Master in Management + Analytics

De ARIMA a TimeGPT: Predicción de demanda en la industria retail de alimentos

Autoría: Spadaro, Juan
Año: 2025

UNIVERSIDAD TORCUATO DI TELLA

Master in Management + Analytics

De ARIMA a TimeGPT:

Predicción de demanda en la industria retail de alimentos

TESIS

Juan Spadaro

Abril 2025

Tutora: Magdalena Cornejo

Resumen

Esta tesis busca comparar distintos modelos de forecasting (ARIMA, XGBoost, Prophet y TimeGPT) en términos de precisión, interpretabilidad y flexibilidad, con el objetivo de identificar cuál proporciona la mejor combinación de estos factores, generando además ahorro económico para el caso de estudio en cuestión. La investigación está motivada por la importancia de anticipar la demanda en el sector retail de alimentos caracterizada por su volatilidad, para fortalecer y volver más eficiente el proceso de compras a proveedores, mediante optimización de inventarios y reducción de costos operativos. Se trabajará con datos históricos de ventas, inventarios y clientes, realizando un análisis comparativo que incluye el ajuste de modelos, análisis de su desempeño con métricas clave y evaluación de su impacto económico. Además, se aplicarán métodos de simulación para medir su robustez ante cambios abruptos en la demanda y se incluirá un análisis cualitativo mediante entrevistas en el sector. La tesis ofrece una perspectiva integral para la toma de decisiones estratégicas en el ámbito del retail y venta al por mayor, integrando enfoques cuantitativos y cualitativos para optimizar la eficiencia y adaptación al mercado.

Abstract

This thesis seeks to compare different forecasting models (ARIMA, XGBoost, Prophet and TimeGPT) in terms of accuracy, interpretability and flexibility, in order to identify which one provides the best combination of these factors, generating economic savings for the case study in question. The research is motivated by the importance of anticipating demand in the food retail sector, characterized by its volatility, in order to strengthen and make more efficient the process of purchasing from suppliers, by optimizing inventories and reducing operating costs. We will work with historical data on sales, inventories and customers, performing a comparative analysis that includes the adjustment of models, analysis of their performance with key metrics and evaluation of their economic impact. In addition, simulation methods will be applied to measure their robustness to abrupt changes in demand and a qualitative analysis will be included through interviews in the sector. The thesis offers a comprehensive perspective for strategic decision making in retail and wholesale, integrating quantitative and qualitative approaches to optimize efficiency and market adaptation.

Índice de contenido:

Índice de tablas

Índice de gráficos:

1 Introducción

1.1 Contexto y Justificación

La predicción de demanda ha ganado una relevancia significativa en el sector retail y de venta al por mayor, particularmente en la industria de alimentos, donde la volatilidad de los mercados y las fluctuaciones en el comportamiento de los consumidores representan grandes desafíos. Empresas de todos los tamaños enfrentan la necesidad de optimizar sus inventarios, mejorar la eficiencia de la cadena de suministro y reducir costos, especialmente en un contexto donde los márgenes de error en la planificación pueden impactar de manera significativa en las finanzas. En este contexto, la capacidad para anticipar la demanda con precisión y adaptarse a cambios inesperados en las condiciones del mercado es esencial para lograr una ventaja competitiva.

La literatura existente sobre modelos de predicción de demanda resalta diferentes enfoques, cada uno con sus fortalezas y limitaciones. ARIMA, un modelo tradicional en el análisis de series temporales, ha demostrado ser efectivo en datos lineales con patrones claros de estacionalidad, pero su capacidad para capturar relaciones complejas es limitada. Por otro lado, modelos más avanzados como XGBoost, que permiten la incorporación de múltiples variables, ofrecen un mejor rendimiento en términos de precisión, aunque suelen ser más difíciles de interpretar. Prophet, desarrollado por Facebook, proporciona una gran flexibilidad en casos de datos irregulares o con eventos especiales, como días festivos, pero su capacidad para ajustarse a cambios abruptos en la demanda sigue siendo un área de estudio. Finalmente, TimeGPT(Hochreiter, 1997; Simmhan & Ranganathan, 2023), una herramienta emergente basada en modelos generativos, ha mostrado un potencial significativo en capturar patrones no lineales y escalables en grandes volúmenes de datos, lo que sugiere que puede ser una opción adecuada para escenarios en constante evolución.

