



Modelos de aprendizaje automático para el pronóstico de la demanda en eventos promocionales en el comercio electrónico: un estudio aplicado a ventas minoristas

Machine learning models for demand forecasting at promotional events on e-commerce: an applied study on retail sales

Tomás Duque Giraldo

tduqueg@eafit.edu.co

Trabajo de Grado

Asesor, docente

Diego Fernando Fonseca Valero

dffonsecav@eafit.edu.co

UNIVERSIDAD EAFIT

ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA

MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA

MEDELLÍN

2025

CONTENIDO

INTRODUCCIÓN.....	7
PLANTEAMIENTO DEL PROBLEMA	9
JUSTIFICACIÓN	11
OBJETIVOS	12
GENERAL.....	12
ESPECÍFICOS.....	12
MARCO CONCEPTUAL.....	13
DISEÑO METODOLÓGICO	15
CONTENIDO DEL DATASET	15
UNIDAD DE MODELADO A NIVEL DE TIENDA.....	17
PROCEDIMIENTO EXPERIMENTAL Y COMPARACIÓN DE MODELOS	17
MÉTRICAS DE EVALUACIÓN DEL PRONÓSTICO	21
RESULTADOS GENERALES.....	24
DESEMPEÑO EN FECHAS ESPECIALES	27
CONCLUSIONES.....	29
CONCLUSIONES PRINCIPALES	29
ALCANCES Y LIMITACIONES DEL ESTUDIO	31
IMPLICACIONES PRÁCTICAS PARA LA GESTIÓN DE LA DEMANDA	33
LÍNEAS DE INVESTIGACIÓN FUTURA	35
REFERENCIAS	36
ANEXOS.....	42

LISTA DE FIGURAS

Tabla 1. Archivos base utilizados y su rol en el pipeline de modelado	15
Tabla 2. Resultados comparativos por tienda y modelo (LightGBM y SARIMAX, con y sin precio)	24
Figura A.1 Flujo metodológico del proceso de pronóstico por tienda	16
Figura B.1 Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda CA_1	25
Figura B.2 Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda CA_2	44
Figura B.3 Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda CA_3	45
Figura B.4 Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda CA_4	45
Figura B.5 Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda TX_1	46
Figura B.6 Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda TX_2	46
Figura B.7 Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda TX_3	47
Figura B.8 Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda WI_1	47

Figura B.9 Comparación de modelos LightGBM y SARIMAX con variable de
precio para la tienda WI_2 48

Figura B.10 Comparación de modelos LightGBM y SARIMAX con variable de
precio para la tienda WI_3 48

RESUMEN

La creciente relevancia del comercio electrónico ha intensificado la necesidad de pronósticos de demanda precisos, especialmente durante eventos promocionales que generan picos anómalos de consumo como Black Friday o Navidad. Estos eventos conducen a incrementos súbitos en las ventas que los métodos estadísticos tradicionales no logran anticipar adecuadamente. Como consecuencia, pueden producirse quiebres de stock o sobreinventarios. Para abordar este problema, el presente trabajo aplica técnicas avanzadas de aprendizaje automático, capaces de incorporar patrones no lineales y variables contextuales, con el objetivo de mejorar la exactitud de las estimaciones de demanda en dichos escenarios. En particular, se evalúa un modelo predictivo de aprendizaje automático basado en árboles conocido como LightGBM con métodos de pronóstico tradicionales, empleando un conjunto de datos público de ventas minoristas de Estados Unidos. De este modo se identifica el enfoque más efectivo bajo condiciones de alta volatilidad y se examina su pertinencia durante períodos festivos y promocionales en los que el comportamiento del consumidor se altera significativamente.

Palabras clave: Pronóstico de demanda, Comercio electrónico, Eventos promocionales, Aprendizaje automático, Modelos de ensamble basado en árboles.

ABSTRACT

The growing importance of e-commerce has intensified the need for accurate demand forecasts, especially during promotional events that generate abnormal consumption peaks such as Black Friday or Christmas. These events lead to sudden increases in sales that traditional statistical methods fail to anticipate adequately. As a result, stock shortages or overstocking may occur. To address this problem, this paper applies advanced machine learning techniques capable of incorporating nonlinear patterns and contextual variables, with the aim of improving the accuracy of demand estimates in such scenarios. In particular, a tree-based machine learning predictive model known as LightGBM is evaluated against traditional forecasting methods, using a public dataset of US retail sales. This identifies the most effective approach under conditions of high volatility and examines its relevance during holiday and promotional periods when consumer behavior changes significantly.

Keywords: Demand forecasting, E-commerce, Promotional events, Machine learning, Tree-based ensemble models.

INTRODUCCIÓN

La última década ha visto un acelerado crecimiento del comercio electrónico a nivel global, transformando los hábitos de consumo y las operaciones minoristas (Statista, 2022g). En 2021 las ventas en línea superaron los 5 billones de dólares mundialmente y se proyecta que en 2026 alcancen más de 8 billones (Statista, 2022g). De igual forma, mercados emergentes como Colombia evidencian incrementos notables; por ejemplo, en 2022 el comercio electrónico colombiano creció 38,4 %, con una facturación de 55,7 billones de pesos, según la Cámara Colombiana de Comercio Electrónico (CCCE) (CCCE, 2023). Este auge del e-commerce plantea nuevos desafíos en la gestión de la cadena de suministro, particularmente en la planificación y el pronóstico de la demanda. Lograr proyecciones precisas se ha vuelto crítico para asegurar la disponibilidad de productos y la satisfacción del cliente en entornos altamente dinámicos (Fildes, Goodwin & Önköl, 2018; Kourentzes, Trapero & Barrow, 2019).

Un desafío específico en comercio electrónico es anticipar la demanda durante eventos promocionales de gran volatilidad, como Black Friday, Cyberlunes o la temporada navideña. Estas campañas especiales generan picos de ventas atípicos en periodos cortos, rompiendo los patrones históricos y dificultando la predicción mediante métodos estadísticos tradicionales (Trapero, Pedregal, Fildes & Kourentzes, 2013). En la práctica, muchas empresas se ven obligadas a realizar ajustes manuales de último momento sobre los pronósticos para incorporar información de promociones, con resultados variables (Fildes & Goodwin, 2007; Fildes et al., 2009). Esto refleja la necesidad de enfoques de pronóstico más robustos que capten objetivamente el efecto de eventos excepcionales de demanda con mínima intervención humana.

En años recientes, modelos de pronóstico basados en aprendizaje automático han surgido como alternativas prometedoras para afrontar estas limitaciones. A diferencia de los métodos tradicionales, las técnicas de *machine learning* pueden modelar relaciones no lineales e incorporar múltiples variables externas

(calendarios de promociones, tendencias de búsqueda, comportamiento del usuario, etc.), mejorando la exactitud predictiva (Bandara et al., 2019; Makridakis, Spiliotis & Assimakopoulos, 2022). Estudios reportan que algoritmos como Random Forest superan consistentemente a modelos estadísticos clásicos en escenarios de ventas con alta fluctuación (Bandara et al., 2019; Vesga Acevedo, 2020). Incluso en competencias globales de pronóstico se han obtenido estimaciones más precisas y menos sesgadas con enfoques de aprendizaje automático que con métodos tradicionales cuando se aprovechan grandes volúmenes de datos (Makridakis et al., 2022).

El presente trabajo de investigación se inscribe en este contexto, uniendo la necesidad práctica y la oportunidad técnica: aprovechar el potencial del *machine learning* para optimizar el pronóstico de la demanda en comercio electrónico durante eventos promocionales de alta volatilidad. Es importante aclarar que, debido a la disponibilidad limitada de datos detallados en el ámbito local, la validación empírica de la propuesta se realiza utilizando un conjunto de datos público de ventas minoristas de Estados Unidos (M5 – Walmart Sales Forecasting). No obstante, los hallazgos obtenidos buscan ser relevantes y aplicables en entornos de comercio electrónico con campañas promocionales intensivas.

PLANTEAMIENTO DEL PROBLEMA

Prever con precisión la demanda durante eventos promocionales de gran magnitud como Black Friday, Cyberlunes y Navidad se ha convertido en un reto crítico para las empresas de comercio electrónico. La incapacidad de anticipar adecuadamente estos picos de demanda tiene consecuencias severas en la gestión: una subestimación puede resultar en quiebres de stock, retrasos en entregas y pérdidas de ventas, afectando la satisfacción del cliente y la reputación del minorista (Seifert et al., 2015), mientras que una sobrestimación conduce a excedentes de inventario con costos elevados de almacenamiento, liquidación de sobrantes y riesgo de obsolescencia de productos (Nenes et al., 2010). Ambos escenarios implican pérdidas financieras considerables y reducen la eficiencia operativa. Estudios de caso documentan que, ante la incertidumbre de la demanda en promociones, muchas compañías optan por inflar sus inventarios en un 20–30 % por encima de lo habitual, incrementando los costos logísticos (Nenes et al., 2010; Ramanathan & Muyldermans, 2011). Además, un pronóstico poco confiable dificulta la planeación de personal, la coordinación con proveedores y la preparación de la infraestructura tecnológica necesaria para soportar picos de tráfico, pudiendo generar cuellos de botella durante los eventos pico (Huang, Fildes & Soopramanien, 2014). En suma, no predecir adecuadamente la demanda en fechas críticas erosiona la competitividad de los minoristas digitales, traduciéndose en pérdidas de ingresos y oportunidades en un mercado cada vez más disputado.

Desde la perspectiva científica, esta problemática evidencia un vacío en la capacidad predictiva de los enfoques tradicionales y plantea preguntas abiertas en la literatura. Los métodos clásicos de series de tiempo, como Holt-Winters o el modelo autorregresivo integrado de media móvil (ARIMA), han resultado insuficientes para capturar la dinámica abrupta de las promociones si no integran información contextual (Ali et al., 2009; Trapero et al., 2013; Fildes et al., 2009). Se ha observado empíricamente que, ante campañas de gran impacto, los modelos univariados basados solo en patrones históricos incurren en errores sustanciales al

subestimar los picos reales de demanda (Ali et al., 2009; Trapero et al., 2013; Fildes et al., 2009). En ausencia de un tratamiento explícito de estos eventos extraordinarios, los pronósticos tienden a rezagarse o a quedar amortiguados frente al nivel real del pico (Fildes et al., 2009). La academia reconoce que la previsión de la demanda en escenarios de alta volatilidad sigue siendo un problema no resuelto que exige el desarrollo de nuevos modelos y enfoques analíticos (Abolghasemi et al., 2020). En este contexto, surge la siguiente pregunta de investigación, focalizada en los eventos promocionales descritos: ¿Cómo mejorar la precisión del pronóstico de la demanda en comercio electrónico durante eventos promocionales de alta volatilidad?

