



**MODELO DE PREDICCIÓN DE VENTA EN UNA COMPAÑÍA TEXTIL CON
TÉCNICAS DE MACHINE LEARNING**

Sales Prediction Model in a Textile Company Using Machine Learning Technics

JHON WILDER LEZCANO ECHEVERRI

Proyecto de Grado

Henry Daniel Puerta Puerta

Docente Asesor

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.....	8
JUSTIFICACIÓN.....	9
OBJETIVOS.....	10
1. GENERAL.....	10
2. ESPECÍFICOS.....	10
MARCO CONCEPTUAL	11
3. El sector moda y el comportamiento del mercado en Colombia	11
4. Fundamentos de la predicción de ventas.....	12
5. Modelos de pronóstico basados en Machine Learning	13
6. Evaluación de modelos predictivos	15
DESARROLLO DEL TRABAJO	16
Fase 1: Análisis Exploratorio y Caracterización de Datos	16
Fase 2: Ingeniería de Características.....	19
Fase 3: Implementación base	21
Fase 4: Optimización y Mejoras	23
Fase 5: Validación y Comparación.....	24
RESULTADOS.....	24
7. Desempeño global de los modelos	24
8. Análisis por modelo.....	25
Prophet.....	25
XGBoost	26

Random Forest.....	28
Regresión lineal.....	29
ANALISIS DE RESULTADOS.....	30
CONCLUSIONES.....	32
REFERENCIAS.....	34

LISTA DE FIGURAS

Figura 1. Flujo de desarrollo del sistema.	16
Figura 2. Análisis de desempeño del modelo XGBoost	27
Figura 3. Diagnósticos del modelo de Regresión Lineal	29

LISTA DE TABLAS

Tabla 1. Comparativa entre los modelos considerados.	14
Tabla 2. Descripción de las variables de la base de datos.	17
Tabla 3. Cantidad de características por modelo.	21
Tabla 4. Métricas de desempeño comparativo de los modelos implementados	25
Tabla 5. Métricas globales del modelo Prophet	25
Tabla 6. Análisis por categoría de tienda (unidades)	26
Tabla 7. Métricas del modelo Random Forest (Unidades)	28
Tabla 8. Análisis por categoría de tienda (Unidades)	29

RESUMEN

Este trabajo explora la implementación de modelos de predicción de ventas en una empresa del sector textil colombiano, a partir del uso combinado de técnicas tradicionales y enfoques basados en Machine Learning. Se analizaron datos de ventas diarias en 187 tiendas entre 2021 y 2025. Se desarrollaron cuatro fases metodológicas: (1) análisis exploratorio, (2) ingeniería de características, (3) implementación de modelos (4) Optimización y mejoras de los modelos y (5) validación comparativa. Se implementaron los modelos: Prophet, XGBoost, Random Forest y Regresión Lineal regularizada. Prophet obtuvo el mejor desempeño general en unidades ($R^2 = 0.7121$), destacándose por su capacidad para capturar patrones estacionales complejos y adaptarse a la variabilidad entre tiendas. Por su lado, XGBoost mostró alta precisión en escenarios no lineales, mientras que Random Forest ofreció robustez ante el ruido y la Regresión Lineal proporcionó mayor interpretabilidad. La ingeniería de características incorporó 83 variables, incluyendo componentes temporales, tendencias, volatilidad y efectos especiales. Mediante un análisis transversal se reveló tendencias comunes como la subestimación de picos, mayor error en tiendas pequeñas y fines de semana, y menor precisión al predecir montos monetarios con respecto a unitarios. Se concluye que la predicción de ventas mediante Machine Learning ofrece mejoras sustanciales respecto a los métodos tradicionales para una mayor eficiencia operativa, optimización de inventario y planificación financiera. Se recomienda adoptar Prophet como modelo principal y establecer ciclos mensuales de recalibración para mantener la precisión.

Palabras clave: Pronóstico de ventas, sector textil, Machine Learning, Prophet, estacionalidad, series temporales, Random Forest, XGBoost, modelos de ensemble, regresión lineal.

ABSTRACT

This study explores the implementation of sales forecasting models in a Colombian textile company, combining traditional techniques with Machine Learning-based approaches. Daily sales data from 187 stores between 2021 and 2025 were analyzed. The methodology followed five stages: (1) exploratory analysis, (2) feature engineering, (3) model implementation, (4) model optimization and fine-tuning, and (5) comparative validation. The models implemented were: Prophet, XGBoost, Random Forest, and regularized Linear Regression. Prophet achieved the best overall performance for units sold ($R^2 = 0.7121$), standing out for its ability to capture complex seasonal patterns and adapt to store-level variability. XGBoost demonstrated high accuracy in non-linear scenarios, Random Forest showed robustness to noise, and Linear Regression provided greater interpretability. Feature engineering resulted in 83 variables, including temporal components, trends, volatility, and special effects. A cross-sectional analysis revealed common patterns such as peak underestimation, higher error in smaller stores and weekends, and lower accuracy in predicting monetary values compared to units. The findings confirm that sales forecasting using Machine Learning offers substantial improvements over traditional methods, enhancing operational efficiency, inventory optimization, and financial planning. Prophet is recommended as the primary model, along with the establishment of monthly recalibration cycles to maintain accuracy.

Keywords: Sales forecasting, textile sector, Machine Learning, Prophet, seasonality, time series, Random Forest, XGBoost, ensemble models, linear regression.

INTRODUCCIÓN

La predicción precisa de ventas constituye uno de los desafíos más relevantes en la gestión operativa del sector textil. En un entorno caracterizado por ciclos de producto acelerados, márgenes de rentabilidad estrechos y consumidores con expectativas crecientes, la capacidad de anticipar la demanda futura determina la supervivencia empresarial. El sector moda en Colombia —que incluye la industria textil— es estratégico para la economía nacional, teniendo además de una amplia demanda mundial, una alta tasa de crecimiento (Colombia Productiva, 2019). Para marzo de 2025 el sector manufacturero, en el cual está inmerso el sector moda, fue el que registró una mayor variación positiva, del 4.9 %, en el Índice de Producción Industrial (DANE, 2025). Así mismo, el sector moda representa el 8.5% del PIB industrial y genera más de 590,000 empleos directos (MINCIT, 2020). Sin embargo, se han identificado cuellos de botella que limitan el crecimiento del sector, entre ellos la falta de alineación entre la oferta y la demanda del consumidor, y el escaso uso de tecnologías de punta (Colombia Productiva, 2019; Oliveros Contreras et al., 2019).

Estas condiciones exigen la incorporación de soluciones analíticas robustas que permitan pronosticar adecuadamente la demanda con el fin de mantener su competitividad por medio del desarrollo de planes de producción y ventas efectivos (Lorente-Leyva et al., 2021a).

Partiendo de esta prioridad en el mercado textil, se observa una oportunidad significativa para incorporar los avances en el modelado de datos que permitan desarrollar un modelo predictivo, y a su vez integre múltiples variables operativas y del entorno. Proporcionando así estimaciones más ajustadas a la realidad de la empresa, adaptadas tanto a sus particularidades internas como a las dinámicas del mercado local y global.

PLANTEAMIENTO DEL PROBLEMA

La ausencia de métodos basados en datos para la estimación de la venta en una compañía puede impactar negativamente en la eficiencia operativa, la planificación financiera y la competitividad en el mercado. La falta de previsión adecuada conlleva a una planificación deficiente de la producción, dificultando la gestión de envíos, provocando pérdidas por desabastecimiento o exceso de inventario, y limitando la capacidad de respuesta ante la demanda u otros cambios del entorno (Rodríguez-León & Pachón-Rincón, 2021; Xu et al., 2021).

Este problema se ve agravado por la complejidad inherente a la predicción de ventas, dada la diversidad de tiendas, cada una con patrones de comportamiento particulares, así como la influencia de factores regulares (semanales, mensuales o por quincenas de pago) e irregulares (festivos, clima, movilizaciones sociales) (Hasan, 2024). Estas variaciones generan escenarios de incertidumbre que dificultan la toma de decisiones operativas, con la consecuente pérdida de competitividad frente al mercado.

