algorithms for forecasting groups of time series: Locality and globality. *International Journal of Forecasting*, 37(4), 1632–1653. <https://doi.org/10.1016/j.ijforecast.2021.03.004>
- Muth, M., Lingenfelder, M., & Nufer, G. (2025). The application of machine learning for demand prediction under macroeconomic volatility: A systematic literature review. *Management Review Quarterly*, 75, 2759–2802. <https://doi.org/10.1007/s11301-024-00447-8>
- Rodrigues, M., Miguéis, V. L., Freitas, S., & Machado, T. (2024). Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste. *Journal of Cleaner Production*, 435, 140265. <https://doi.org/10.1016/j.jclepro.2023.140265>
- Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- Seyam, A., Mathew, S. S., Du, B., El Barachi, M., & Shen, J. (2025). A stacking ensemble model for food demand forecasting: A preventative approach to food waste reduction. *Cleaner Logistics and Supply Chain*, 15, 100225. <https://doi.org/10.1016/j.clscn.2025.100225>
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>

- van Donselaar, K., van Woensel, T., Broekmeulen, R., & Fransoo, J. (2006). Inventory control of perishables in supermarkets. *International Journal of Production Economics*, 104(2), 462–472. <https://doi.org/10.1016/j.ijpe.2004.10.019>
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324–342. <https://doi.org/10.1287/mnsc.6.3.324>

Anexo 1 – Prototipo académico para consulta de predicciones de demanda

1. Propósito del prototipo

Como complemento al desarrollo de la investigación, se implementó un prototipo académico orientado a representar de manera funcional la consulta de las predicciones generadas por los modelos desarrollados. Su propósito fue facilitar la visualización de la relación entre los datos utilizados, los modelos entrenados y las estimaciones de demanda horaria obtenidas para los locales y productos incluidos en el estudio.

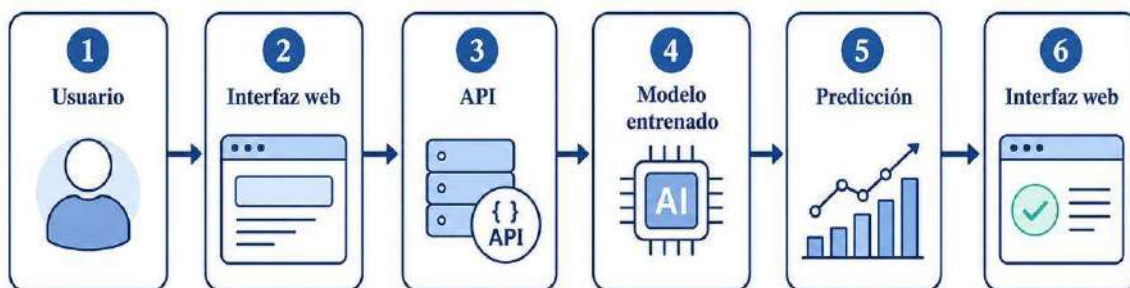
El prototipo fue concebido con una finalidad demostrativa y académica. Permite mostrar cómo los resultados del proceso predictivo pueden presentarse mediante una interfaz de consulta, aproximando el desarrollo técnico a un escenario de utilización por parte de un usuario. Su implementación no modifica el alcance metodológico de la investigación ni constituye una solución desplegada dentro de la operación de la empresa.

2. Arquitectura y componentes

El prototipo fue desarrollado mediante una arquitectura compuesta por una interfaz web, una interfaz de programación de aplicaciones (API) y los artefactos generados durante el entrenamiento de los modelos predictivos. De manera simplificada, la Figura 10 representa el flujo general de funcionamiento del prototipo académico. El usuario realiza una consulta mediante la interfaz web, la cual envía la solicitud a la API. Esta procesa los parámetros ingresados y utiliza el modelo entrenado correspondiente para generar la predicción de demanda. El resultado es devuelto a la interfaz web para su visualización.