JUSTIFICACIÓN

Este estudio aborda la predicción de la demanda bajo volatilidad extrema en comercio electrónico, un ámbito que presenta vacíos importantes en la literatura. Diversos trabajos han investigado promociones, eventos especiales y métodos avanzados de pronóstico, pero aún es limitada la evidencia sobre la efectividad relativa de enfoques de aprendizaje automático frente a los tradicionales en tales situaciones, especialmente en contextos con demanda irregular y datos históricos escasos propios de mercados emergentes (Ali et al., 2009; Nikolopoulos et al., 2015). En respuesta a ello, se plantea una evaluación comparativa rigurosa entre técnicas de aprendizaje automático y métodos estadísticos convencionales para capturar los “shocks” de promociones y cambios abruptos. El estudio aportará evidencia empírica sobre el desempeño de cada enfoque y delineará las condiciones bajo las cuales unos u otros resultan más adecuados, contribuyendo así al avance de modelos teóricos y aplicaciones prácticas de pronóstico (Bandara et al., 2019; Abolghasemi et al., 2020).

Desde la perspectiva empresarial, lograr mayor precisión en el pronóstico durante campañas como Black Friday habilita decisiones más informadas en compras, producción y logística; permite optimizar inventarios, reducir quiebres de stock y evitar sobreaprovisionamientos. Esto se traduce en una mejor eficiencia operativa y, según estudios previos, en disminuciones de costos de inventario y distribución de hasta un 15 % (Huang et al., 2014; Nenes et al., 2010; Kourentzes et al., 2019; Ramanathan, 2012). Anticipar adecuadamente los picos de demanda también posibilita alinear con suficiente antelación recursos críticos como personal, infraestructura tecnológica, materiales de empaque y transporte, manteniendo la calidad del servicio durante los periodos de alto tráfico. Adicionalmente, el presente trabajo proporciona un enfoque analítico de pronóstico replicable que puede

fortalecer la ventaja competitiva y la satisfacción del cliente en las empresas de comercio electrónico (Seifert et al., 2015; Fildes et al., 2018).

OBJETIVOS

GENERAL

Identificar y evaluar un modelo de pronóstico de demanda basado en técnicas de machine learning para eventos promocionales de alta volatilidad en comercio electrónico, comparando su desempeño en precisión predictiva frente a métodos tradicionales de pronóstico.

ESPECÍFICOS

1. Analizar las características de la demanda en eventos promocionales de gran magnitud en comercio electrónico, identificando patrones históricos, niveles de volatilidad y limitaciones de los enfoques de pronóstico actuales a través de una revisión de la literatura y del análisis de datos reales de ventas.
2. Diseñar e implementar un modelo predictivo de aprendizaje automático capaz de incorporar variables exógenas relevantes para capturar los aumentos anómalos de demanda durante eventos especiales.
3. Comparar el desempeño del modelo de machine learning desarrollado con el de métodos de pronóstico tradicionales en escenarios de alta volatilidad, empleando métricas de error y análisis estadístico para determinar si el modelo basado en aprendizaje automático (machine learning, ML) ofrece mejoras significativas en la precisión del pronóstico.

MARCO CONCEPTUAL

En el contexto del comercio electrónico, caracterizado por un crecimiento acelerado y campañas promocionales masivas, surgen desafíos particulares para el pronóstico de la demanda. Los eventos promocionales de alta volatilidad se definen como aquellos periodos breves en los que las ventas se disparan muy por encima de los niveles habituales debido a campañas comerciales específicas (por ejemplo, Black Friday), generando picos de demanda difíciles de predecir. La alta volatilidad de estos eventos rompe los patrones históricos de ventas: las promociones inducen desplazamientos temporales y aumentos atípicos que los modelos basados solo en promedios no pueden anticipar si no incorporan información específica del suceso. La literatura documenta que estas promociones son una causa frecuente de grandes desvíos en los pronósticos, los cuales a menudo deben corregirse con ajustes de juicio experto de último momento (Fildes & Goodwin, 2007; Fildes et al., 2009; Ali et al., 2009). Además, la estacionalidad propia de ciertas épocas (p. ej., la temporada navideña) añade complejidad, ya que su variabilidad anual se superpone con las campañas masivas de fin de año, dificultando separar el efecto estacional normal del impacto de la promoción táctica (Heizer & Render, 2014; Forbes Staff, 2024; Ali et al., 2009).

Los métodos tradicionales de pronóstico como el suavizamiento exponencial, ARIMA o modelos de regresión se basan en extrapolar los patrones pasados bajo supuestos de estabilidad y de relaciones lineales o aditivas. Por tal motivo, con frecuencia omiten información contextual relevante (promociones, shocks, tendencias emergentes) y ven incrementado su error ante cambios estructurales bruscos en la serie (Hyndman, 2014; Lawrence et al., 2006). Si bien existen extensiones que incorporan variables exógenas (por ejemplo, ARIMA con variables exógenas (ARIMAX) o modelos de suavizamiento exponencial en espacio de estados (ETS) con regresores), su parametrización se vuelve compleja y no garantiza mejoras sustanciales. De hecho, la evidencia de competencias internacionales de pronóstico advierte que una mayor sofisticación estadística no

conlleva necesariamente mayor precisión, e incluso puede degradarla por problemas de sobreajuste. En la práctica, esto obliga a producir un pronóstico base que luego debe ajustarse manualmente, un proceso oneroso y poco escalable (Trapero et al., 2013; Makridakis & Hibon, 2000; Green & Armstrong, 2015; Fildes et al., 2009).

En contraste, el aprendizaje automático ofrece una vía prometedora al detectar patrones complejos directamente de los datos sin asumir una forma funcional predeterminada. Las técnicas de machine learning pueden incorporar múltiples predictores como precios, calendarios promocionales, tendencias de búsqueda, entre otros y modelar interacciones no lineales así como efectos temporales con mayor flexibilidad, logrando reducir el error de pronóstico en contextos donde los métodos clásicos no capturan adecuadamente la complejidad (Bandara et al., 2019; Carbonneau, Laframboise & Vahidov, 2008; Kremer, Siemsen & Thomas, 2015). Estudios empíricos han mostrado que incluso mejoras moderadas en la precisión mediante machine learning pueden traducirse en beneficios económicos relevantes (Choi, Chan & Yue, 2017; Mitrea, Lee & Wu, 2009, citado en Teo, 2020).

Entre las técnicas de machine learning, los métodos de ensamble basados en árboles de decisión han demostrado particular eficacia en tareas de pronóstico. En particular, los algoritmos de gradient boosting como LightGBM destacan por su capacidad para modelar relaciones no lineales y manejar gran cantidad de predictores sin requerir supuestos estadísticos rígidos. En el sector minorista, estos enfoques han mostrado ventajas frente a modelos univariados de series de tiempo; por ello, en entornos con promociones intensivas, estacionalidades superpuestas y datos históricos limitados, un modelo de boosting de árboles como LightGBM se perfila como una alternativa idónea para complementar o incluso superar a los métodos clásicos. Su aplicación puede, además, fundamentar decisiones

operativas más precisas en contextos de comercio electrónico (Breiman, 2001; Zhang et al., 2019; Fildes, Goodwin & Önköl, 2018).

DISEÑO METODOLÓGICO

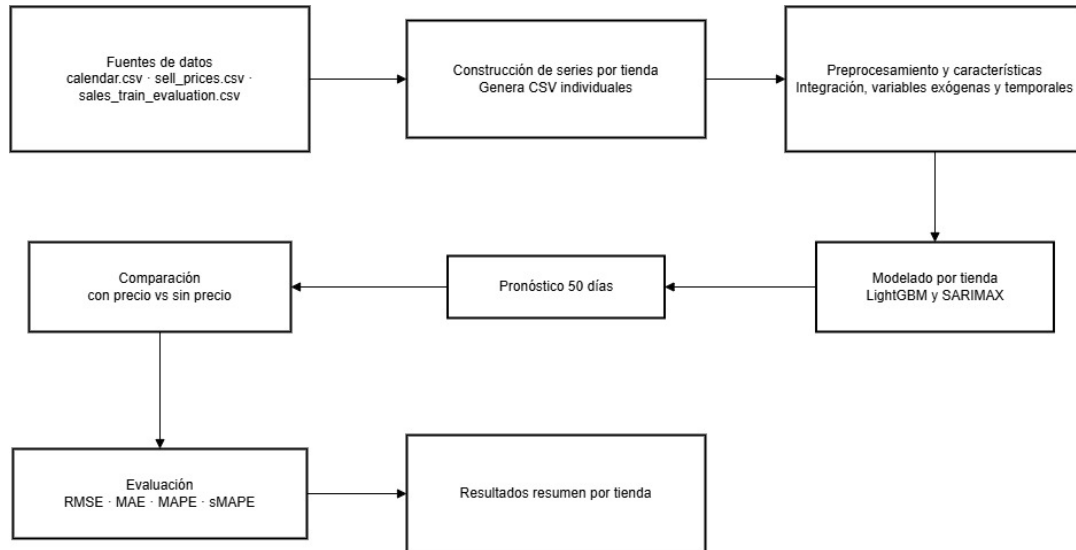


Figura A.1 Flujo metodológico del proceso de pronóstico por tienda

Fuente: Elaboración propia.

Como se resume en la Figura A.1, el proceso metodológico contempla la construcción de series por tienda, la integración de variables exógenas y la evaluación comparativa entre LightGBM y SARIMAX.

CONTENIDO DEL DATASET

Para llevar a cabo el experimento se utilizó el conjunto de datos público M5 Walmart Sales Forecasting, que contiene series de ventas diarias de una cadena minorista en EE. UU. Este dataset, si bien proviene de tiendas físicas, resulta pertinente para estudiar picos de demanda comparables a los del comercio electrónico: incluye eventos promocionales y festivos que generan volatilidad similar en las ventas. Dada la falta de datos abiertos de empresas de e-commerce con tales eventos, M5 se adoptó como caso representativo para probar los métodos propuestos, asumiendo que los patrones de demanda bajo promociones en retail físico pueden

extrapolarse conceptualmente al entorno online. En concreto, M5 abarca 10 tiendas de Walmart ubicadas en tres estados de Estados Unidos de América los cuales son California, Texas y Wisconsin, con datos diarios de unidades vendidas por producto durante 5 años, precios de venta promedio semanales y un calendario detallado de fechas especiales. La base se compone de tres archivos principales: un histórico de ventas diarias titulado `sales_train_evaluation.csv`, un archivo de calendario titulado `calendar.csv` con información de cada fecha como día de la semana, feriados y eventos e indicadores del programa SNAP (Supplemental Nutrition Assistance Program) de asistencia alimentaria, y un archivo de precios titulado `sell_prices.csv` que registra el precio de venta semanal de cada producto en cada tienda. La Tabla 1 ilustra la relación entre estos archivos mediante sus llaves y resume el contenido de cada archivo junto con sus variables clave y el rol que cumplen en la preparación de los datos para el modelo.