En este contexto, la implementación de técnicas de machine learning ofrecen una alternativa metodológica robusta. Estas herramientas permiten integrar múltiples variables relevantes, generar estimaciones de frecuencia diaria y ofrecer insumos accionables para una toma de decisiones más informada y eficiente.

De esta forma, se plantea la pregunta de investigación:

¿Cómo se puede desarrollar e implementar un sistema de predicción de ventas basado en machine learning que integre eficazmente factores múltiples —como promociones, estacionalidad, días festivos y fines de semana— para proporcionar predicciones diarias y accionables en una compañía del sector textil?

JUSTIFICACIÓN

El crecimiento sostenido de las ventas en la industria textil la posiciona como un motor relevante del desarrollo económico nacional. No obstante, se trata de un sector altamente dinámico y volátil, fuertemente influenciado por cambios en las tendencias de moda y variaciones en las preferencias y comportamientos del consumidor (Ren et al., 2020). Por tanto, contar con un sistema de inventario ajustado a la demanda real es decisivo en una compañía, ya que permite mejorar la satisfacción del cliente y, en consecuencia, la rentabilidad de la empresa (Loureiro et al., 2018). Para lograrlo, es necesario contar con estimaciones de ventas de alta calidad, y así sustentar decisiones operativas efectivas en los procesos de planificación y ejecución (Lorente-Leyva et al., 2021b; Xu et al., 2021).

La estimación de ventas en el sector textil enfrenta una elevada complejidad debido a su extensa cadena operativa y a la influencia simultánea de factores internos — como la disponibilidad de materias primas, la variabilidad de las colecciones y la gestión de promociones y precios (Lorente-Leyva et al., 2021b; Loureiro et al., 2018)—, y factores externos como las condiciones macroeconómicas, la competencia, fenómenos sociales e incluso climáticos (Raizada & Saini, 2021). Estas particularidades hacen que los métodos de pronóstico para esta industria requieran enfoques altamente especializados (Ren et al., 2020).

Los métodos tradicionales de predicción, basados en promedios históricos e intuición gerencial, han demostrado ser insuficientes para atender las necesidades del mercado actual (DuBreuil & Lu, 2020). En contraste, las técnicas analíticas, especialmente aquellas basadas en machine learning, ofrecen un alto potencial por su capacidad de procesar grandes volúmenes de datos heterogéneos y en distintos niveles de granularidad (Wellens et al., 2024).

En el caso particular de la empresa objeto de estudio, los errores en la previsión de ventas generan pérdidas anuales estimadas en \$6.6 millones de pesos colombianos, distribuidos entre exceso de inventario (\$2.5 millones) y ventas no concretadas (\$4.1 millones), lo cual representa el 3.2 % de sus ventas totales. Este

panorama evidencia la necesidad urgente de una solución analítica que permita transformar la gestión operativa en aras a una administración más inteligente de inventarios, recursos y campañas promocionales.

La implementación exitosa de un sistema de predicción de ventas contribuirá a mejorar la resiliencia operativa de la organización y consolidar una ventaja competitiva sostenible.

OBJETIVOS

1. GENERAL

Implementar y validar un modelo de predicción de ventas utilizando técnicas de machine learning, que supere el modelo tradicional de regresión lineal múltiple, para esto, se utilizará factores importantes para el negocio textil (promociones, ciclos en las ventas, fines de semana, pagos en quincenas) para optimizar la planificación estratégica y operativa de una compañía en el sector textil.

2. ESPECÍFICOS

- Recolectar datos históricos de ventas diarias por tienda, junto con información relevante sobre promociones, días festivos, fines de semana, y realizar el preprocesamiento necesario para asegurar la calidad y coherencia de los datos.
- Desplegar diferentes modelos de machine learning, incluyendo XGBoost, Random Forest, Prophet, y regresión lineal múltiple, para evaluar su capacidad de predecir ventas diarias a nivel de tienda.
- Evaluar el rendimiento de cada modelo utilizando métricas de precisión como R^2 , RMSE, MAE y MAPE, comparar los resultados para identificar el modelo que proporciona las métricas más fiables.
- Implementar el modelo con el mejor desempeño en un entorno de producción, integrando las predicciones en los sistemas de planificación y toma de decisiones de la compañía, y establecer un marco para el monitoreo continuo y ajuste del modelo según sea necesario.

MARCO CONCEPTUAL

3. El sector moda y el comportamiento del mercado en Colombia

El sector moda en Colombia abarca la industria textil, confección, calzado y marroquinería. La industria textil y de confecciones provee una dinámica productiva y comercial a la economía en todo el mundo, y en Colombia tiene más de 100 años aportando a los indicadores económicos del país, concentrada en regiones como Antioquia y Cundinamarca (Espinel González et al., 2018; Oliveros Contreras et al., 2019). La empresa objeto de estudio es una de las empresas pioneras, con más de 50 años en el mercado, con una participación importante en producción, presencia en las ciudades principales e intermedias y posicionada dentro del top 10 de su segmento, la cual prefiere mantenerse confidencial en este proyecto.

Una investigación de Oliveros Contreras et al. (2019) resalta que el sector ha sido impactado por los efectos de la globalización, con una fuerte presión por parte de productos importados, especialmente de bajo costo provenientes de Asia. Este fenómeno ha contribuido a una pérdida de competitividad en el mercado interno. En esa investigación se resalta que la incertidumbre en la demanda debido a factores externos como la competencia desleal, el contrabando, la elevada estacionalidad y la inestabilidad en la economía nacional, acarrea fallas en la planificación logística que obligan a las empresas a liquidar inventarios por debajo del costo, lo cual afecta su rentabilidad.

La industria textil-confecciones exhibe marcados ciclos estacionales de demanda, siendo este un aspecto clave a considerar en este trabajo. Las ventas tienden a concentrarse en ciertas temporadas pico, especialmente en los meses de fin de año. La temporada navideña es tradicionalmente la de mayor consumo de vestuario. Ciertos eventos y promociones como el Black Friday, Madrugones, combos, entre otros, suelen representar una proporción significativa de las ventas anuales. Otras estacionalidades relevantes incluyen las liquidaciones de temporada, fechas comerciales especiales (Día de la Madre, Amor y Amistad, etc.) y ferias regionales (Feria de las Flores, Carnaval de Barranquilla, entre otros), que generan picos

adicionales de demanda (DANE, 2024). Entre estos períodos de alta actividad, se observan valles o temporadas bajas donde el consumo disminuye, obligando a las marcas a ajustar estrategias para mantener el flujo de ventas.

4. Fundamentos de la predicción de ventas

La naturaleza estacional y volátil del mercado de moda hace decisiva la implementación de pronósticos de ventas acertados. Una planificación basada en datos permite a las empresas anticipar los picos de demanda y prepararse para las variaciones, evitando problemas de sobrestock o quiebres de inventario y facilitando la rotación de estos (Loureiro et al., 2018; Oliveros Contreras et al., 2019).

La predicción de ventas, también conocida como pronóstico de ventas permite estimar el comportamiento futuro de la demanda de productos. Es un proceso que busca reducir el nivel de incertidumbre al que está sujeto el pronóstico, proporcionando resultados ajustados a la realidad (Lorente-Leyva et al., 2021b).

La predicción de ventas tiene como objetivo principal apoyar la toma de decisiones informadas mediante la asignación eficiente de recursos, la planificación estratégica y financiera, y la optimización del inventario (Lorente-Leyva et al., 2021b; Raizada & Saini, 2021; Xu et al., 2021). Permite anticipar patrones de consumo, mejorar la eficiencia operativa y alinear la producción con la demanda real. Asimismo, facilita la programación logística, reduce costos, mejora la satisfacción del cliente y fortalece la fidelización (Hasan, 2024; Loureiro et al., 2018).