Esta tesis se enmarca en la creciente necesidad de comparar y evaluar estas herramientas desde un enfoque integral que no solo priorice la precisión predictiva, sino que también considere la interpretabilidad de los resultados y su aplicabilidad práctica en términos de ahorro económico y flexibilidad. Al basarse en datos reales de la industria retail de alimentos, se pretende generar conocimiento aplicable para la toma de decisiones estratégicas que permitan optimizar tanto los recursos humanos (medido como costo hora persona), como los costos de infraestructura en la nube y el tiempo total invertido. Con esta investigación, se espera agregar valor al debate académico y empresarial sobre qué herramientas son más adecuadas para enfrentar los retos contemporáneos en la predicción de demanda, aportando una visión más completa que ayude a las empresas (particularmente a nuestro caso de estudio) a mejorar su eficiencia operativa y adaptarse a los cambios del mercado.

Este problema se intensifica cuando se tiene en cuenta la presión por optimizar los recursos, reducir costos y aumentar la eficiencia operativa, lo que implica que un modelo no solo debe ser preciso, sino también fácil de interpretar, flexible y capaz de generar ahorros económicos tangibles para las empresas. La falta de un enfoque comparativo integral que analice la efectividad de estos modelos en un contexto real, considerando tanto la precisión como la interpretabilidad y los costos asociados, crea una barrera para la implementación

óptima de estas herramientas en el sector para las PyMEs[1]. Siendo así, ya que el problema que se intenta resolver debe tener en cuenta la parte técnica como la visión de negocio.

Este estudio se desarrolla a partir de un caso real de una empresa mediana del sector retail de alimentos ubicada en Uruguay. Este contexto regional añade particular relevancia, ya que economías latinoamericanas como la uruguaya enfrentan desafíos estructurales asociados a la volatilidad macroeconómica, la incertidumbre en el consumo y restricciones operativas típicas de las PyMEs. En este sentido, la implementación de herramientas de predicción de demanda no solo mejora la eficiencia interna, sino que también representa una ventaja estratégica para aumentar la resiliencia empresarial frente a escenarios inestables. El análisis busca ser aplicable a otros casos similares dentro de la región.

Este análisis no solo resulta relevante para la empresa estudiada, sino que puede extenderse a otras PyMEs latinoamericanas que enfrentan desafíos similares en la gestión de la demanda.

1.2 Objetivos de la Investigación

La tesis tiene como objetivo responder a la siguiente pregunta de investigación: ¿Qué modelo de predicción de demanda —ARIMA, XGBoost, Prophet o TimeGPT— proporciona la mejor combinación de precisión, interpretabilidad y flexibilidad, generando además el mayor ahorro económico en una empresa mediana del sector retail y de venta al por mayor[2] [3]?

El criterio de éxito de este estudio se basará en la evaluación de cada uno de los modelos en función de su desempeño en tres áreas clave:

1. **Precisión predictiva**, medida a través de métricas[4] como MAE y RMSE.
2. **Interpretabilidad**, es decir, la capacidad del modelo para ser comprendido y explicado por usuarios no técnicos, facilitando la toma de decisiones informadas.
3. **Flexibilidad y ahorro económico**, evaluados en términos de costos operativos (como el uso de infraestructura en la nube y horas de trabajo requeridas), y la capacidad del modelo para adaptarse a cambios en el entorno de predicción.

El modelo que mejor se desempeñe en estas tres dimensiones será considerado el más adecuado para abordar el problema de predicción de demanda para la empresa puntual.

1.3 Alcance y Limitaciones

El alcance de esta tesis se centra en evaluar y comparar el desempeño de distintos modelos de forecasting (ARIMA, XGBoost, Prophet y TimeGPT) en el sector retail y venta al por

[1] Pequeñas y Medianas Empresas.

[2] Estos trabajos representan investigación fundamental y aplicaciones de modelos de pronóstico relevantes para el contexto de las pequeñas y medianas empresas del sector minorista: Cohen et al. (2022), Hastie (2009), Hochreiter (1997), Hyndman (2018), Makridakis y Hibon (2000) y Taylor y Letham (2018).