Archivo	Llaves principales	Variables clave	Rol en el pipeline
<code>sales_train_evaluation.csv</code>	<code>id</code> , <code>item_id</code> , <code>dept_id</code> , <code>cat_id</code> , <code>store_id</code> , <code>state_id</code>	Columnas diarias <code>d_1</code> , <code>d_2</code> , ... con unidades vendidas representando una columna por día	Serie base de ventas. Se pivotea de <i>wide</i> a <i>long</i> y se agrega al nivel elegido (tienda en este trabajo).
<code>calendar.csv</code>	<code>d</code> , <code>date</code> , <code>wm_yr_wk</code>	<code>weekday</code> , <code>event_name_1/2</code> , <code>event_type_1/2</code> , <code>snap_CA/TX/WI</code>	Mapea <code>d</code> →fecha y aporta exógenas de calendario/eventos y SNAP.
<code>sell_prices.csv</code>	<code>store_id</code> , <code>item_id</code> , <code>wm_yr_wk</code>	<code>sell_price</code> (precio promedio semanal)	Exógena opcional de precio; se une por semana y nivel tienda-ítem y luego se agrega.

Tabla 1 Archivos base utilizados y su rol en el pipeline de modelado

Fuente: Elaboración propia

UNIDAD DE MODELADO A NIVEL DE TIENDA

Si bien el contexto motivador se enmarca en empresas de comercio electrónico, para los fines del experimento se eligió la tienda física individual como unidad de análisis del pronóstico. Este nivel representa un equilibrio entre granularidad operativa y estabilidad estadística. Al agregar las ventas de todos los ítems de una misma tienda, se obtiene una serie de tiempo más regular: se reducen la intermitencia y el ruido aleatorio propios de productos individuales, emergiendo patrones temporales mejor definidos, lo cual facilita el uso de rezagos e identificadores calendáricos en el modelado. Además, muchas decisiones logísticas (reabastecimiento de inventario, dotación de personal, promociones locales) se toman a escala de tienda; por tanto, trabajar al nivel tienda hace que el pronóstico sea directamente accionable para la gestión. Los picos de demanda debidos a promociones o feriados se manifiestan y se gestionan típicamente en cada tienda, de modo que este enfoque conserva la sensibilidad a eventos especiales (como los festivos nacionales y el efecto SNAP propio de cada estado) sin diluirlos en promedios regionales o globales. En resumen, el nivel tienda permite al modelo capturar los efectos de calendario (festividades, días de pago de subsidios, etc.) de forma localizada y relevante para las decisiones operativas.

PROCEDIMIENTO EXPERIMENTAL Y COMPARACIÓN DE MODELOS

Sobre la base de los datos descritos, se diseñó un experimento de pronóstico con el siguiente esquema: para cada serie temporal de cada tienda, se reservó el segmento final de 50 días como conjunto de prueba (holdout) destinado a la evaluación, utilizando el resto de las observaciones anteriores para el entrenamiento. Es decir, por cada tienda se entrenaron los modelos con datos hasta 50 días antes del final del historial, y luego se generaron pronósticos día a día para esos últimos 50 días, los cuales se compararon contra las ventas reales observadas en dicho periodo. Cabe anotar que este horizonte de prueba cubre principalmente

el inicio de la primavera de 2016 e incluyó la festividad de Pascua de ese año, proporcionando al menos un evento especial dentro de la ventana de evaluación.

Se implementaron y compararon dos enfoques de modelado de la demanda, El primero un modelo de Gradient Boosting basado en árboles de decisión llamado LightGBM, y el segundo un modelo estadístico ARIMA con términos estacionales y exógenos llamado SARIMAX. Cada modelo fue calibrado por separado para las 10 tiendas, capturando así los patrones específicos de cada ubicación. Adicionalmente, para cada enfoque se consideraron dos variantes, una que incorpora la variable de precio y otra que no, con el fin de determinar el impacto de conocer la información de precios en la capacidad predictiva:

- LightGBM es un método de aprendizaje automático de ensamble de árboles que puede aprovechar múltiples características explicativas. Se entrenó un modelo de LightGBM independiente por tienda. En la variante “LGBM con precio”, se incluyó como predictor exógeno el precio de venta `sell_price` junto con diversas variables de entrada derivadas de la fecha, rezagos de la demanda como las ventas de días anteriores como features e indicadores de eventos especiales, entre otras características. Por su parte, en la variante ‘LGBM sin precio’, se excluyó la información de las variaciones de precio; para mantener la estructura del dataset se dejó el campo de precio con un valor constante (el precio medio histórico), lo que equivale a un predictor sin señal adicional. De este modo, el modelo no puede explotar la variación de precio, y el escenario representa un caso en el que solo se utiliza la información temporal y de calendario, evaluando cuánto mejora o no la predicción al conocer las fluctuaciones de precios. Los hiperparámetros de LightGBM, principalmente una tasa de aprendizaje pequeña (0,05), un número máximo de 5.000 iteraciones de boosting con parada temprana, 63 hojas por árbol (`num_leaves`), un mínimo de 20 observaciones por hoja (`min_data_in_leaf`), y fracciones de muestreo de características y observaciones del 90 % (`feature_fraction` = 0,9 y `bagging_fraction` = 0,9) se

fijaron de manera conservadora y se validaron por tienda mediante un esquema de holdout temporal: en cada serie se reservaron los últimos 28 días del conjunto de entrenamiento como ventana de validación interna, utilizándose dicha ventana para ajustar el número efectivo de árboles vía early stopping y verificar la estabilidad del error de validación, priorizando la comparabilidad entre las variantes con y sin precio frente a la optimización extrema de cada modelo. LGBM con precio: Incluye el precio de venta como variable predictora exógena junto con las variables de fecha, lags de demanda, indicadores de eventos especiales, etc.

- SARIMAX es un modelo ARIMA que incorpora estacionalidad multiplicativa y covariables exógenas. Específicamente, se utilizó una configuración SARIMAX (1,0,1)×(1,0,1)_7 (orden ARIMA no estacional 1–0–1 con estacionalidad semanal 7 días) seleccionada tras analizar las funciones de autocorrelación, buscando capturar la dinámica básica de tendencia y estacionalidad semanal presente en las series. Se estimó un modelo SARIMAX separado para cada tienda. En la variante “SARIMAX con precio”, al modelo ARIMA de cada tienda se le incorporó la serie de precio promedio semanal como regresor exógeno (alineando el precio con la misma semana correspondiente a cada observación de ventas). En la variante “SARIMAX sin precio”, se excluyó dicho regresor y, de forma análoga al caso de LightGBM, se empleó una serie de precio fija (igual al valor medio) para que el modelo opere esencialmente solo con la componente temporal de ventas. Todos los modelos ARIMA se ajustaron con máxima verosimilitud; la especificación elegida se mantuvo homogénea para las 10 tiendas por motivos de simplicidad y consistencia en la comparación.

En cuanto a la estrategia de pronóstico multi-paso, esta difirió según el tipo de modelo, atendiendo a la naturaleza de cada uno. Con LightGBM, que no genera pronósticos multihorizonte de forma directa, se adoptó un enfoque autoregresivo

iterativo: el modelo predice un día adelante, y luego ese valor pronosticado se realimenta como parte de las entradas como lag del día anterior para predecir el siguiente día, repitiendo el proceso sucesivamente hasta completar los 50 días del horizonte. Este procedimiento emula cómo se usaría el modelo en producción, recalculando la predicción día a día y propagando cualquier error cometido a futuros pasos lo que permite evaluar también la robustez del modelo al acumular error. En cambio, con SARIMAX se aprovechó la capacidad del modelo estadístico para proyectar varios pasos en una sola corrida, se generó directamente el pronóstico de los 50 días futuros en bloque, dado que el ARIMA puede extrapolar la serie usando su estructura autorregresiva-estacional. Para asegurar una comparación coherente, ambos enfoques recibieron el mismo tratamiento de las variables exógenas en el horizonte de pronóstico; en las configuraciones con precio se suministraron al modelo los valores futuros reales del precio, mientras que en las versiones “sin precio” se continuó utilizando el precio constante. Cabe señalar que esta diferencia en la estrategia multi-paso constituye una limitación de la comparabilidad estricta entre modelos: el enfoque autoregresivo iterativo de LightGBM tiende a acumular el error paso a paso, mientras que el pronóstico en bloque de SARIMAX proyecta directamente los 50 días futuros de una sola vez. En la práctica, cada esquema refleja un uso típico de estas familias de modelos, pero es posible que favorezcan o perjudiquen marginalmente a uno u otro enfoque; esta restricción debe tenerse en cuenta al interpretar los resultados comparativos.

Es importante aclarar las condiciones de conocimiento de las variables exógenas de cara a un uso práctico de estos modelos. Ciertas señales de calendario son determinísticas y se conocen de antemano con certeza; por ejemplo, los días festivos nacionales, las fechas promocionales fijas como Black Friday o las festividades navideñas, así como el calendario de pagos del programa SNAP en EE. UU. que sigue un patrón predefinido cada mes para cada estado. Por tanto, incluir estas variables exógenas en el modelo es a la vez metodológicamente válido

y conceptualmente fundado, el precio de venta se espera que afecte la cantidad demandada a través de la elasticidad precio-demanda y de las promociones asociadas a descuentos; los indicadores de eventos especiales capturan cambios bruscos en el tráfico y en las ventas y los indicadores SNAP reflejan días en los que ciertos segmentos de consumidores disponen de mayor capacidad de compra debido al pago del subsidio, de modo que estas señales conocidas de antemano permiten ajustar los pronósticos al alza o a la baja en fechas donde es razonable esperar un comportamiento distinto al habitual. En contraste, otras variables exógenas como el precio futuro de los productos no siempre están aseguradas al momento de pronosticar la demanda. En la práctica, las empresas pueden tener planes promocionales previstos como descuentos programados en cuyo caso esa información sí podría alimentarse al modelo pero también enfrentan incertidumbre o decisiones de precio que se toman sobre la marcha. En este estudio se proporcionaron los precios reales futuros para poder medir el máximo beneficio potencial de conocerlos; sin embargo, en un entorno operativo real, si el precio no estuviera fijado por anticipado, habría que optar por enfoques alternativos: o bien excluir el precio del modelo como en las variantes analizadas “sin precio”, o bien generar un pronóstico adicional del precio o escenarios supuestos de precio, reconociendo que cualquier error en esa predicción influirá en el pronóstico de demanda. Esta distinción es crítica para interpretar los resultados: un desempeño superior de la variante con precio en nuestro análisis supone que la organización cuenta con una estimación confiable del precio futuro.