Los enfoques tradicionales en la estimación de ventas incluyen métodos estadísticos como promedio móvil, suavizado exponencial (Wellens et al., 2024), ARIMA (Hasan, 2024; Lorente-Leyva et al., 2021b; Loureiro et al., 2018) y regresión lineal (Raizada & Saini, 2021), así como el juicio humano basado en experiencia y creatividad (DuBreuil & Lu, 2020). Sus principales limitaciones son la dependencia del instinto y el juicio subjetivo (DuBreuil & Lu, 2020) y las dificultades para capturar patrones no lineales o estacionales complejos (Hasan, 2024). En cambio, los enfoques basados en machine learning superan estas barreras al modelar

relaciones complejas y no lineales (Loureiro et al., 2018; Wellens et al., 2024), incorporar múltiples variables —como el clima, promociones, entre otros— (Raizada & Saini, 2021), manejar grandes volúmenes de datos de distinto tipo (Loureiro et al., 2018) y mejorar la precisión con respecto a los métodos tradicionales (Wellens et al., 2024), aunque requieren más recursos computacionales y una implementación técnica más compleja.

5. Modelos de pronóstico basados en machine learning

El paradigma en el pronóstico de las ventas ha evolucionado alineándose con las últimas tendencias de los enfoques basados en datos. En este trabajo se explorará, por un lado, un modelo estadístico clásico: la regresión lineal múltiple, como línea base, pues ha sido ampliamente utilizada por su simplicidad y facilidad de interpretación (Raizada & Saini, 2021). Por el otro lado, se explorará una categoría de técnicas de machine learning conocida como Modelos de Aprendizaje por Conjuntos, que combinan las predicciones de múltiples modelos para mejorar la robustez del pronóstico (Loureiro et al., 2018), en particular los siguientes enfoques basados en árboles de decisión:

- **Random Forest (RF):** Construye múltiples árboles de decisión durante el entrenamiento y utiliza el promedio de sus predicciones para la salida final (Hasan, 2024; Raizada & Saini, 2021). De forma que se crea un árbol en una muestra aleatoria de los datos y por medio de la agregación Bootstrap o 'bagging', se crea un gran número de árboles no correlacionados que de forma individual tienen alta varianza, pero unidos, sus errores tienden a cero. Su implementación es sencilla porque casi no requiere tuneado en los parámetros del modelo y es ampliamente usada en predicciones (Bolhuis & Rayner, 2020).
- **Gradient Boosting (GBT):** Entrena árboles de manera secuencial, usando los errores de predicción de los árboles anteriores para entrenar el siguiente, llegando a sumar cero en el último árbol. Mientras que RF combina árboles relativamente profundos con un sesgo bajo y una varianza alta, GBT combina

árboles relativamente poco profundos con un sesgo alto y una varianza baja (Bolhuis & Rayner, 2020). Chen & Guestrin (2016) introdujeron una versión optimizada llamada XGBoost, diseñada para ser altamente eficiente, flexible y portátil, características que lo hacen particularmente adecuado para problemas de forecasting con múltiples características temporales (Wellens et al., 2024).

Adicionalmente, se aprovechará el potencial de las herramientas del enfoque empírico Prophet, desarrollada por Facebook para la previsión de series temporales. Este enfoque se basa en un modelo aditivo donde tendencias no lineales se ajustan a los datos mediante descomposición en componentes de tendencia, estacionalidad y días festivos, teniendo la capacidad de incluir componentes adicionales como efectos de eventos especiales (Taylor & Letham, 2017).

A continuación, se presenta una tabla comparativa con los aspectos clave de cada modelo: su capacidad de modelado, manejo de estacionalidad, robustez ante el ruido, interpretabilidad y facilidad de uso.

Tabla 1. Comparativa entre los modelos considerados.

Criterio	Regresión Lineal	Random Forest	XGBoost	Prophet
Tipo de modelo	Estadístico clásico	Conjunto de árboles de decisión	Boosting de árboles de decisión	Modelo empírico de series temporales
Capacidad de modelado	Relaciones lineales simples	Relaciones no lineales y complejas	Relaciones no lineales, alta precisión	Tendencia, estacionalidad y eventos
Manejo de estacionalidad	Limitado (requiere ingeniería manual de variables)	Moderado (depende de ingeniería de datos)	Moderado (depende de ingeniería de datos)	Alta (incorpora automáticamente estacionalidades)

Robustez ante ruido	Baja	Alta (reduce sobreajuste mediante promedios)	Muy alta (usa regularización y corrección de errores)	Moderada (sensible a valores atípicos si no se ajusta)
Interpretabilidad	Alta	Media	Baja	Alta (representación clara de componentes de la serie)
Facilidad de uso	Alta	Media	Media-baja	Alta

6. Evaluación de modelos predictivos

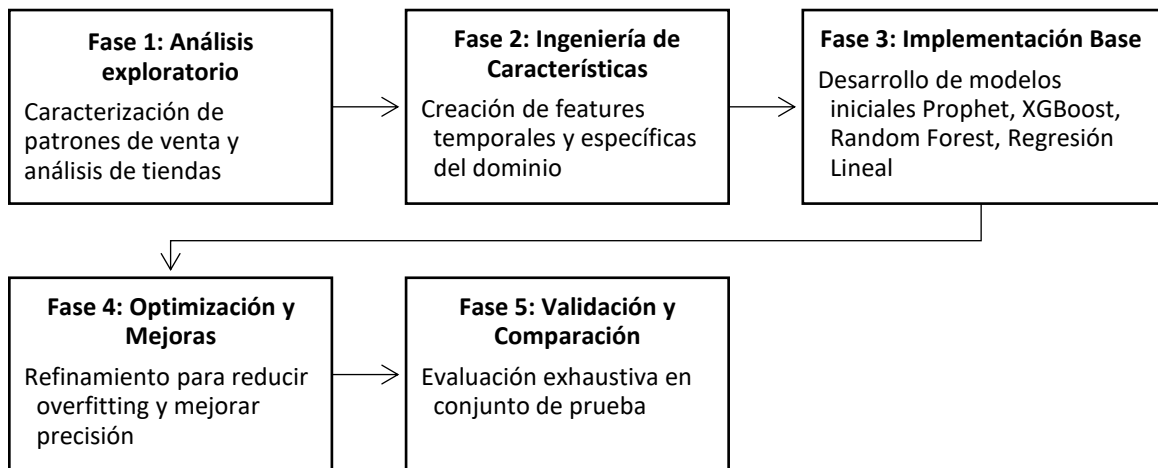
Las métricas comunes de evaluación para modelos de pronóstico, especialmente en el ámbito de estimación de ventas, incluyen:

- MAE (Mean Absolute Error): Mide el error promedio absoluto entre las predicciones y los valores reales; útil por su fácil interpretación y sensibilidad uniforme a los errores (Raizada & Saini, 2021).
- RMSE (Root Mean Squared Error): Calcula la raíz cuadrada del promedio de los errores al cuadrado; penaliza más los errores grandes y es ideal para comparar modelos de alta precisión (Lorente-Leyva et al., 2021b).
- MAPE (Mean Absolute Percentage Error): Expresa el error en porcentaje respecto al valor real; permite comparar la precisión entre diferentes series de datos (Lorente-Leyva et al., 2021b).
- R² (Coeficiente de Determinación): Indica qué proporción de la variabilidad en la variable dependiente es explicada por el modelo. Cuanto más cercano a 1, mejor el ajuste (Hasan, 2024).

DESARROLLO DEL TRABAJO

En esta sección se describe, paso a paso, el diseño y la implementación del sistema de pronóstico de ventas basado en machine learning que responde al objetivo general, en el cual se considera las unidades vendidas y los ingresos monetarios. El proceso trató los desafíos inherentes a los datos de ventas, incluyendo la alta variabilidad entre tiendas, los patrones estacionales complejos y la necesidad de capturar tendencias a largo plazo y fluctuaciones de corto plazo. En la Figura 1 se esquematiza las fases de desarrollo del sistema, las cuales se detallarán en esta sección.

Figura 1. Flujo de desarrollo del sistema.