[3] A lo largo de toda la tesis, se trabajará con una empresa mediana de Uruguay, la cual reservamos su nombre por confidencialidad.

[4] MAE (Mean Absolute Error - Error Absoluto Medio); RMSE (Root Mean Squared Error - Raíz del Error Cuadrático Medio).

mayor, considerando precisión, interpretabilidad, flexibilidad y ahorro económico. Para ello, se emplean datos históricos de ventas, inventarios y clientes, incorporando variables exógenas relevantes.

Sin embargo, el estudio está limitado al contexto de la industria alimentaria, lo que podría reducir la generalización de los hallazgos a otros sectores. Además, el análisis se basa en los datos disponibles y los supuestos realizados durante la limpieza y preprocesamiento, lo que podría influir en la representatividad de los resultados.

1.4 Estructura de la Tesis

La tesis está estructurada en los siguientes capítulos:

- *Introducción + Definición del problema y contexto de negocio:* Contextualiza el problema de investigación, establece los objetivos y describe la relevancia del estudio.
- *Marco Teórico:* Proporciona una visión integral de los métodos de forecasting y sus aplicaciones en el sector retail.
- *Metodología:* Detalla los datos utilizados, las técnicas de preprocesamiento y el diseño experimental para evaluar los modelos.
- *Implementación de Modelos:* Describe la configuración y ajuste de los modelos ARIMA, XGBoost, Prophet y TimeGPT.
- *Resultados y Análisis:* Presenta la comparación de los modelos con métricas clave y su análisis contextual en la industria.
- *Conclusiones:* Resume los hallazgos principales y propone recomendaciones prácticas y líneas futuras de investigación.

Estas secciones están diseñadas para ofrecer un análisis exhaustivo y fundamentado que apoye la toma de decisiones estratégicas en forecasting de demanda.

2 Definición del problema y contexto de negocio

La presente investigación se desarrolla en el contexto de una empresa mediana de Uruguay perteneciente al sector retail de alimentos. Por motivos de confidencialidad, se reserva el nombre de la organización. La empresa cuenta con aproximadamente 50 empleados y un volumen anual de ventas superior a los 2 millones de dólares, operando tanto en ventas minoristas como mayoristas. Su catálogo de productos abarca principalmente alimentos secos, productos de almacén, bebidas, lácteos, conservas y artículos de consumo masivo.

El desafío central que enfrenta la empresa es la alta volatilidad en la demanda de sus productos, motivada por factores como estacionalidades, promociones comerciales, eventos especiales (por ejemplo, festividades locales) y cambios en los patrones de consumo. Actualmente, los procesos de reposición de inventario se basan en métodos heurísticos y proyecciones manuales, lo que ha generado problemas frecuentes de roturas de stock, exceso de inventario, obsolescencia de productos y costos operativos elevados.

El objetivo de este estudio es modelar y anticipar la demanda diaria de los productos principales de la empresa, con especial foco en aquellos que representan un porcentaje significativo de las ventas totales. Se busca construir modelos de predicción que permitan realizar una planificación más precisa de compras y stock, reduciendo así tanto los costos de almacenamiento como las pérdidas por desabastecimiento.

El análisis se realizará utilizando un conjunto de datos históricos que abarca desde abril de 2022 hasta mayo de 2024, en frecuencia diaria. Este período incluye tanto comportamientos de demanda regulares como fluctuaciones asociadas a eventos extraordinarios, permitiendo capturar estacionalidades, tendencias y patrones cíclicos relevantes para el negocio.

Predecir la demanda con mayor precisión resulta estratégico para la empresa, ya que permite optimizar inventarios, mejorar la eficiencia de la cadena de suministro, reducir costos de operación y aumentar la capacidad de respuesta frente a cambios en el mercado. Además, una mejor planificación puede traducirse en una mejora sustancial en la satisfacción del cliente y la rentabilidad general del negocio.

En este contexto, la tesis propone comparar el desempeño de distintos modelos de forecasting (ARIMA, XGBoost, Prophet y TimeGPT) no sólo en términos de precisión, sino también considerando su interpretabilidad, flexibilidad operativa y el ahorro económico que puedan generar en un entorno real de retail de alimentos.