MÉTRICAS DE EVALUACIÓN DEL PRONÓSTICO

El desempeño de cada modelo se cuantificó mediante cuatro métricas de error, calculadas sobre el período de holdout de 50 días en cada tienda. Las métricas utilizadas fueron:

- **MAE Error Absoluto Medio:** Es el promedio de las desviaciones absolutas entre el valor pronosticado y el valor real. Se mide en las mismas unidades de la variable en este caso, unidades vendidas. Un MAE más bajo indica un pronóstico más preciso en términos absolutos. Por ejemplo, un MAE = 200 significa que, en promedio, el modelo se está equivocando en 200 unidades vendidas por día.
- **RMSE Raíz del Error Cuadrático Medio:** Es la raíz cuadrada de la media de los errores al cuadrado. Penaliza más fuertemente los errores grandes debido al cuadrático, por lo que un RMSE elevado puede indicar que hubo algunos días con errores muy altos. Al igual que el MAE, se expresa en unidades de venta. Un RMSE cercano al MAE sugiere que los errores son relativamente uniformes; en cambio, si el RMSE es mucho mayor que el MAE, implica presencia de outliers o algunos días con grandes diferencias entre pronóstico y realidad.
- **MAPE Porcentaje de Error Absoluto Medio:** Es el promedio del error absoluto expresado como porcentaje del valor real. Se calcula como $MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$ donde Y_t es el valor real y $\hat{Y}_t$ el pronóstico. Esta métrica es independiente de la escala de la serie, lo que permite comparaciones relativas. El mape indica que, en promedio, el modelo se desvía cierto porcentaje del valor real. Valores extremos o cercanos a cero en la demanda pueden distorsionar el MAPE produciendo valores muy altos, por lo que debe interpretarse con cuidado en esas situaciones.
- **sMAPE Porcentaje de Error Absoluto Medio Simétrico:** Es una variación del MAPE que utiliza una fórmula simétrica: $sMAPE = \frac{100\%}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{(Y_t + \hat{Y}_t)/2}$. Esta métrica está acotada entre 0% y 200% y trata de mitigar problemas cuando Y_t es muy pequeño, ya que normaliza el error respecto al promedio entre real y pronosticado. El sMAPE suele tener un comportamiento similar al MAPE para errores moderados, pero aporta estabilidad ante valores muy bajos o

cuando se sobrestima completamente la demanda. Un sMAPE más bajo indica mejor ajuste, siendo 0% un pronóstico perfecto.

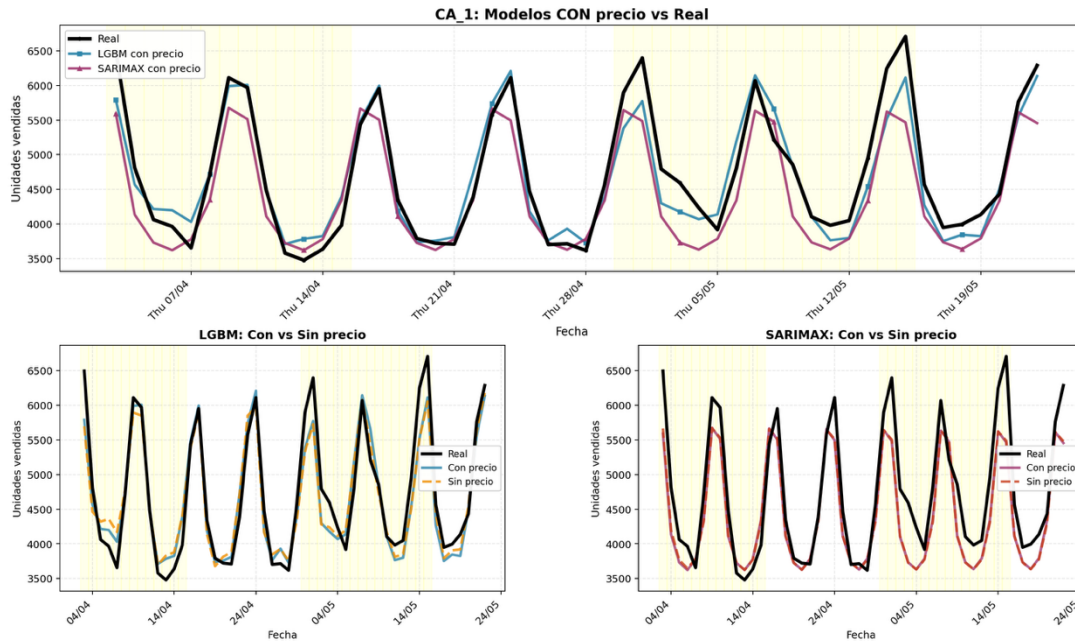
En conjunto, estas métricas ofrecen una visión complementaria del desempeño predictivo. El MAE y RMSE cuantifican el error en términos absolutos, lo cual es útil para evaluar el impacto operativo de las desviaciones; por ejemplo, saber en promedio en cuántas unidades se está fallando por día (MAE) y si existen días con errores mucho mayores a lo usual (RMSE). Por otro lado, el MAPE y sMAPE expresan el error en términos relativos al nivel de demanda, facilitando la comparación de la precisión entre tiendas de distinto tamaño o entre periodos de venta alta y baja. La comparación entre métricas absolutas y porcentuales también permite detectar inconsistencias, idealmente, un modelo con mejor desempeño debería mostrar tanto un MAE bajo como un MAPE bajo. Las definiciones y el uso de estas métricas de evaluación se basan en Hyndman y Athanasopoulos (2021).

RESULTADOS GENERALES

Tienda	Modelo	MAE	RMSE	MAPE (%)	sMAPE (%)
CA_1	LGBM con precio	235.79	299.03	4.95	4.99
	LGBM sin precio	255.17	317.94	5.39	5.42
	SARIMAX con precio	376.64	462.55	7.59	7.96
	SARIMAX sin precio	373.33	455.28	7.53	7.90
WI_2	LGBM con precio	415.86	529.01	7.58	7.47
	LGBM sin precio	407.72	536.77	7.30	7.30
	SARIMAX con precio	820.15	1113.32	13.42	15.01
	SARIMAX sin precio	819.65	1112.98	13.41	15.00
CA_2	LGBM con precio	375.31	477.07	7.81	8.11
	LGBM sin precio	396.71	524.55	8.03	8.45
	SARIMAX con precio	360.08	476.64	7.01	7.33
	SARIMAX sin precio	362.02	478.78	7.05	7.38
TX_2	LGBM con precio	462.99	591.76	11.56	10.59
	LGBM sin precio	378.11	471.66	9.52	8.88
	SARIMAX con precio	271.54	350.54	6.55	6.71
	SARIMAX sin precio	266.71	345.31	6.45	6.58

Tabla 2 Muestra parcial de resultados por tienda y configuración de modelo

La Tabla 2 presenta los resultados para un subconjunto de tiendas representativas; los resultados completos para las 10 tiendas analizadas se presentan en la Tabla A.1 en el anexo.



En la Figura B.1 se ilustra un ejemplo de la demanda real frente a los pronósticos por modelo para una tienda representativa (CA_1). Arriba: comparación entre modelos con precio vs. ventas reales. Abajo izquierda: LightGBM con y sin precio. Abajo derecha: SARIMAX con y sin precio. Las comparaciones gráficas completas para las 10 tiendas se presentan en el anexo (Figuras B.1–B.10).

La Tabla 2 resume el desempeño cuantitativo de ambos enfoques en el conjunto de 10 tiendas analizadas, en particular, el modelo de LightGBM exhibió un desempeño superior al de SARIMAX en 7 de las 10 tiendas analizadas, lo que sugiere que, al menos en este conjunto de casos, los enfoques flexibles con múltiples covariables y no linealidades pueden capturar mejor la dinámica de la demanda que un modelo ARIMA estacional estándar. Ejemplos representativos incluyen CA_1, donde LGBM con precio alcanzó $MAPE \approx 4,95\%$ frente a $\approx 7,6\%$ de SARIMAX, y WI_2, donde LGBM sin precio registró $\approx 7,3\%$ contra $\approx 13,4\%$ de SARIMAX. Estos resultados sugieren que, cuando la señal proviene de interacciones y efectos no triviales como calendario, rezagos, potenciales asimetrías, los métodos de ensamblado basados

en árboles logran una mejor aproximación funcional que los modelos lineales con estructura ARIMA estacional.

No obstante, se identificaron casos en que SARIMAX fue competitivo o marginalmente superior, particularmente en series dominadas por patrones base de tendencia y estacionalidad y con menor aporte marginal de covariables. En CA_2 y CA_3, la variante SARIMAX con precio obtuvo los errores más bajos; por ejemplo, en CA_2 el MAPE $\approx 7,01\%$ de SARIMAX superó al $\approx 7,8\%$ de LightGBM. La simplicidad de SARIMAX, su representación explícita de la estacionalidad y la correcta identificación de órdenes parecen suficientes cuando la estructura temporal explica la mayor parte de la variabilidad y la elasticidad a factores exógenos es débil o inconsistente.

El impacto de incluir el precio fue heterogéneo al comparar cada modelo consigo mismo. En varias tiendas de California, añadir el precio mejoró ligeramente la precisión de LGBM; en CA_1, por ejemplo, el MAPE disminuyó de $\approx 5,39\%$ sin precio a $\approx 4,95\%$ con precio, lo que sugiere cierta elasticidad precio/demanda aprovechable por el algoritmo. En contraste, en TX_1 y TX_3 las métricas con y sin precio fueron prácticamente indistinguibles $\approx 6\text{--}7\%$ MAPE, indicando que el precio semanal no aportó señal incremental respecto a la ya capturada por rezagos y estacionalidades. Más aún, en TX_2 la inclusión del precio empeoró el rendimiento de LightGBM de $\approx 9,5\%$ sin precio a $\approx 11,6\%$ con precio, apuntando a ruido o sobreajuste: variaciones de precio poco alineadas con la respuesta como descuentos de corto plazo sin efecto persistente pueden introducir fluctuaciones espurias que el modelo aprende indebidamente; en ese contexto, un precio estabilizado actúa como regularizador efectivo.