Figura 10

Flujo general del prototipo académico para consulta de predicciones de demanda



La interfaz web permite al usuario seleccionar los parámetros necesarios para realizar una consulta. Estas solicitudes son enviadas a una API desarrollada con FastAPI, encargada de procesar la petición y acceder a los componentes requeridos para generar la respuesta. Los modelos entrenados y los elementos necesarios para su utilización se almacenan como artefactos serializados. La solución incluye modelos de Ridge Regression, Random Forest, XGBoost, LightGBM, Support Vector Regression y Holt-Winters, además de los codificadores y métricas generados durante el proceso de entrenamiento.

La interfaz fue desarrollada utilizando React y permite presentar información relacionada con el análisis exploratorio, métricas de los modelos, predicciones y ejercicios de validación. La infraestructura del prototipo se organizó mediante Docker Compose, lo que permite ejecutar de manera reproducible los componentes correspondientes al análisis, entrenamiento, API e interfaz web.

3. Funcionamiento general

Para realizar una consulta de predicción, el usuario selecciona principalmente el local, uno o varios productos, la fecha sobre la cual desea obtener la estimación y el modelo que desea utilizar, de acuerdo con las opciones disponibles en el prototipo.

Una vez ingresados los parámetros, la interfaz envía la solicitud a la API. Esta procesa la información requerida, utiliza las transformaciones y características definidas durante el desarrollo del modelo y accede al artefacto correspondiente al algoritmo seleccionado. A partir de este proceso se genera una estimación de la demanda horaria observada para la combinación consultada.

El resultado se devuelve a la interfaz web, donde puede visualizarse de manera estructurada para facilitar su interpretación. La predicción representa una estimación cuantitativa basada en los patrones históricos identificados en los datos y debe interpretarse como información de apoyo para la planificación del horneo, no como una instrucción automática de producción.

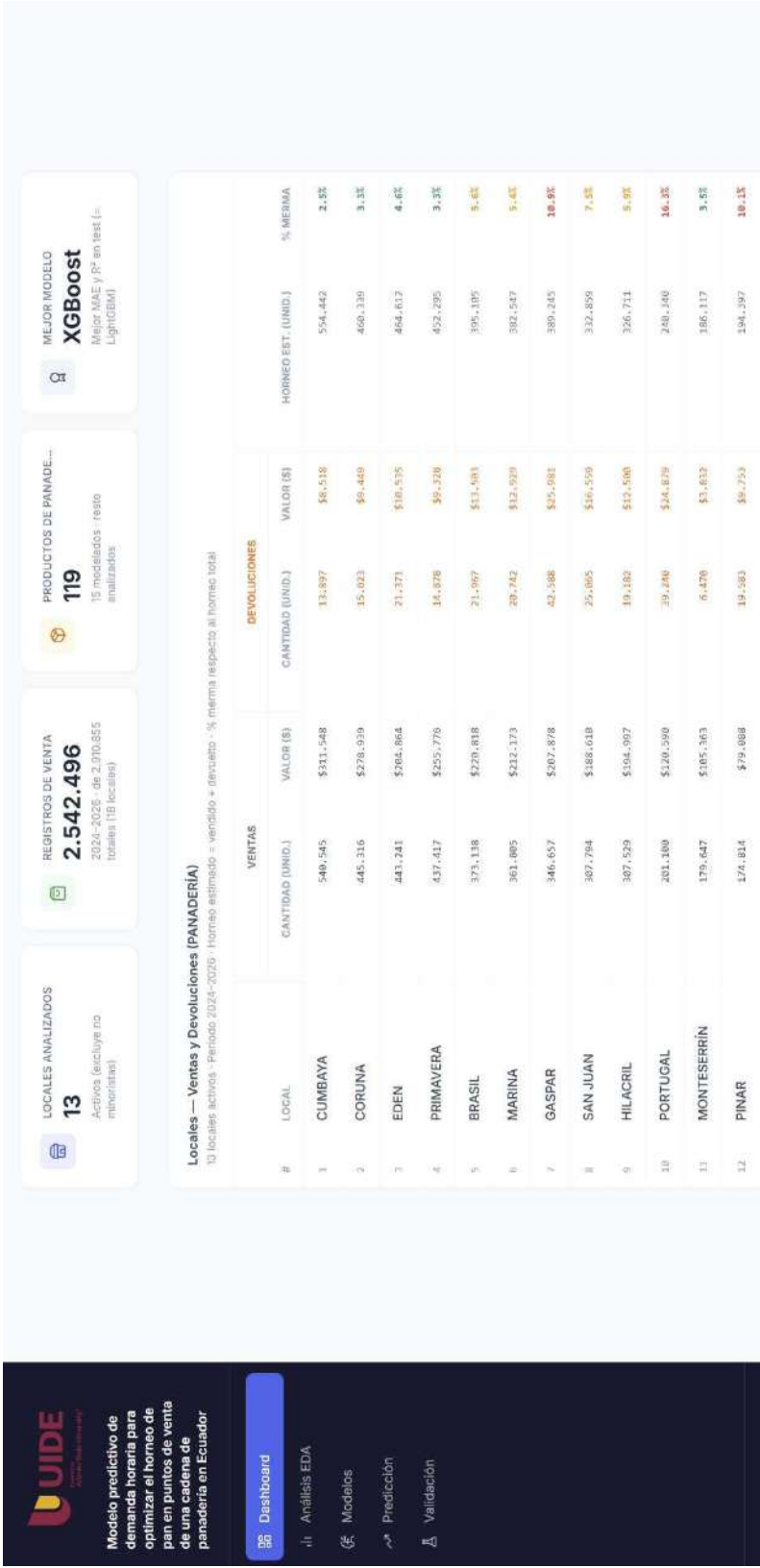
El prototipo también permite consultar información complementaria relacionada con las métricas de los modelos y realizar comparaciones con datos posteriores cuando estos se encuentran disponibles. Estas funcionalidades tienen una finalidad demostrativa y permiten evidenciar el flujo existente entre los datos procesados, los modelos entrenados y los resultados presentados al usuario.