Fase 1: Análisis Exploratorio y Caracterización de Datos

Se inicia con un análisis exhaustivo del comportamiento de ventas en 187 tiendas durante el período 2021-2025. El conjunto de datos cuenta con las variables definidas en la Tabla 2, las cuales permiten descomponer la dimensión temporal, identificar patrones estacionales y evaluar el impacto de variables externas como festividades y promociones.

Tabla 2. Descripción de las variables de la base de datos.

Variable	Descripción	Tipo de dato
Tienda	Identificador único de la tienda	Categórica nominal
FechaVenta	Fecha exacta de la venta	Temporal (fecha)
VtaUnds	Unidades vendidas	Numérica continua
PrecioVta	Precio unitario del producto	Numérica continua
mes	Mes de la venta (1 a 12)	Numérica discreta
dia_semana_nombre	Nombre del día de la semana	Categórica nominal
dia_semana_num	Número del día de la semana (1 = lunes, ..., 7 = domingo)	Numérica discreta
es_fin_de_semana	Indicador binario de fin de semana (1 = sí, 0 = no)	Binaria (categórica)
dia_mes	Día del mes (1 a 31)	Numérica discreta
anio	Año de la venta	Numérica discreta
semana_anio	Número de la semana en el año (1 a 52/53)	Numérica discreta
dia_anio	Día del año (1 a 365/366)	Numérica discreta
es_festivo	Indicador binario si es día festivo (1 = sí, 0 = no)	Binaria (categórica)
Oferta	Indicador binario de si hubo oferta/promoción (1 = sí, 0 = no)	Binaria (categórica)
Notas	Indica si se presentó algo que afecte las ventas (Paro, clima, violencia) (1 = sí, 0 = no)	Binaria (categórica)

Para caracterizar los datos se definieron cuatro pasos de modelado basados en el conocimiento del negocio: (1) coeficiente de variación (CV) para medir la volatilidad relativa de ventas, (2) análisis de tendencia mediante regresión lineal aplicada a ventanas móviles de 90 días, (3) estimación de tendencias y tasas de crecimiento,

y (4) detección de anomalías mediante el criterio de 3-sigma sobre ventanas móviles.

El CV se calcula como la razón entre la desviación estándar y la media de las ventas, proporcionando una medida estandarizada de variabilidad. Se definió como umbral crítico un $CV > 1.5$ para catalogar alta variabilidad, condición que se presentó en el 28% de las tiendas analizadas.

La tendencia fue analizada mediante tres métricas:

- Momentum: evalúa la diferencia entre las ventas actuales (V_T) y las ventas en un período de tiempo t anterior (V_{T-t}) o también ventas rezagadas, lo cual refleja si las ventas aumentan o disminuyen en ese plazo.
- Ratio de ventas promedio: Se construye de la forma $V_T/\text{Promedio}_t$ donde la venta actual se relaciona con el promedio en un período t . Este enfoque permite detectar desviaciones relativas a la tendencia en un periodo de tiempo.
- Tasa de crecimiento: se definió como el porcentaje de cambio entre las ventas actuales respecto a las ventas rezagadas, utilizando la fórmula:

$$\text{Tasa de crecimiento (\%)} = \frac{V_T - V_{T-t}}{V_{T-t}} \times 100$$

Para la detección de anomalías, se consideraron como atípicas aquellas observaciones de ventas que se encontraban fuera del rango definido por:

$$rolling_{mean} \pm 3 \times rolling_{std}$$

dentro de una ventana móvil de siete unidades de tamaño (correspondiente a los días de la semana). A partir de este criterio, se calculó un índice de anomalía por tienda como la proporción entre el número de días con ventas anómalas y el total de registros. Este análisis permitió identificar que el 15% de las tiendas presentaban tasas de anomalía superiores al 5%, lo cual fue determinante para la posterior parametrización adaptativa en la fase de modelado predictivo.

Fase 2: Ingeniería de Características

Durante esta etapa se desarrolló un conjunto de variables derivadas, complementando los 15 campos originales del dataset, con el objetivo de mejorar la capacidad predictiva de los modelos. La creación de features siguió un enfoque para intentar capturar la complejidad temporal y contextual de las ventas, al evitar el sobreajuste mediante el control de la multicolinealidad. Durante esta etapa se implementaron más de siete categorías principales, tomando en cuenta el contexto del dominio y las mejores prácticas estadísticas.

1. **Características temporales:** Estas características se dividieron en dos segmentos: básicos y cíclicos. Las características temporales básicas ya estaban descritas en su mayoría en la base de datos original, y se decide agregar una variable de días desde el inicio de la campaña, un indicador de trimestre y otro de quincena. Mientras que para las características temporales cíclicas se utilizó una transformación trigonométrica mediante funciones seno y coseno para codificar variables como el mes del año. Esta codificación permite capturar la naturaleza periódica de la variable sin introducir discontinuidades, e incluye ciclos mensuales y semanales.
2. **Lags adaptativos:** Se incorporaron rezagos (lags) en ventanas de 1, 7, 14 y 28 días de la variable objetivo, permitiendo que el modelo tenga acceso a patrones históricos recientes de ventas. Estos rezagos fueron cuidadosamente ajustados para evitar la filtración de datos futuros, y se aplicaron filtros para gestionar valores faltantes.
3. **Medias móviles ponderadas:** Se utilizaron medias móviles (MA) y medias móviles exponenciales (EMA) para captar tendencias recientes en las ventas (con ventanas de 7, 14, 28 y 30), asignando mayor peso a las observaciones más cercanas en el tiempo. Estas medias fueron desplazadas temporalmente para evitar la filtración de información futura en el proceso de entrenamiento.

4. Poder adquisitivo dinámico: Se modeló el efecto de los días de pago sobre el consumo mediante una función logística suave que asigna mayor poder adquisitivo a los días cercanos a fechas comunes de pago (1, 15 y 30 del mes). Esto permitió modelar la influencia de la liquidez en el comportamiento de compra.
5. Características de tienda: Se incluye el cálculo de CV para cada tienda. Se realiza también una categorización de las tiendas según el tamaño dado por el número de ventas (pequeña, mediana y grande), y se calcularon estadísticos básicos de la media, mediana y desviación estándar para las unidades vendidas y el precio de venta.
6. Tendencias de crecimiento: Usando la ecuación de Tasa de Crecimiento, fue calculada para $t = \{7, 28\}$ días y anual. Se calculó también el momentum a $t = \{1, 3, 7\}$ en unidades de días. Además del ratio de venta promedio, usando un promedio histórico, y un promedio de ventanas móviles a 7 y 28 días.
7. Otros: Entre otros se encuentran las categorías de volatilidad, patrones especiales e interacciones, que agrupan variables diseñadas para capturar comportamientos no lineales y efectos específicos en la demanda. Las variables de volatilidad (como desviaciones estándar móviles o el coeficiente de variación) miden la inestabilidad de las ventas en ventanas recientes, reflejando entornos de alta incertidumbre. Los patrones especiales incluyen indicadores similares a "es_festivo", cercanía a quincenas o fines de mes, y efectos post-pago, los cuales influyen en el comportamiento de compra. Finalmente, las interacciones combinan variables relevantes (por ejemplo, fin de semana con quincena), permitiendo modelar efectos compuestos que impactan la demanda de manera diferenciada.

El conjunto total incluye 83 características generadas con este enfoque de ingeniería de datos. Estas se distribuyen entre los modelos según su aplicabilidad y compatibilidad. La Tabla 3 muestra el número de variables empleadas por cada

modelo, lo que permite entender la carga de información procesada en cada enfoque.

Tabla 3. Cantidad de características por modelo.

Modelo	Numero de características
Prophet	28
XGBoost	60
Random Forest	45
Regresión Lineal	16

Desde el punto de vista técnico, se aplicaron algunos criterios para la reducción de dimensionalidad con el fin de evitar la multicolinealidad. Se empleó el Factor de Inflación de Varianza (VIF), manteniendo valores inferiores a 10, y se eliminaron iterativamente variables hasta cumplir este umbral, especialmente en el modelo de regresión lineal. Asimismo, se controló que las correlaciones entre variables no superara 0.85 y que se preservara al menos el 80% de la varianza explicada tras la selección de características. Este proceso redujo el conjunto de variables de 45 a 28, mejorando la interpretabilidad sin comprometer el desempeño predictivo.