En SARIMAX, la contribución del precio fue todavía más tenue. En la mayoría de tiendas, las diferencias entre especificaciones con y sin precio se limitaron a unas décimas de punto de MAPE, y en algunos casos fueron prácticamente idénticas como en TX_2. Esto es consistente con dos explicaciones no excluyentes: Primero, el efecto precio/demanda es intrínsecamente bajo en estas series; y/o segundo, el

tratamiento lineal del precio como regresor exógeno no captura posibles no linealidades, rezagos múltiples o efectos umbral.

Las métricas de error mostraron coherencia interna. MAE y RMSE mantuvieron la misma jerarquía entre modelos, con $RMSE/MAE \approx 1,2-1,3$, compatible con una dispersión moderada sin colas extremas. Al contrastar métricas absolutas con porcentuales, no se detectaron inconsistencias: una tienda con MAPE bajo puede presentar RMSE absoluto mayor si su escala de ventas es alta como en CA_3 con $MAPE \approx 5\%$ y $RMSE \approx 391$, donde el error absoluto luce grande pero representa una fracción estable del volumen. En tiendas pequeñas ocurre lo inverso: errores absolutos modestos pueden traducirse en MAPE relativamente altos. No se observó ningún caso con MAPE bajo y RMSE inflado de forma incoherente, lo que refuerza la validez comparativa de las estimaciones.

DESEMPEÑO EN FECHAS ESPECIALES

Dada la motivación central de este trabajo resulta crucial examinar cómo se comportaron los modelos específicamente en días asociados a picos de consumo. En el conjunto de prueba de 50 días utilizado, la principal fecha especial presente fue el domingo de Pascua de 2016 (27 de marzo de 2016), ya que eventos de fin de año como Black Friday y Navidad quedaron fuera del rango de evaluación directa. No obstante, se aprovechó la información histórica y las gráficas generadas para inferir el desempeño en ese tipo de eventos. La Figura B.1 ilustra las ventas reales versus los pronósticos de cada modelo para la tienda CA_1 durante el periodo de prueba, destacando en amarillo el día correspondiente a Pascua. Se observa que en esa fecha ambos modelos subestimaron el aumento de demanda respecto al valor real; sin embargo, el modelo LightGBM gracias a incorporar la variable indicadora del feriado anticipó en mayor medida el pico en comparación con SARIMAX, que proyectó un valor más cercano a la media habitual. En términos numéricos, el error porcentual de LightGBM en el domingo de Pascua fue menor que el de SARIMAX en dicha tienda, evidenciando la ventaja de contar con información exógena de eventos para ajustar al alza el pronóstico cuando

corresponde. De manera cualitativa, al analizar otros picos de demanda presentes en el historial como las ventas del día Black Friday en años previos dentro del dataset, se aprecia un patrón similar; LightGBM reaccionó mejor a los incrementos súbitos, elevando sus predicciones más agresivamente que SARIMAX, el cual tiende a mantenerse cercano a la tendencia media a menos que disponga de un cambio marcado en sus componentes estacionales o exógenas lineales. Ningún modelo, con todo, logró predecir con exactitud perfecta la magnitud de estos saltos, incluso con variables de evento, la volatilidad extrema de días como Black Friday implicó que quedara un error remanente significativo, tanto LightGBM como SARIMAX subestimaron en diferente grado los picos más agudos. Un caso distinto es el de eventos que redujeron drásticamente las ventas: por ejemplo, el dataset refleja que el 25 de diciembre, fecha de la navidad, las ventas caen a prácticamente cero debido al cierre de tiendas. Ambos enfoques pueden aprender este comportamiento recurrente. Esto sugiere que para eventos cíclicos conocidos, sean picos positivos o caídas como cierres, esto sugiere que, para eventos cíclicos conocidos, un modelo que incorpore variables exógenas definidas con criterios teóricos y operativos; por ejemplo, indicadores de calendario/feriados, variables de evento, SNAP, rezagos de demanda y, cuando sea viable, el precio, puede adelantarse razonablemente al cambio. En este estudio, la pertinencia de dichas variables se evaluó empíricamente mediante comparación controlada de variantes con y sin precio y manteniendo constantes las demás señales exógenas. En contraste, ante eventos inesperados o atípicos, es mucho más difícil lograr anticipación. Cualquiera de los modelos aquí estudiados enfrentaría limitaciones importantes ya que dependen de patrones aprendidos o señales predefinidas.

CONCLUSIONES

CONCLUSIONES PRINCIPALES

A la luz de los resultados obtenidos, se concluye que los métodos basados en machine learning pueden mejorar la precisión del pronóstico en contextos de demanda altamente volátil: el modelo LightGBM superó sistemáticamente a la aproximación ARIMA en la mayoría de las tiendas analizadas, venciendo en 7 de 10 tiendas y, en promedio, logró errores absolutos y porcentuales menores. Esto indica que, al incorporar múltiples factores explicativos y capturar relaciones no lineales, el enfoque de árboles en ensamble fue más eficaz para anticipar variaciones complejas en las ventas. En particular, durante los días de eventos especiales que generan picos de demanda, LightGBM mostró un mejor desempeño relativo al lograr ajustar al alza sus predicciones cuando correspondía, mientras que el modelo SARIMAX tendió a quedarse más rezagado respecto al valor real en esos picos. Dicho de otro modo, la incorporación explícita de variables indicativas de promociones/feriados en conjunto con la flexibilidad del algoritmo de boosting permitió capturar parte de los aumentos abruptos de demanda más efectivamente que la técnica tradicional.

No obstante, las conclusiones también matizan que no existe un modelo universalmente superior para todos los contextos. El rendimiento óptimo dependió de la estructura de cada serie de demanda: en aproximadamente un tercio de las tiendas evaluadas, aquellas con patrones muy regulares y estables, el modelo SARIMAX alcanzó resultados tan buenos como o incluso mejores que LightGBM. Esto evidencia que los modelos estadísticos parsimoniosos siguen representando una frontera eficiente cuando el comportamiento subyacente de la serie es suficientemente estable, periódico y lineal, de modo que puede capturarse bien con unos pocos parámetros y donde añadir complejidad no conlleva mejora sustancial. En escenarios así, un enfoque tradicional bien ajustado puede ofrecer pronósticos confiables con menor esfuerzo computacional y mayor interpretabilidad. En cambio,

en situaciones con alta volatilidad o múltiples factores incidentes, típicamente el caso en grandes campañas de venta online, un modelo de machine learning demostró ajustarse mejor a las variaciones. Los hallazgos respaldan la idea de que la elección del modelo ideal debe considerar la naturaleza de la serie, las técnicas avanzadas de ML sobresalen bajo condiciones de demanda complejas, mientras que los modelos clásicos conservan ventaja en series “tranquilas” o altamente estructuradas.

Respecto al objetivo específico de las fechas especiales de venta, los resultados permiten afirmar que se logró mejorar la capacidad predictiva durante estos eventos al emplear modelos con variables exógenas informativas. En días de picos de consumo análogos a Black Friday, Cyber Lunes o eventos navideños, los análisis realizados sugieren que un modelo como LightGBM, alimentado con indicadores de dichas fechas, puede reducir el error de pronóstico en comparación con un modelo ARIMA estacional estándar que no incorpora esa información. Por ejemplo, aunque ninguno de los métodos anticipó perfectamente la magnitud de los picos de demanda, LightGBM generalmente se aproximó más a los niveles observados durante eventos promocionales, cerrando en parte la brecha entre pronóstico y realidad. Esto confirma la pertinencia de incorporar conocimiento contextual como calendarios promocionales y feriados en los algoritmos para enfrentar cambios de régimen súbitos. Al mismo tiempo, se debe reconocer que persisten desafíos: incluso el mejor modelo presentó cierta subestimación en los días de demanda extrema, lo que indica que estos eventos siguen siendo difíciles de predecir con total exactitud debido a su naturaleza excepcional. Por ende, mejorar la previsión en fechas especiales no significa eliminar por completo el error, sino mitigarlo y acercar lo más posible la predicción al escenario real, combinando técnicas cuantitativas con una planificación proactiva.

ALCANCES Y LIMITACIONES DEL ESTUDIO

Este trabajo tiene limitaciones importantes que conviene resaltar, las cuales abren a su vez oportunidades de investigación futura. En primer lugar, aunque la motivación y la narrativa se enmarcaron en el contexto del comercio electrónico e-commerce, la implementación empírica se realizó con un dataset de retail físico estadounidense dado que no se contaba con datos públicos reales de una empresa de e-commerce local para los fines del experimento. Esto implica que las conclusiones deben interpretarse con cautela fuera del ámbito de los datos analizados: las características de demanda de Walmart en EE. UU. Como tipo de productos, comportamiento del consumidor ante promociones, calendarios de festividades pueden diferir de las de un minorista en línea colombiano u otro contexto. Si bien se argumentó que el dataset M5 proporciona un análogo razonable de picos promocionales, no deja de ser una aproximación. Como trabajo futuro, sería valioso validar los hallazgos de este estudio con datos de empresas de comercio electrónico concretas y en distintos mercados geográficos, para confirmar que las ventajas observadas de los modelos de ML se mantienen y que las conclusiones son generalizables. En segundo lugar, la ventana de evaluación de 50 días utilizada fue relativamente corta y, por diseño, no incluyó eventos como Black Friday o Navidad en el año de prueba. La performance de los modelos en esos eventos clave se infirió indirectamente a través de resultados en eventos análogos y análisis visual de historiales, pero idealmente se desearía probar los modelos prospectivamente en un periodo que abarque dichos picos masivos para medir directamente su eficacia. Esto tampoco fue posible por la naturaleza del dataset; por ello, otra extensión futura sería aplicar y testear la metodología en un horizonte que cruce por las principales temporadas pico y evaluar allí sí de forma explícita el error en esas fechas. En tercer lugar, existe la suposición de disponibilidad de las variables exógenas que ya se discutió, en el experimento se proporcionó a los modelos información que en un entorno real podría no estar disponible, particularmente los precios futuros. Aunque se analizó el caso sin precios para reflejar la falta de dicha información, es importante enfatizar que cualquier