3 Marco teórico

3.1 Predicción de Demanda en el Sector Retail

La predicción de demanda en el sector retail es fundamental para la gestión eficiente de inventarios y la optimización de la cadena de suministro. Este sector, especialmente en la industria alimentaria, se caracteriza por su alta volatilidad y la variabilidad en el comportamiento del consumidor, lo que plantea desafíos significativos para las empresas que buscan equilibrar oferta y demanda. La predicción de demanda en el sector retail ha sido ampliamente estudiada en la literatura, con enfoques que van desde modelos tradicionales hasta técnicas avanzadas de aprendizaje automático y modelos generativos. A continuación, se presentan algunos estudios relevantes que han abordado esta problemática.

Estudios recientes han demostrado que la combinación de datos de ventas con variables macroeconómicas mejora significativamente la precisión del pronóstico. Haque et al. (2023) exploraron el uso de variables como el Índice de Precios al Consumidor (ICP) y la tasa de desempleo en modelos multivariados, obteniendo mejoras en la estabilidad de las predicciones.

En un análisis comparativo de modelos de forecasting aplicados a retail, Hasan et al. (2022) evaluaron la efectividad de enfoques tradicionales como ARIMA frente a modelos más avanzados como Prophet y LightGBM en datos de ventas de Walmart. Los resultados indicaron que los modelos basados en aprendizaje automático proporcionan una mayor precisión en escenarios con alta variabilidad y estacionalidad.

Otro aspecto clave en la predicción de demanda es la relación entre precio y volumen de ventas. Liu y Sustik (2021) abordaron esta problemática mediante un modelo de elasticidad de demanda, optimizando precios minoristas para maximizar ingresos bajo incertidumbre.

Además, la edad del producto ha sido identificada como una variable crucial en la planificación de inventarios en la industria de la moda. Vashishtha et al. (2020) propusieron un modelo basado en la evolución del ciclo de vida del producto, demostrando su impacto en la gestión de inventarios en una multinacional del sector.

Estos estudios destacan la importancia de utilizar enfoques híbridos que combinen información histórica con factores exógenos para mejorar la capacidad predictiva en el retail. La presente investigación toma como referencia estos hallazgos para evaluar el desempeño de distintos modelos en el sector de alimentos, considerando precisión, interpretabilidad y flexibilidad como criterios clave.

3.1.1 Marco de la problemática en retail

El sector retail enfrenta fluctuaciones constantes en la demanda debido a cambios en el comportamiento del consumidor, estacionalidades y eventos externos como festividades o promociones. Según un estudio reciente (Cohen et al., 2022), la predicción de demanda en este sector tiene un impacto directo en la rentabilidad, pues permite ajustar inventarios y minimizar costos operativos, como los asociados al exceso de stock o las pérdidas por faltantes.

3.1.2 Variables de interés

Las variables más comúnmente utilizadas incluyen (Hyndman, 2018) el historial de ventas con datos pasados que reflejan patrones de consumo. La estacionalidad entendiendo a esta como la influencia de períodos específicos del año, por ejemplo festividades o cambios climáticos. Promociones como podrían ser impacto de descuentos o campañas de marketing en la demanda. Y finalmente, regresores exógenos como el clima, la competencia y la macroeconomía.

3.1.3 Técnicas típicamente usadas

Las técnicas más relevantes en forecasting de demanda incluyen (Garza et al., 2023; Hastie, 2009; Taylor & Letham, 2018):

- **Modelos tradicionales:** ARIMA y variantes como SARIMA, efectivos para capturar patrones estacionales y tendencias lineales.
- **Machine Learning:** XGBoost, Random Forest y redes neuronales que permiten integrar múltiples variables y capturar relaciones no lineales.
- **Modelos híbridos:** Combinan enfoques estadísticos y de machine learning para mejorar precisión en escenarios específicos.
- **Modelos probabilísticos:** Herramientas como *Conformal Prediction*, que permiten cuantificar la incertidumbre en las predicciones generando intervalos de confianza válidos. Aunque esta técnica no fue implementada en el presente estudio, se

menciona por su creciente relevancia en escenarios donde la interpretabilidad y la confiabilidad de los pronósticos son clave.