Fase 3: Implementación base

En estas fases se desarrollaron cuatro arquitecturas complementarias de modelado, cada una diseñada para capturar distintas dimensiones del problema, teniendo en cuenta que no existe un único modelo óptimo que aporte una solución integral. De esta forma, equilibrando precisión, generalización, robustez y explicabilidad.

Prophet

Se configuró el modelo Prophet con múltiples componentes estacionales (anual, mensual, semanal y quincenal) y algunos ajustes adaptativos según la tienda. Para tiendas con alta variabilidad ($CV > 1.5$), se incrementó la flexibilidad del modelo ajustando la cantidad de puntos de cambio y su penalización. En casos de alta inestabilidad ($CV > 2.0$), se aplicó una transformación de raíz cuadrada a la variable

objetivo para estabilizar la varianza. Además, se incorporó una corrección de residuos utilizando Gradient Boosting, y se aplicó un posprocesamiento con límites basados en percentiles históricos (hasta un 20% sobre el percentil 99) para mitigar outliers. Para tiendas con anomalías frecuentes, se combinó el pronóstico con promedios históricos mediante técnicas de ensamblado.

XGBoost

Este modelo se entrenó mediante una búsqueda sistemática de combinaciones de hiperparámetros orientada a minimizar el riesgo de sobreajuste. Se mejoró la calidad del entrenamiento aumentando el tamaño del conjunto de validación al 20% y utilizando imputación con la mediana para valores los faltantes. La selección de características fue guiada por medidas de información mutua, y se aplicaron correcciones dinámicas en las predicciones, restringiendo las variaciones dentro de un rango controlado (factores multiplicativos entre 0.85 y 1.3) según la dispersión histórica.

Random Forest

Se implementó un proceso de optimización basado en búsqueda aleatoria con énfasis en minimizar la brecha de desempeño entre los conjuntos de entrenamiento y prueba. En casos donde el gap de R^2 superaba 0.15, se endurecieron los parámetros del modelo (como la profundidad máxima de los árboles y los mínimos de muestras por nodo) con el fin de controlar el sobreajuste.

Regresión Lineal Regularizada

La regresión lineal se enriqueció mediante la transformación de Box-Cox sobre la variable objetivo, para cumplir los supuestos de normalidad y homocedasticidad. Se aplicó una eliminación sistemática de variables con alta multicolinealidad utilizando el criterio de VIF, y se incorporó regularización Ridge con penalización baja ($\alpha = 0.01$) para estabilizar los coeficientes.

Cada algoritmo aborda aspectos distintos: Prophet se destaca en la captura de estacionalidades complejas, XGBoost por tener un modelado no lineal de alta precisión, Random Forest en su robustez frente al ruido y la Regresión Lineal por tener una mayor explicabilidad. Este enfoque combinado permite aprovechar las fortalezas individuales de cada modelo para lograr un sistema de pronóstico más sólido y confiable.

Fase 4: Optimización y Mejoras

Se implementó un framework de validación riguroso para asegurar la calidad y generalización de los modelos antes de su evaluación final. Para esto, se estableció un corte temporal fijo en diciembre 2024, reservando enero 2025 como conjunto de prueba completamente no visto.

Con el fin de controlar el sobreajuste se implementaron varias estrategias teniendo en cuenta la naturaleza de cada modelo:

- Prophet: Se implementó el modelo de corrección de residuos mediante Gradient Boosting, además de una parametrización adaptativa según las características de cada tienda (CV, presencia de anomalías), y una validación mediante backtesting en ventanas móviles.
- XGBoost: Se realizó una búsqueda sistemática de hiperparámetros con penalización por gap train-validation, junto con un aumento del conjunto de validación al 20%, y early stopping basado en métricas de validación.
- Random Forest: Se incorporó una validación out-of-bag (OOB), también un ajuste dinámico de parámetros de regularización cuando se detecta sobreajuste, y se limitó la profundidad máxima adaptativa.
- Regresión Lineal: Se eliminó la multicolinealidad ($VIF < 10$), además de la regularización Ridge ya mencionada, y validación cruzada para selección de α . Finalmente se diagnosticaron los supuestos del modelo.

Fase 5: Validación y Comparación

En esta fase se establecieron los protocolos y controles necesarios para asegurar que los modelos desarrollados pudieran ser evaluados correctamente, mediante el conjunto de métricas: MAE, RMSE, R^2 , MAPE, y SMAPE. Adicionalmente, se diseñaron análisis segmentados por categoría de tienda, rango de ventas, y período temporal.

Así mismo se aplicaron diversos diagnósticos estadísticos, como la normalidad y homocedasticidad de los residuos mediante las pruebas Shapiro-Wilk y Breusch-Pagan, principalmente para validar los supuestos de modelos lineales. La autocorrelación de los errores se analizó usando el estadístico de Durbin-Watson, para identificar dependencias temporales no capturadas. Además, se inspeccionaron patrones en los errores de predicción para detectar sesgos persistentes. Finalmente, se evaluó la estabilidad temporal de los parámetros del modelo mediante validaciones por ventanas móviles, asegurando que su comportamiento fuera consistente a lo largo del tiempo.

RESULTADOS

Esta sección presenta los resultados obtenidos en la implementación y evaluación de los cuatro modelos de pronóstico de ventas desarrollados. Se analizan las métricas de desempeño, los patrones de error y las características de cada aproximación.

7. Desempeño global de los modelos

La evaluación de los modelos se realizó sobre el conjunto de prueba correspondiente a enero 2025, conservando las mismas métricas para poder comparar directamente entrenamiento y prueba. La Tabla 4 presenta las métricas principales alcanzadas por cada modelo.

Tabla 4. Métricas de desempeño comparativo de los modelos implementados

Modelo	R² Unidades	MAE Unidades	RMSE Unidades	MAPE Unidades	R² Pesos	MAE Pesos
Prophet	0.7121	7.99	11.43	35.2%	0.6060	\$828,153
XGBoost	0.6547	9.17	13.82	42.5%	0.5894	\$985,432
Random Forest	0.5437	9.45	14.37	56.3%	0.5305	\$882,365
Regresión Lineal	0.4979	11.44	17.61	52.3%	0.5142	\$1,037,593

8. Análisis por modelo

Prophet

El modelo Prophet obtuvo el mejor desempeño global entre las arquitecturas evaluadas, alcanzando un $R^2 = 0.7121$ para unidades y $R^2 = 0.6060$ para pesos, lo que evidencia su solidez tanto en volumen como en valor monetario. Esto es especificado en la Tabla 5.

Tabla 5. Métricas globales del modelo Prophet

Métrica	Unidades	Pesos
MAE	7.99	\$ 828,153
RMSE	11.43	\$ 1,186,103
R ²	0.7121	0.606
MAPE	35.2% (≥5 uds)	34.6% (≥\$765,364)
SMAPE	32.50%	31.60%
Sesgo	0.00 (0.0%)	\$-13,895 (-0.5%)

Su capacidad para manejar estacionalidades múltiples —incluyendo patrones anuales, mensuales, semanales y quincenales— permitió capturar la complejidad

cíclica del entorno. Además, la parametrización adaptativa por tienda (basada en CV y presencia de anomalías), contribuyó a mejorar la precisión en contextos de alta variabilidad. Un elemento clave fue el modelo de corrección de residuos con Gradient Boosting, que redujo el sesgo a cero en las predicciones unitarias.

Así mismo, se realiza un análisis por categoría de tienda (ver Tabla 6), el cual mostró una diferencia en el rendimiento según el tamaño de tienda. Las tiendas grandes alcanzaron un MAPE del 27.2%, frente al 34.7% en medianas y 44.3% en pequeñas. Esto sugiere que la mayor estabilidad y volumen de datos en tiendas grandes favorece una mejor capacidad predictiva, mientras que las pequeñas, al presentar mayor volatilidad, siguen representando un reto para los modelos.