implementación práctica enfrentará la necesidad de pronosticar o fijar de antemano las variables de entrada que no sean plenamente conocidas. Por ejemplo, si una empresa desea usar un modelo como el aquí propuesto para pronosticar durante el próximo Black Friday, deberá alimentar al modelo con un calendario de promociones definido y posiblemente con supuestos o proyecciones de precio; de lo contrario, la eficacia real del sistema podría verse reducida frente a lo aquí reportado. En cuarto lugar, el alcance metodológico de la comparación fue limitado: se eligió un algoritmo de bosques de decisión por ser representativo de técnicas de ML eficientes en series temporales, y se lo contrastó contra un modelo ARIMA tradicional. Si bien esta elección estuvo justificada y alineada con los objetivos y con la literatura que sugiere las ventajas de los árboles de decisión en entornos promocionales, no se exploró un espectro más amplio de modelos. Queda pendiente investigar cómo se comportarían otras metodologías avanzadas por ejemplo, redes neuronales recurrentes o modelos LSTM o modelos híbridos que combinen componentes estadísticos y de ML. Asimismo, el estudio se centró en pronósticos a nivel de tienda individual de forma independiente; no se abordaron enfoques multivariados o jerárquicos que pudieran explotar información cruzada entre múltiples series, por ejemplo, pronosticar conjuntamente todas las tiendas o todas las categorías de producto. Tales enfoques podrían potencialmente mejorar la precisión al compartir información común entre series especialmente para eventos que afectan a muchas series a la vez, por lo que representan otra área de exploración futura. Por último, se reconoce que la muestra de datos utilizada es reducida en términos estadísticos. Si bien permitió un análisis detallado y controlado, una muestra mayor podría proveer evidencia más robusta de las conclusiones. Esta limitación se debió en parte a restricciones computacionales y de tiempo, por lo que un trabajo futuro podría expandir el experimento a un número mayor de series para validar la consistencia de los resultados a mayor escala.

IMPLICACIONES PRÁCTICAS PARA LA GESTIÓN DE LA DEMANDA

Desde el punto de vista práctico, los hallazgos de este proyecto conducen a recomendaciones concretas para la gestión de la demanda en entornos de venta con alta estacionalidad promocional. En primer lugar, las empresas deberían considerar adoptar modelos de aprendizaje automático como herramienta base de pronóstico, especialmente para aquellos periodos o líneas de producto sujetos a picos de demanda impredecibles. En este estudio, LightGBM demostró ser una alternativa sólida al capturar patrones complejos, por lo que su implementación podría ayudar a reducir errores en situaciones donde los métodos tradicionales tienden a fallar. Sin embargo, esto no implica descartar las técnicas clásicas, los resultados sugieren más bien un enfoque híbrido donde, para series muy estables, un modelo parsimonioso tipo ARIMA puede seguir utilizándose como referencia o incluso como opción preferible si ofrece un ajuste equivalente con menor variabilidad. En segundo lugar, es fundamental integrar explícitamente el calendario de eventos especiales y promociones en el proceso de predicción. Esto significa que la planificación de ventas debe nutrirse con información sobre fechas clave de manera estructurada, proveer al modelo dichas variables exógenas se ha comprobado beneficioso para que anticipe alzas o bajas fuera de lo común. Una implementación práctica de esta recomendación es mantener actualizado un calendario comercial e incorporarlo en las bases de datos de pronóstico con suficiente anticipación, de modo que el algoritmo conozca cuándo esperar comportamiento anómalo. De igual forma, se recomienda evaluar cuidadosamente la inclusión de la variable de precio en los modelos operativos. Si la empresa cuenta con una estrategia de precios predefinida y se ha evidenciado mediante validación histórica que incluir esos cambios mejora la precisión, entonces aprovechar un modelo con precio puede ser valioso. Por el contrario, si el precio futuro es incierto o altamente volátil, o si las pruebas no muestran un beneficio claro, es preferible no alimentar dicha variable al modelo o hacerlo de forma controlada por ejemplo usando el último precio conocido o un promedio, para evitar introducir ruido.

Adicionalmente, se desprende como recomendación la implementación de un monitoreo constante del error de pronóstico y la adopción de mecanismos de ajuste manual durante los eventos de demanda extraordinarios. En la práctica, esto implica que durante jornadas como Black Friday el equipo de planificación de la demanda siga en tiempo real las ventas vs. pronósticos y pueda introducir correcciones puntuales cuando se observen desviaciones significativas. Es importante señalar que estos mecanismos de monitoreo y ajuste manual no formaron parte del diseño experimental de este trabajo, sino que se plantean como lineamientos operativos complementarios, inspirados en la literatura y en la constatación de que incluso los modelos avanzados mantienen errores residuales en eventos de demanda extrema. Ningún modelo automático será infalible en contextos de cambios abruptos como se evidenció en Black Friday y eventos similares, por lo que complementar el sistema con supervisión humana puede mitigar riesgos. En la práctica, esto implica que durante jornadas como Cyber Monday el equipo de planificación de la demanda siga en tiempo real las ventas vs. pronósticos; si se detectan desviaciones mayores a cierto umbral, se pueden reajustar al alza las previsiones de los días restantes del evento, evitando quedarse cortos en inventario. Del mismo modo, si un modelo predijera un pico que no se materializa, es necesaria la intervención para corregir a la baja y prevenir sobreinventarios. Este esquema de tracking diario y reglas de decisión asociadas actúa como red de seguridad ante los inevitables errores que persisten incluso con modelos avanzados, y es altamente recomendable incorporarlo en la operación. De hecho, muchas empresas líderes combinan sus sistemas algorítmicos con ajustes de juicio experto en momentos críticos, lo cual se alinea con las recomendaciones de la literatura para mejorar los pronósticos en la planificación de la cadena de suministro (Fildes et al., 2009).

LÍNEAS DE INVESTIGACIÓN FUTURA

En cuanto a investigaciones futuras, se proponen varias direcciones para ampliar y profundizar lo aquí logrado. Una extensión natural es replicar este estudio en un contexto puramente de e-commerce por ejemplo, con datos de una tienda en línea regional para verificar si las conclusiones permanecen y evaluar si surgen nuevos desafíos como diferentes hábitos de compra, menor lealtad a fechas específicas, entre otros. Otra vía es explorar modelos de aprendizaje más avanzados, como redes neuronales LSTM o modelos basados en transformers especializados en series temporales, que podrían captar dependencias de larga duración o interacciones entre múltiples series de manera más efectiva. Asimismo, sería interesante desarrollar enfoques híbridos o en ensamblado que combinen lo mejor de ambos mundos como por ejemplo, usar pronósticos ARIMA como una característica adicional dentro de un modelo de boosting, o emplear promedios ponderados entre LightGBM y un método de red neuronal para aprovechar distintas perspectivas de la serie. Igualmente, se podría investigar la aplicación de modelos globales o jerárquicos, entrenando un único modelo sobre muchas series a la vez, para aprovechar patrones comunes y garantizar coherencia entre distintos niveles. Por último, otra dimensión relevante sería incorporar la medición del costo del error en la evaluación de los modelos, por ejemplo, penalizando más las subestimaciones que llevan a quiebres de stock, o asociando el error a métricas financieras, de forma que la selección del modelo más adecuado responda no solo a criterios estadísticos sino también económicos.

REFERENCIAS

- Abolghasemi, M., Eshragh, A., Hurley, J., & Fahimnia, B. (2020). Demand forecasting in the presence of systematic events: Cases in capturing sales promotions. *International Journal of Production Economics*, 230, 107892. <https://doi.org/10.1016/j.ijpe.2020.107892>
- Ali, Ö. G., Sayın, S., van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36(10), 12340–12348. <https://www.sciencedirect.com/science/article/abs/pii/S0957417409004035>
- Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., & Seaman, B. (2019). Sales demand forecast in e-commerce using long short-term memory neural network methodology. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*. <https://arxiv.org/abs/1901.04028>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cámara Colombiana de Comercio Electrónico (CCCE). (2022a). Informe anual de comercio electrónico 2021. Bogotá, Colombia. <https://incodice.org/algunas-cifras-y-hechos-sobre-el-comercio-electronico-2021-2022/>
- Cámara Colombiana de Comercio Electrónico. (2024, 28 mayo). Informe de cierre 2023 (versión pública). https://ccce.org.co/gestion_gremial/informe-de-cierre-2023-version-publica/

- Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 184(3), 1140–1154. <https://www.sciencedirect.com/science/article/abs/pii/S0377221706012057>
- Choi, T. M., Chan, H. K., & Yue, X. (2017). Recent development in big data analytics for business operations and risk management. *IEEE Transactions on Cybernetics*, 47(1), 81–92.
- Departamento Administrativo Nacional de Estadística (DANE). (2025, 14 de mayo). *Encuesta mensual de comercio (EMC), marzo de 2025* [Boletín técnico]. DANE <https://www.dane.gov.co/files/operaciones/EMC/bol-EMC-mar2025.pdf>
- Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37(6), 570–576. <https://www.jstor.org/stable/20141547>
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), 3–23. <https://doi.org/10.1016/j.ijforecast.2008.11.010>
- Fildes, R., Goodwin, P., & Önköl, D. (2018). Use and misuse of information in supply chain forecasting of promotion effects. *International Journal of Forecasting*, 34(4), 572–583. <https://doi.org/10.1016/j.ijforecast.2017.12.006>
- Forbes Staff. (2024, 7 marzo). Comercio electrónico en Colombia creció un 12,58% en 2023. *Forbes Colombia*. <https://forbes.co/2024/03/07/actualidad/comercio-electronico-en-colombia-crecio-un-1258-en-2023/>
- Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68(8), 1678–1685. <https://www.sciencedirect.com/science/article/abs/pii/S014829631500140X>

- Heizer, J., & Render, B. (2014). Principios de administración de operaciones (7.a ed.). Pearson Education. <https://matrixeducacion.com/wp-content/uploads/2023/07/Libro-5-Heizer-and-Render-Principios-de-Operaciones.pdf>
- Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, 237(2), 738–748. <https://www.sciencedirect.com/science/article/abs/pii/S0377221714001374>
- Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: Principles and Practice* (2nd ed.). OTexts. <https://otexts.com/fpp2/>
- Kourentzes, N., Trapero, J. R., & Barrow, D. K. (2019). Optimising forecasting models for inventory planning. *International Journal of Production Economics*, 209, 89–94. <https://doi.org/10.1016/j.ijpe.2019.107597>
- Kremer, M., Siemsen, E., & Thomas, D. J. (2015). The sum and its parts: Judgmental hierarchical forecasting. *Management Science*, 62(9), 2745–2764. <https://doi.org/10.1287/mnsc.2015.2259>
- Lawrence, M., Goodwin, P., O'Connor, M., & Önköl, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3), 493–518. <https://www.sciencedirect.com/science/article/abs/pii/S0169207006000501>
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions, and implications. *International Journal of Forecasting*, 16(4), 451–476. <https://www.sciencedirect.com/science/article/abs/pii/S0169207000000571>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The 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>