- **Herramientas emergentes:** TimeGPT y otras técnicas generativas que explotan grandes volúmenes de datos para identificar patrones complejos.

3.1.4 Métricas de éxito

Las métricas más relevantes en forecasting de demanda incluyen (Makridakis & Hibon, 2000) el *MAE (Mean Absolute Error)*, el cual evalúa errores absolutos promedio. El *RMSE (Root Mean Squared Error)* que penaliza grandes errores. El *MAPE (Mean Absolute Percentage Error)* útil para comparar errores relativos. Y finalmente el *bias* que evalúa la desviación sistemática de las predicciones.

3.2 Importancia de la Predicción de Demanda

Anticipar la demanda con precisión permite a las empresas **optimizar inventarios** manteniendo niveles adecuados de stock para minimizar costos asociados al exceso o falta de productos. **Mejorar la eficiencia operativa** mediante una planificación adecuada que reduce los tiempos de respuesta y mejora el servicio al cliente. Y finalmente **reducir costos** para minimizar desperdicios y costos operativos es esencial para mantener márgenes de ganancia en un entorno competitivo.

3.3 Aplicaciones en economías latinoamericanas y comparación internacional

Diversos estudios han evaluado la aplicación de modelos de predicción de demanda en economías latinoamericanas. López-Machado (2024) explora modelos econométricos aplicados al crecimiento económico en la región, remarcando la necesidad de adaptar estos enfoques a contextos con alta volatilidad e infraestructura limitada. Por otro lado, García y Pérez (2022) realiza un análisis comparativo entre modelos tradicionales y modernos, destacando la efectividad de herramientas como ARIMA y redes neuronales en entornos con restricciones de datos. Estos antecedentes contrastan con la mayoría de los estudios realizados en países desarrollados, donde se dispone de datos más ricos y entornos operativos más estables. En este sentido, el presente trabajo aporta evidencia empírica desde Uruguay, ofreciendo un análisis comparativo de modelos clásicos y de última generación (como TimeGPT), y evaluando su aplicabilidad en el contexto del retail latinoamericano.

4 Métodos de Pronóstico

4.1 Modelos Tradicionales

ARIMA (Autoregressive Integrated Moving Average)(Hyndman, 2018):

Su función objetivo implica descomponer una serie temporal en tres componentes principales: autorregresión (p), integración para hacer la serie estacionaria (d), y promedios móviles (q). Este modelo predice el valor futuro basado exclusivamente en valores pasados y errores de predicción anteriores.

Las fortalezas de este tipo de modelos se evidencian en que son eficientes para series temporales con patrones estacionales simples y tendencias lineales. Además ofrecen interpretabilidad al modelar explícitamente las relaciones temporales y su baja demanda computacional lo hace ideal para escenarios con recursos limitados.

Por el lado de las debilidades, suelen presentar cierta incapacidad de manejar relaciones complejas o no lineales entre múltiples variables. También son sensibles a la falta de estacionariedad en los datos, lo que requiere preprocesamiento (e.g., diferenciaciones).

Ejemplo Práctico: Utilizado en el pronóstico de ventas diarias de un supermercado pequeño con patrones estacionales claros, como incrementos en días específicos del mes.

4.2 Modelos de Machine Learning

XGBoost (Extreme Gradient Boosting)(Chen & Guestrin, 2016):

Su función objetivo implica construir predicciones basadas en un conjunto de árboles de decisión que se optimizan iterativamente. Cada árbol reduce los errores del anterior mediante un enfoque de *boosting*.

Las fortalezas de este tipo de modelos se evidencian en que son flexibles para incorporar múltiples variables explicativas (e.g., clima, promociones, días festivos). Además son altamente eficaces en capturar relaciones no lineales y detectar interacciones complejas entre variables. Y también presentan herramientas avanzadas como importancia de características (*feature importance*) y valores SHAP ayudan a explicar los *drivers* de las predicciones.

Por el lado de las debilidades, presentan un costo computacional significativamente mayor que los modelos tradicionales y también requieren un cuidadoso ajuste de hiperparámetros para evitar *overfitting* o *underfitting*.