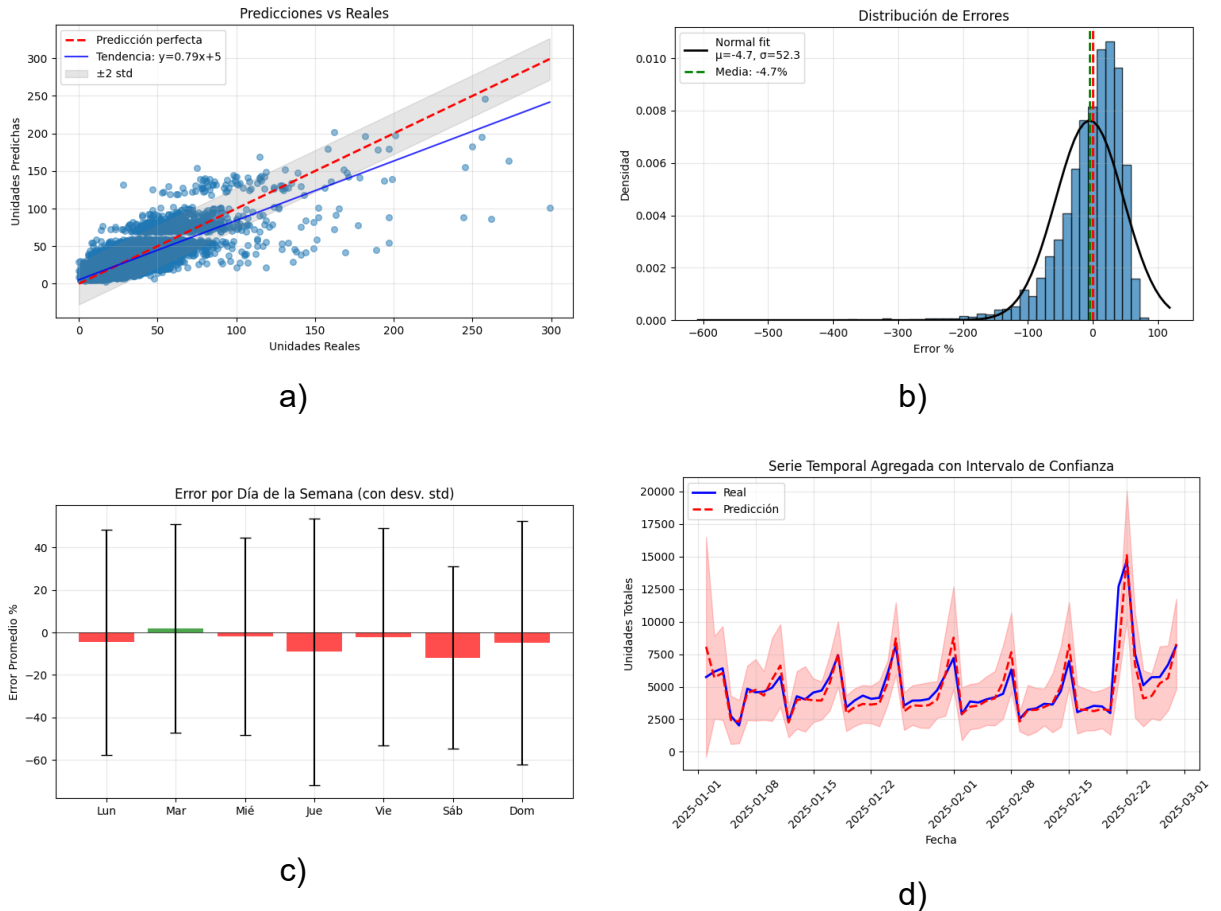
Tabla 6. Análisis por categoría de tienda (unidades)

Categoría de Tienda	MAE	Sesgo	MAPE
Pequeñas	7	2.5	44.30%
Medianas	7.6	0.4	34.70%
Grandes	9.4	-2.9	27.20%

XGBoost

El modelo XGBoost alcanzó un notable desempeño en la fase de validación, logrando un coeficiente de determinación $R^2 = 0.6547$ y una diferencia de rendimiento entre entrenamiento y prueba de 0.1746, lo cual evidencia un control efectivo del sobreajuste debido a las estrategias implementadas. La optimización de hiperparámetros permitió determinar que una profundidad máxima de 6 y una tasa de aprendizaje de 0.02 eran los más adecuados. Entre las variables más influyentes se destacaron *mes_ratio* (25.66%) e *is_peak* (25.04%), confirmado así la importancia de los factores estacionales y la detección de picos de demanda. Además, se incorporaron ajustes dinámicos a las predicciones, aplicando factores de corrección entre 0.85 y 1.14 para mejorar la precisión en distintos rangos de venta, particularmente en casos de subestimación.

Figura 2. Análisis de desempeño del modelo XGBoost



La Figura 2 ofrece un análisis visual del desempeño del modelo. En la *Figura 2.a)* se observa un gráfico de dispersión entre valores predichos y reales, cuya tendencia (línea $y = 0.79x + 5$) refleja una leve subestimación, enmarcada dentro de una banda de ± 2 desviaciones estándar. En la *Figura 2.b)* se muestra la distribución de errores porcentuales con un ajuste normal ($\mu = -4.7$, $\sigma = 52.3$), revelando una leve asimetría negativa y colas pesadas. En la *Figura 2.c)* se reporta el error promedio según el día de la semana, con una subestimación del 5% al 10%, excepto los martes donde hay errores positivos. En general se observa alta dispersión de los errores, en especial el jueves que fue el día con mayor variabilidad. Finalmente, en la *Figura 2.d)* se presenta la serie temporal agregada, en la que se aprecia, de forma

general, un buen seguimiento de los patrones, aunque con cierta dificultad para capturar los picos extremos, como el observado el 22 de febrero.

Random Forest

El modelo Random Forest alcanzó un $R^2 = 0.5437$ en la predicción de unidades, evidenciando un desempeño moderado con ciertas fortalezas clave. El control del sobreajuste fue eficaz, demostrado por la baja diferencia entre el rendimiento en entrenamiento ($R^2 = 0.8754$) y validación ($R^2 = 0.7432$), todo esto gracias al ajuste dinámico de hiperparámetros cuando el gap superaba el umbral de 0.15. Además, el Out-of-Bag (OOB) score de 0.8150 confirmó una buena capacidad de generalización interna. En la Tabla 7 se evidencia las métricas con este control de sobre ajuste y sin control.

Tabla 7. Métricas del modelo Random Forest (Unidades)

Corrección	MAE	RMSE	R²	MAPE	Sesgo
Sin corrección	9.45	14.36	0.5441	56.30%	-0.94 (-3.4%)
Con corrección	9.45	14.37	0.5437	56.30%	-0.94 (-3.3%)

En cuanto a la importancia de variables, “mes_ratio” fue la más influyente (28.71%), seguida por las medias móviles de ventas en una ventana de 7 días (13.29%), lo que resalta la relevancia de los patrones estacionales y de tendencia. No obstante, el modelo presentó un sesgo negativo constante del -3.3% y un MAPE elevado (56.3%), lo cual puede sugerir dificultades para capturar los valores extremos de demanda, a pesar de la robustez inherente del enfoque ensemble. Estos resultados indican que, aunque Random Forest puede ser confiable en escenarios con variabilidad moderada, requiere un refuerzo adicional en el manejo de picos y comportamientos extremos en las series temporales.

Por otro lado, el análisis por categoría de tienda, descrito en la Tabla 8, reveló que las tiendas grandes presentaron el mejor desempeño con un MAPE de 47.8%,

seguidas por las tiendas medianas con 51.5%, mientras que las tiendas pequeñas alcanzaron un MAPE considerablemente más alto de 69.8%. Esta variabilidad sugiere que el modelo es más eficaz en entornos con mayor volumen y estabilidad de ventas, como ocurre en tiendas grandes, donde los patrones son más consistentes. En contraste, la mayor volatilidad y menor densidad de datos en tiendas pequeñas dificultan la predicción. Esto resaltando resalta la necesidad de enfoques diferenciados según el tamaño de la tienda.