4. Evidencia visual

La Figura 11 presenta la interfaz desarrollada para realizar consultas de predicción de demanda horaria por local y producto. El usuario selecciona los parámetros de la consulta y el sistema devuelve las estimaciones generadas por el modelo seleccionado. La interfaz tiene una finalidad demostrativa y no corresponde a una aplicación desplegada dentro de los sistemas productivos de la organización.

Figura 11

Interfaz del prototipo académico para consulta de predicciones de demanda



5. Alcance y limitaciones

El prototipo constituye un desarrollo académico complementario destinado a representar de forma funcional los resultados generados durante la investigación. Su alcance se limita a la consulta y visualización de información derivada de los modelos desarrollados. El prototipo no se encuentra integrado con el sistema de planificación de recursos empresariales ni con el sistema de punto de venta de la organización. Los datos utilizados corresponden a los conjuntos procesados durante la investigación y no se obtienen mediante una conexión en tiempo real con los sistemas transaccionales.

Asimismo, la aplicación no genera órdenes automáticas de horneado ni determina de forma autónoma las cantidades que deben prepararse. Las estimaciones constituyen referencias cuantitativas que deben ser interpretadas conjuntamente con las condiciones operativas y el criterio del personal responsable de cada local.

La solución no fue desplegada en un ambiente productivo ni sometida a una evaluación formal de usabilidad. Tampoco se realizó un piloto operativo que permitiera medir efectos sobre las cantidades efectivamente horneadas, la disponibilidad de producto, las devoluciones, las ventas o la demanda no atendida.

Por esta razón, el prototipo no constituye evidencia adicional para la validación del desempeño predictivo ni para la comprobación de la hipótesis de investigación. La validación del modelo se sustenta en las métricas obtenidas sobre los conjuntos de validación y prueba y en la evaluación temporal posterior descrita en el Capítulo 4. La función del prototipo es exclusivamente demostrar, de manera aplicada, el flujo existente entre los datos, los modelos entrenados y las estimaciones presentadas al usuario.