- Nenes, G., Panagiotidou, S., & Tagaras, G. (2010). Inventory management of multiple items with irregular demand: A case study. *European Journal of Operational Research*, 205(2), 313–324. <https://doi.org/10.1016/j.ejor.2009.12.022>
- Nikolopoulos, K., Litsa, A., Petropoulos, F., Bougioukos, V., & Khammash, M. (2015). Relative performance methods for forecasting special events. *Journal of Business Research*, 68(8), 1785–1791. <https://doi.org/10.1016/j.jbusres.2015.03.034>
- Ramanathan, U. (2012). Supply chain collaboration for improved forecast accuracy of promotional sales. *International Journal of Operations & Production Management*, 32(6), 676–695. <https://www.researchgate.net/publication/241970086>
- Seifert, M., Siemsen, E., Hadida, A. L., & Eisingerich, A. B. (2015). Effective judgmental forecasting in the context of fashion products. *Journal of Operations Management*, 36, 33–45. <https://doi.org/10.1016/j.jom.2015.02.001>
- Statista. (2022g). Retail e-commerce sales worldwide from 2014 to 2026 (in billion U.S. dollars). Statista Research Department. <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/>
- Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society*, 66(2), 299–307. https://www.researchgate.net/publication/258840311_On_the_identification_of_sales_forecasting_models_in_the_presence_of_promotions
- Trapero, J. R., Pedregal, D. J., Fildes, R., & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting*, 29(2), 234–243. <https://doi.org/10.1016/j.ijforecast.2012.09.001>
- United Nations Conference on Trade and Development. (2024, 7 junio). Making e-commerce and the digital economy work for all. <https://unctad.org/news/making-e-commerce-and-digital-economy-work-all>

Vesga Acevedo, D. N. (2020). Modelo de planeación de inventarios para e-commerce, utilizando herramientas de inteligencia artificial para hacer pronósticos de demanda y clasificación de inventarios [Tesis de Maestría, Universidad de los Andes]. Repositorio Uniandes.
<https://repositorio.uniandes.edu.co/entities/publication/25a6e1bd-661d-4dbf-9880-0215d7f667f8>

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35–62.
<https://www.sciencedirect.com/science/article/abs/pii/S0169207097000447>

Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68(8), 1678–1685.
<https://www.sciencedirect.com/science/article/abs/pii/S014829631500140X>

Babai, M. Z., & Syntetos, A. A. (2020). Demand forecasting and inventory control: A practitioner's primer. *Foundations and Trends® in Technology, Information and Operations Management*, 12(3–4), 134–219. <https://doi.org/10.1561/02000000084>

Bandara, K., Bergmeir, C., & Hewamalage, H. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications*, 140, 112896.
<https://doi.org/10.1016/j.eswa.2019.112896>

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3), 1937–1967.
<https://doi.org/10.1007/s10462-020-09896-5>

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
<https://doi.org/10.1016/j.ins.2011.12.028>

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. En *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining (pp. 785–794). ACM.
<https://doi.org/10.1145/2939672.2939785>

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
<https://doi.org/10.1198/073500102753410444>

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

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3.^a ed.). OTexts. <https://otexts.com/fpp3/>

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
<https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889. <https://doi.org/10.1371/journal.pone.0194889>

ANEXOS

Tabla A.1 Resultados completos por tienda y configuración de modelo (LightGBM y SARIMAX, con y sin precio)

Tienda	Modelo	MAE	RMSE	MAPE (%)	sMAPE (%)
CA_1	LGBM con precio	235.79	299.03	4.95	4.99
	LGBM sin precio	255.17	317.94	5.39	5.42
	SARIMAX con precio	376.64	462.55	7.59	7.96
	SARIMAX sin precio	373.33	455.28	7.53	7.90
CA_2	LGBM con precio	375.31	477.07	7.81	8.11
	LGBM sin precio	396.71	524.55	8.03	8.45
	SARIMAX con precio	360.08	476.64	7.01	7.33
	SARIMAX sin precio	362.02	478.78	7.05	7.38
CA_3	LGBM con precio	357.53	439.43	5.95	5.72
	LGBM sin precio	396.09	467.55	6.63	6.36
	SARIMAX con precio	323.29	391.38	5.08	5.09
	SARIMAX sin precio	326.57	395.53	5.10	5.14
CA_4	LGBM con precio	208.81	261.16	7.24	7.61

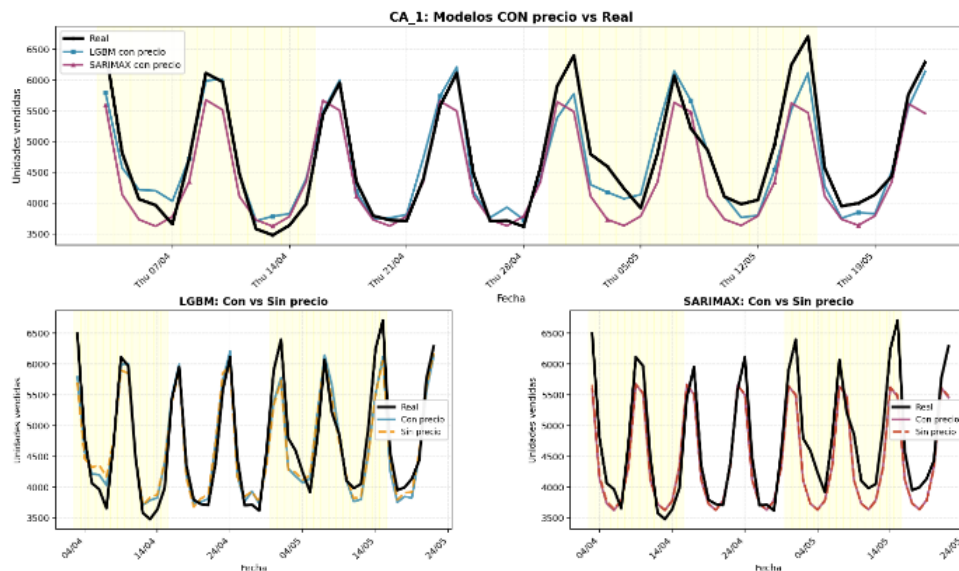
	LGBM sin precio	195.56	244.15	6.84	7.17
	SARIMAX con precio	221.65	265.60	7.75	8.11
	SARIMAX sin precio	215.17	258.66	7.52	7.86
TX_1	LGBM con precio	208.97	284.69	6.06	6.01
	LGBM sin precio	210.51	284.44	6.12	6.06
	SARIMAX con precio	304.40	389.09	8.50	8.98
	SARIMAX sin precio	304.82	389.40	8.51	8.99
TX_2	LGBM con precio	462.99	591.76	11.56	10.59
	LGBM sin precio	378.11	471.66	9.52	8.88
	SARIMAX con precio	271.54	350.54	6.55	6.71
	SARIMAX sin precio	266.71	345.31	6.45	6.58
TX_3	LGBM con precio	310.53	398.59	7.49	7.84
	LGBM sin precio	298.74	379.65	7.29	7.55
	SARIMAX con precio	376.54	485.80	8.92	9.50
	SARIMAX sin precio	375.73	485.10	8.89	9.48
WI_1	LGBM con precio	323.41	423.34	7.77	8.08
	LGBM sin precio	311.53	404.26	7.51	7.81
	SARIMAX con precio	353.07	433.47	8.70	9.20
	SARIMAX sin precio	352.78	433.16	8.69	9.19
WI_2	LGBM con precio	415.86	529.01	7.58	7.47
	LGBM sin precio	407.72	536.77	7.30	7.30
	SARIMAX con precio	820.15	1113.32	13.42	15.01
	SARIMAX sin precio	819.65	1112.98	13.41	15.00
WI_3	LGBM con precio	219.30	287.95	5.26	5.44

	LGBM sin precio	204.44	270.91	5.01	5.16
	SARIMAX con precio	424.42	571.71	9.68	10.47
	SARIMAX sin precio	421.55	568.29	9.62	10.39

Nota. Los valores corresponden a las métricas promedio de error obtenidas para cada tienda bajo las configuraciones de modelos SARIMAX y LightGBM. Los valores en **negrita** indican el mejor desempeño por tienda.

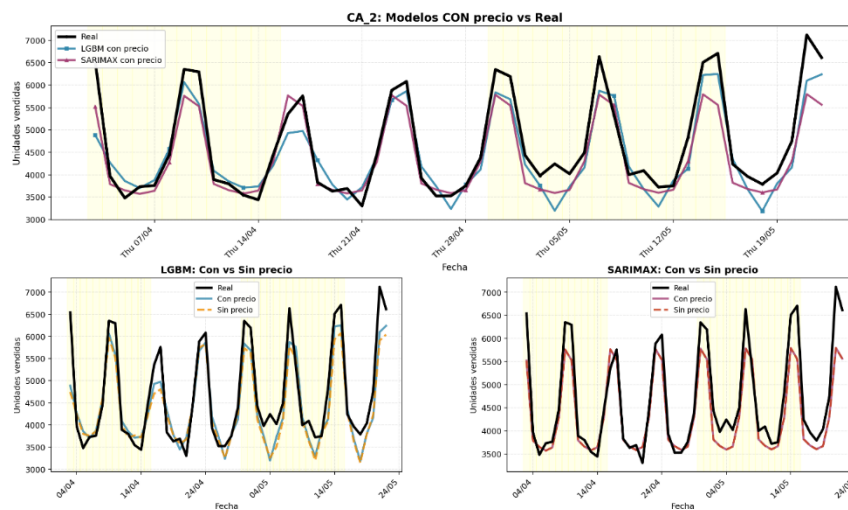
Fuente: Elaboración propia con base en los resultados del modelo

Figura B.1 - Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda CA_1.



Fuente: Elaboración propia.

Figura B.2 - Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda CA_2.



Fuente: Elaboración propia.

Figura B.3 - Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda CA_3.