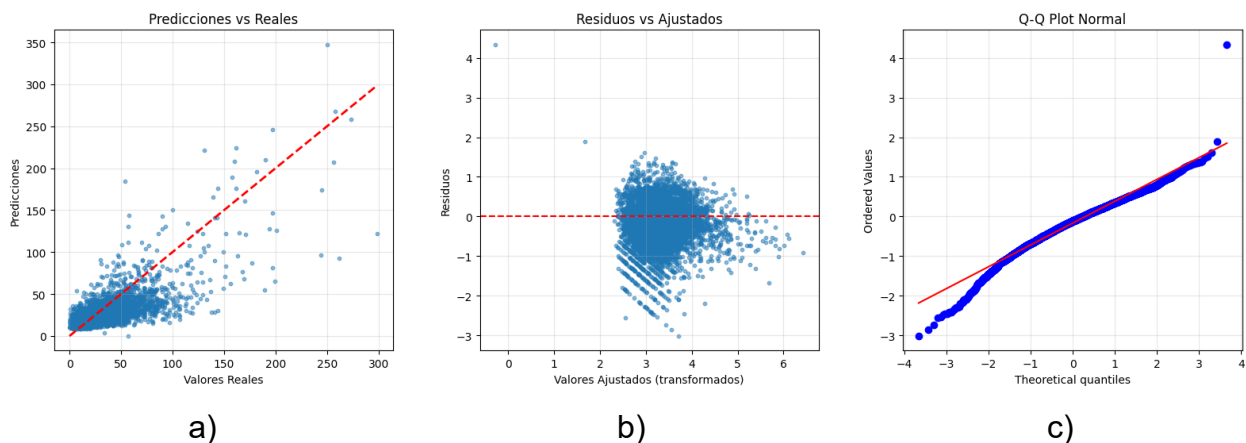
Tabla 8. Análisis por categoría de tienda (Unidades)

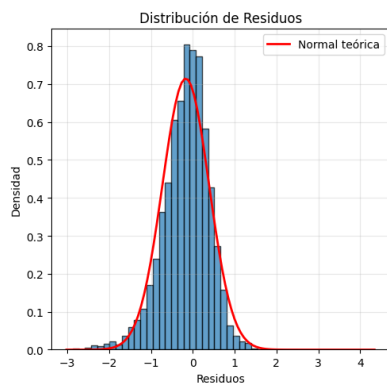
Categoría de Tienda	MAE	Sesgo	MAPE
Pequeñas	7.3	0.4	69.80%
Medianas	9	1.2	51.50%
Grandes	12	1.2	47.80%

Regresión lineal

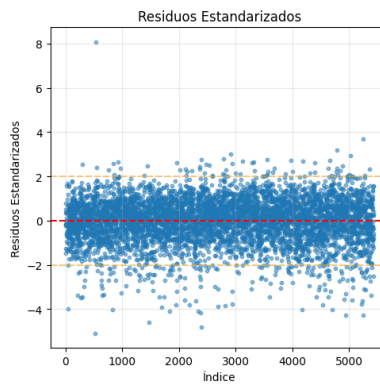
El modelo de Regresión Lineal regularizada obtuvo un $R^2 = 0.4979$, el más bajo entre los modelos evaluados. A pesar de su menor capacidad predictiva, este modelo sirvió como base para el análisis de errores y la validación de supuestos clásicos.

Figura 3. Diagnósticos del modelo de Regresión Lineal

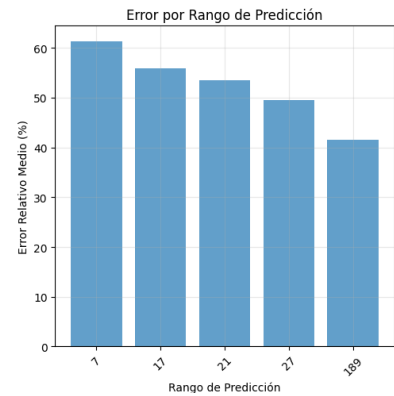




d)



e)



f)

La Figura 3 presenta un conjunto detallando los diagnósticos. En la *Figura 3.a*), se observa la dispersión de las predicciones frente a los valores reales en escala original, mostrando un ajuste general a la tendencia pero bastante dispersión. El gráfico de la *Figura 3.b*), realizado en el espacio transformado, muestra una varianza aproximadamente constante, lo que sugiere homocedasticidad. El Q-Q plot de la *Figura 3.c*) revela desviaciones respecto a la normalidad, sobre todo en los extremos. El histograma de residuos (*Figura 3.d*)), con una curva normal superpuesta, corrobora la no normalidad, consistente con los resultados de las otras pruebas estadísticas. En la *Figura 3.e*), los residuos estandarizados frente al índice temporal muestran algunos valores atípicos aislados, sin patrones sistemáticos. Por último, el error relativo por rango de predicción (*Figura 3.f*)) indica mayores errores en niveles bajos de ventas.

Estas observaciones son respaldadas con las pruebas realizadas: el estadístico Durbin-Watson fue de 1.90, lo que sugiere ausencia de autocorrelación; la correlación entre los valores ajustados y el cuadrado de los residuos fue de -0.087, lo cual apoya la tesis de la homocedasticidad; y la prueba de Shapiro-Wilk arrojó un valor de $p < 0.001$, lo que implica rechazo de la hipótesis de normalidad en los residuos.

ANÁLISIS DE RESULTADOS

El análisis global del desempeño de los modelos reveló patrones en el comportamiento predictivo. En primer lugar, se observó una tendencia generalizada

a subestimar los picos de demanda, especialmente durante días festivos y promociones. Este comportamiento fue analizado también por Xu et al. (2021), donde se menciona que durante las promociones se aumentan drásticamente las ventas y esta variación puede no ser capturada por los modelos, tendiendo a la subestimación. Por ello los autores proponen una función de pérdida consistente de la promoción (*promotion-aware loss function*) para mejorar el rendimiento del pronóstico durante estos periodos.

En segundo lugar, los fines de semana presentaron errores generalizados, con una subestimación promedio del 5 al 10 %, más evidente en modelos como XGBoost. El artículo (Hasan, 2024) también sugiere dificultades sistemáticas en la predicción de valores altos, los cuales se presentan principalmente los fines de semana y festivos.

En cuanto al tamaño de tienda, los modelos mostraron un desempeño consistentemente superior en tiendas grandes frente a pequeñas, con diferencias de MAPE de hasta 20 puntos porcentuales. Por ejemplo, Prophet alcanzó un 27.2% en tiendas grandes frente a 44.3% en pequeñas, y Random Forest mostró 47.8% frente a 69.8%, respectivamente. Este fenómeno se ha evidenciado en otros estudios, que consideran el tamaño de la tienda como una variable importante (Loureiro et al., 2018; Raizada & Saini, 2021; Xu et al., 2021).

Además, se identificó una mayor dificultad en la predicción de ventas en pesos que en unidades, con valores de R^2 más bajos (0.51–0.61), lo cual refleja la complejidad adicional introducida por la variabilidad de los precios.

Desde el punto de vista de calidad y validación, los modelos superaron criterios importantes. Por ejemplo, los gaps entre entrenamiento y prueba por debajo de 0.20 en todos los casos demuestran un buen control del sobreajuste. La estabilidad temporal se confirmó mediante validación en un periodo futuro (enero de 2025), sin degradaciones críticas de desempeño. Asimismo, los modelos que incorporaron

estrategias específicas, como Prophet y XGBoost, demostraron una mayor robustez ante anomalías y outliers.

En conclusión, Prophet se posiciona como el modelo más adecuado para una implementación práctica, al combinar alta precisión predictiva con una interpretación clara de las componentes estacionales. XGBoost surge como una opción potente para escenarios con relaciones no lineales complejas, mientras que Random Forest aporta robustez en contextos de alta incertidumbre. Finalmente, aunque la Regresión Lineal es la menos precisa, sigue siendo útil como modelo base por su interpretabilidad y capacidad para diagnosticar la estructura del problema.