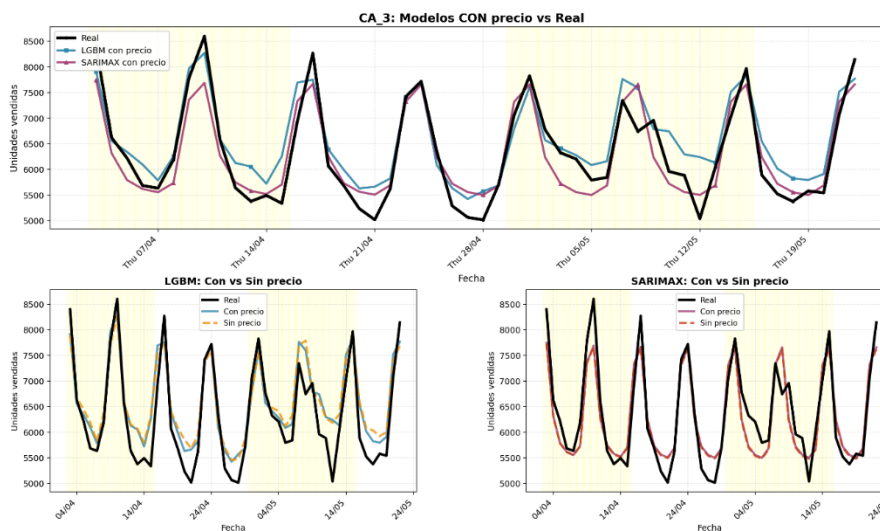


Figura B.4 - Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda CA_4.

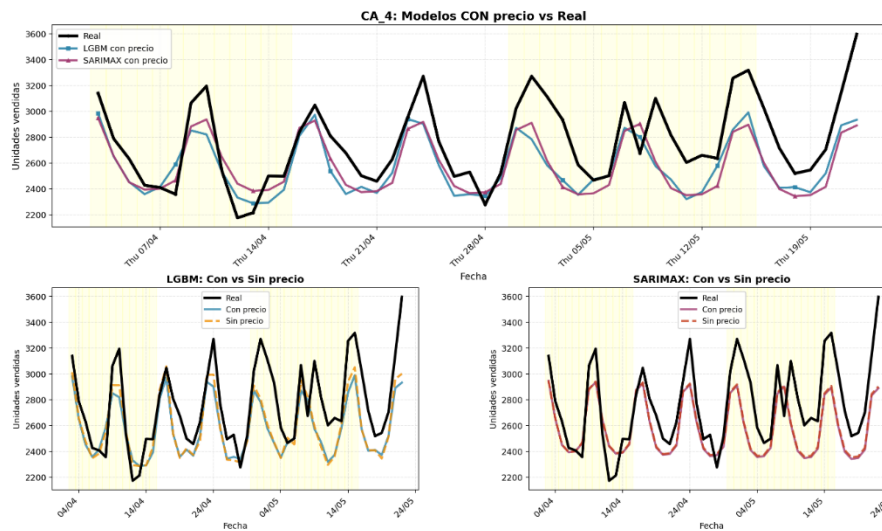
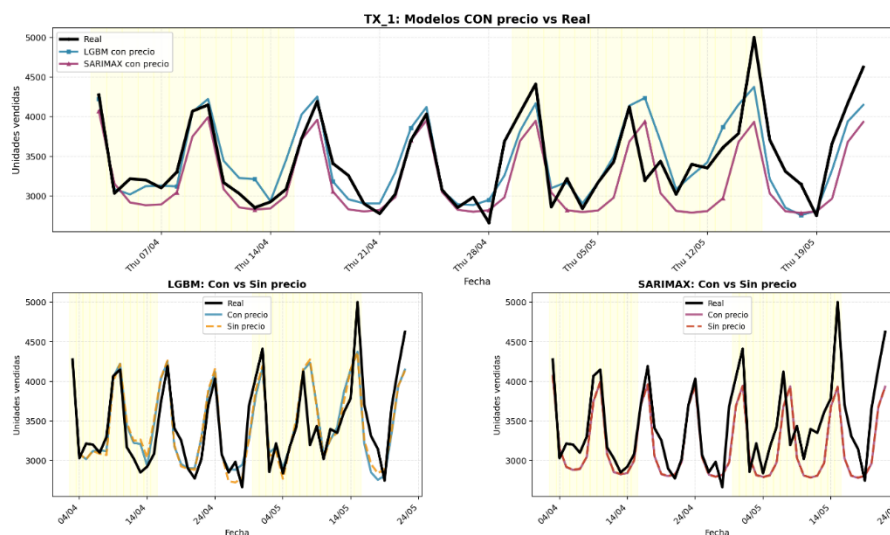


Figura B.5 - Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda TX_1.



Fuente: Elaboración propia.

Figura B.6 - Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda TX_2.

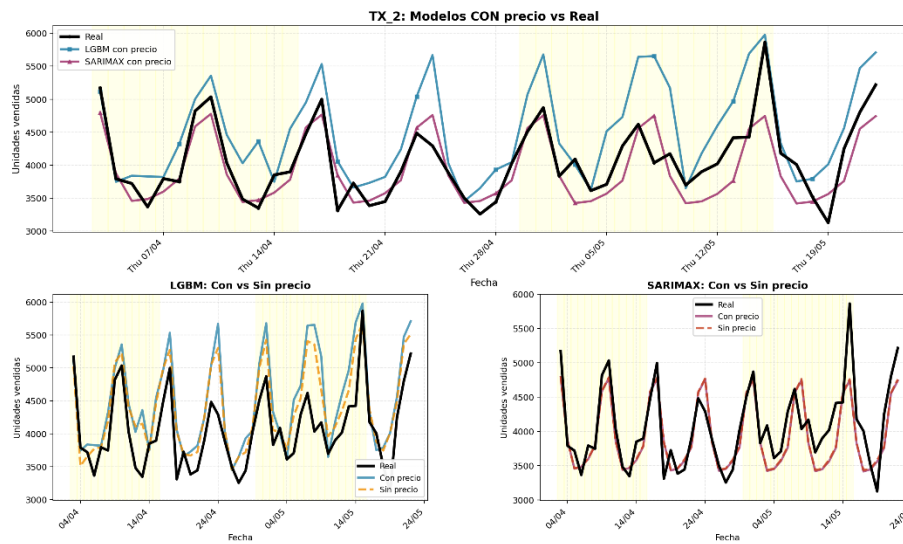
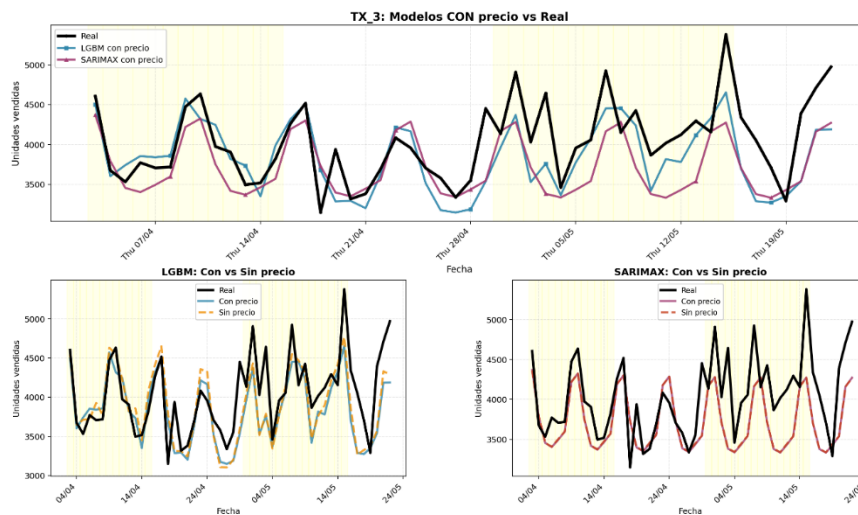
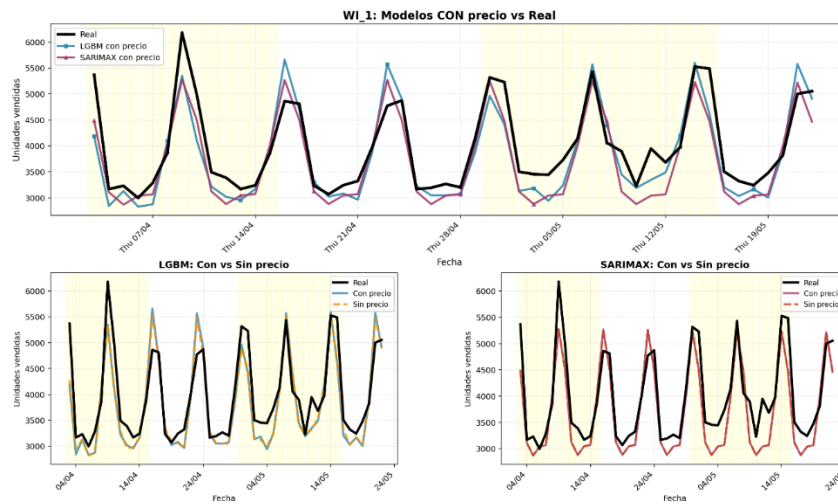


Figura B.7 - Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda TX_3.



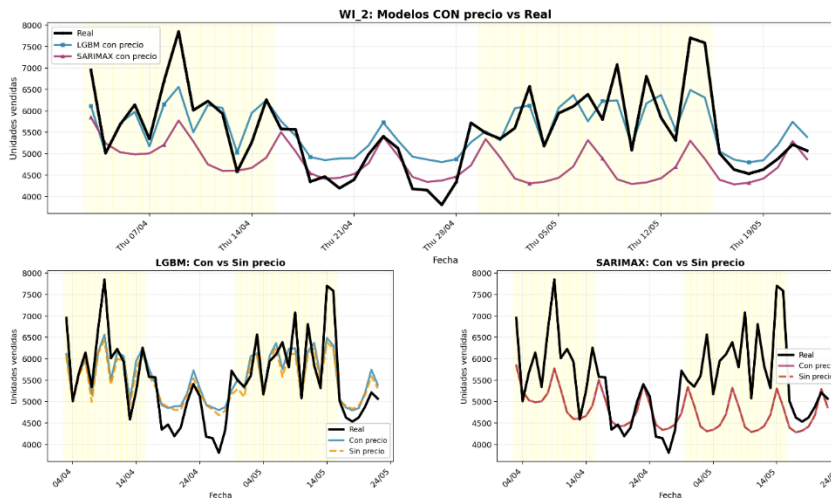
Fuente: Elaboración propia.

Figura B.8 - Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda WI_1.



Fuente: Elaboración propia.

Figura B.9 Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda WI_2.



Fuente: Elaboración propia.

Figura B.10 - Comparación de modelos LightGBM y SARIMAX con variable de precio para la tienda WI_3.