CONCLUSIONES

El presente trabajo logró desarrollar e implementar exitosamente un sistema integral de predicción de ventas basado en técnicas de machine learning (i.e. Prophet, XGBoost, Random Forest y Regresión Lineal), respondiendo directamente a la necesidad del sector textil de anticipar la demanda para optimizar tanto la gestión de inventarios como la planificación operativa. Los resultados obtenidos demuestran que es posible superar las limitaciones de los métodos tradicionales mediante la aplicación de algoritmos que permiten capturar la complejidad inherente a los patrones de venta en el sector textil. Esto queda demostrado en la mejora del 43% para la predicción de las ventas en unidades con Prophet (que logró un R^2 de 0.71), en contraste de la regresión lineal ($R^2 = 0.49$).

La investigación reveló que la integración efectiva de factores múltiples (promociones, estacionalidad, días festivos, fines de semana y ciclos de pago quincenal) requiere enfoques diferenciados según la naturaleza del algoritmo. Prophet destacó por su capacidad nativa de modelar estacionalidades complejas superpuestas, mientras que XGBoost sobresalió en la captura de interacciones no lineales entre variables. La implementación de correcciones adaptativas según las características de cada tienda (tamaño, variabilidad, presencia de anomalías)

resultó fundamental para mejorar la predicción, reduciendo el MAPE promedio de 52.3% en regresión lineal a 35.2% en Prophet.

Los resultados obtenidos en este estudio tienen importantes implicaciones para la gestión operativa del sector textil colombiano. La mejora en la precisión de los modelos, con una reducción del error de predicción de hasta 17 puntos porcentuales, permite optimizar los inventarios, reducir costos logísticos y evitar pérdidas por desabastecimiento. Además, facilita una presupuestación más realista por tienda, lo que mejora la planificación y motiva al personal al fijar metas alcanzables. La segmentación por tamaño de tienda reveló que las tiendas pequeñas presentan mayor volatilidad, lo cual justifica estrategias diferenciadas con mayores márgenes de seguridad. También se fortaleció la capacidad para anticipar picos de demanda, especialmente en fechas clave como quincenas y fines de semana post-pago, lo que mejora la planificación logística y operativa.

Sin embargo, persisten desafíos. La predicción en valores monetarios sigue siendo menos precisa que en unidades, lo que sugiere la necesidad de incorporar modelos de pricing. Además, los eventos extremos aún tienden a ser subestimados, por lo que se recomienda explorar enfoques híbridos y jerárquicos que combinen múltiples algoritmos y capturen tanto patrones globales como locales.

Para una implementación efectiva se recomienda adoptar Prophet como modelo base, por su equilibrio entre precisión e interpretabilidad, complementado con XGBoost en decisiones críticas. También se sugiere establecer actualizaciones mensuales del sistema, integrar dashboards con intervalos de confianza y capacitar al personal en el uso de las predicciones como una herramienta de apoyo.

En síntesis, este trabajo demuestra que el uso adecuado de técnicas de machine learning puede transformar la capacidad predictiva del sector textil, fortaleciendo la toma de decisiones, reduciendo la incertidumbre y facilitando una gestión más ágil, eficiente y basada en datos.

REFERENCIAS

- Bolhuis, M. A., & Rayner, B. (2020). *Deus ex Machina? A Framework for Macro Forecasting with Machine Learning* (Working Paper No. 2020/045). <https://www.imf.org/en/Publications/WP/Issues/2020/02/28/Deus-ex-Machina-A-Framework-for-Macro-Forecasting-with-Machine-Learning-49094>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Colombia Productiva. (2019). *Pacto por el crecimiento y para la generación de empleo del sector moda* (p. 21). <https://www.colombiaproductiva.com/ptp-capacita/publicaciones/pactos-por-el-crecimiento/pacto-por-el-crecimiento-y-para-la-generacion-8/>
- DANE. (2024). *Encuesta Mensual de Comercio - EMC - 2020 - 2023*. Departamento Administrativo Nacional de Estadística. <https://microdatos.dane.gov.co/index.php/catalog/829>
- DANE. (2025). *Boletín técnico Producto Interno Bruto (PIB)* (p. 47). DANE. <https://www.dane.gov.co/files/operaciones/PIB/bol-PIB-ltrim2025.pdf>
- DuBreuil, M., & Lu, S. (2020). Traditional vs. big-data fashion trend forecasting: an examination using WGSN and EDITED. *International Journal of Fashion Design, Technology and Education*, 13(1), 68-77. <https://doi.org/10.1080/17543266.2020.1732482;CTYPE:STRING:JOURNAL>
- Espinel González, P. A., Aparicio Soto, D. M., & Mora, A. J. (2018). SECTOR TEXTIL COLOMBIANO Y SU INFLUENCIA EN LA ECONOMÍA DEL PAÍS. *Punto de vista*, 9(13). <https://doi.org/10.15765/PDV.V9I13.1118>
- Hasan, M. R. (2024). Addressing Seasonality and Trend Detection in Predictive Sales Forecasting: A Machine Learning Perspective. *Journal of business and*

management studies, 6(2), 100-109.
<https://doi.org/10.32996/JBMS.2024.6.2.10>

Lorente-Leyva, L. L., Alemany, M. M. E., Peluffo-Ordóñez, D. H., & Araujo, R. A. (2021a). Demand Forecasting for Textile Products Using Statistical Analysis and Machine Learning Algorithms. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12672 LNAI, 181-194. https://doi.org/10.1007/978-3-030-73280-6_15/TABLES/5

Lorente-Leyva, L. L., Alemany, M. M. E., Peluffo-Ordóñez, D. H., & Araujo, R. A. (2021b). Demand Forecasting for Textile Products Using Statistical Analysis and Machine Learning Algorithms. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12672 LNAI, 181-194. https://doi.org/10.1007/978-3-030-73280-6_15/TABLES/5

Loureiro, A. L. D., Miguéis, V. L., & da Silva, L. F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, 114, 81-93. <https://doi.org/10.1016/J.DSS.2018.08.010>

MINCIT. (2020). *Gobierno e industria de la moda invitan a comprar lo nuestro*. [mincit.gov.co](https://www.mincit.gov.co). <https://www.mincit.gov.co/prensa/noticias/industria/gobierno-sector-moda-invitan-a-comprar-lo-nuestro>

Oliveros Contreras, D., Guzman Duque, A. P., Mendoza García, E. M., & Blanco Alvarez, T. A. (2019). *La eficiencia y productividad del sector textil y confecciones en Colombia* (Universida). https://www.cpae.gov.co/sites/default/files/documentos/2022-09/EFICIENCIA_TEXTIL.pdf

Raizada, S., & Saini, J. R. (2021). Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting. *International Journal of Advanced*

Computer Science and Applications, 12(11), 102-110.
<https://doi.org/10.14569/IJACSA.2021.0121112>

Ren, S., Chan, H. L., & Siqin, T. (2020). Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. *Annals of Operations Research*, 291(1-2), 761-777. <https://doi.org/10.1007/S10479-019-03148-8/FIGURES/1>

Rodríguez-León, J., & Pachón-Rincón, M. (2021). Estudio de pronóstico para la planeación, caso de estudio empresa distribuidora del sector farmacéutico. *Revista UIS Ingenierías*, 20(4), 59-78.
<https://doi.org/10.18273/REVUIN.V20N4-2021005>

Taylor, S. J., & Letham, B. (2017). Forecasting at scale. *PeerJ Preprints*.
<https://doi.org/10.7287/PEERJ.PREPRINTS.3190V2>

Wellens, A. P., Boute, R. N., & Udenio, M. (2024). Simplifying tree-based methods for retail sales forecasting with explanatory variables. *European Journal of Operational Research*, 314(2), 523-539.
<https://doi.org/10.1016/J.EJOR.2023.10.039>

Xu, J., Zhou, Y., Zhang, L., Wang, J., & Lefloch, D. (2021). Sportswear retailing forecast model based on the combination of multi-layer perceptron and convolutional neural network. *Textile Research Journal*, 91(23-24), 2980-2994.
<https://doi.org/10.1177/00405175211020518>