Ejemplo Práctico: Predicción de demanda semanal en una cadena de tiendas de alimentos, considerando promociones, días festivos y tendencias climáticas.

4.3 Modelos Generativos para Series Temporales

TimeGPT (Garza et al., 2023; Simmhan & Ranganathan, 2023):

Su función objetivo implica aplicar redes generativas entrenadas en grandes volúmenes de datos para identificar patrones no lineales y estacionales en series temporales. Funciona como una caja negra altamente automatizada que optimiza predicciones sin necesidad de ajustes manuales extensivos.

Las fortalezas de este tipo de modelos se evidencian en su escalabilidad ya que predice eficientemente millones de series simultáneamente. Además de su gran capacidad para capturar dinámicas de mercado complejas en escenarios con alta volatilidad. Asimismo, no requiere conocimientos avanzados de modelado estadístico gracias a su API intuitiva.

Por el lado de las debilidades, presentan dependencia de infraestructura externa para el procesamiento (e.g., servidores de TimeGPT), menor control e interpretabilidad en comparación con modelos explicativos como Prophet o XGBoost y costos elevados asociados a su uso en implementaciones a gran escala.

Ejemplo Práctico: Uso en la predicción de demanda global para una empresa multinacional de alimentos, combinando datos climáticos, económicos y de ventas regionales.

4.4 Desafíos y Oportunidades

A pesar de los avances en las técnicas de pronóstico, las empresas enfrentan varios desafíos tales como la variabilidad del mercado, en donde llos cambios inesperados en la demanda pueden afectar drásticamente los resultados. Por otro lado, la interpretabilidad debido a que la complejidad de algunos modelos puede dificultar su adopción por parte de usuarios no técnicos. Y finalmente, los costos asociados en donde la implementación de modelos avanzados debe ser evaluada no solo por su precisión, sino también por los costos operativos que implican.

La investigación actual busca no solo evaluar la precisión predictiva de estos modelos, sino también su aplicabilidad práctica en términos económicos y operativos. Este enfoque integral es fundamental para ayudar a las empresas a adaptarse a un entorno dinámico y competitivo. Según Makridakis et al. (2018), es esencial comparar diferentes modelos como ARIMA, XGBoost, Prophet y TimeGPT para determinar cuál proporciona la mejor combinación entre precisión, interpretabilidad y flexibilidad, generando ahorros económicos significativos para el sector retail.

4.5 Patrones en Series Temporales:

4.5.1 Estacionalidades

Las **estacionalidades** en series temporales se refieren a patrones recurrentes que se producen en intervalos regulares a lo largo del tiempo, como días, meses o años. Estos patrones son influenciados por factores estacionales, como cambios climáticos, festividades o ciclos económicos. La identificación de estacionalidades es crucial para mejorar la precisión de los pronósticos, ya que permite ajustar los modelos predictivos para capturar estas variaciones periódicas. Existen diversas técnicas para detectar estacionalidades, entre las que destacan el uso de descomposición de series temporales y modelos como SARIMA (Seasonal Autoregressive Integrated Moving Average) que incorporan componentes estacionales explícitamente. La capacidad de un modelo para ajustarse a estos patrones puede significar una mejora significativa en la precisión de las predicciones.

4.5.2 Tendencias

Las **tendencias** representan la dirección general en la que se mueve una serie temporal a lo largo del tiempo, ya sea hacia arriba (crecimiento), hacia abajo (decrecimiento) o permaneciendo estable. Estas tendencias pueden ser influenciadas por factores económicos, cambios en el comportamiento del consumidor, innovaciones tecnológicas, entre otros. Para identificar tendencias, se pueden utilizar métodos como la suavización exponencial y el ajuste de modelos de regresión. Es fundamental distinguir entre tendencias a corto y largo

plazo, ya que esto puede afectar la estrategia de pronóstico. Un enfoque adecuado para modelar tendencias permite a las empresas anticipar cambios en la demanda y ajustar sus operaciones en consecuencia, optimizando así su cadena de suministro y recursos.

4.5.3 Ciclos y Aleatoriedad

Los **ciclos** en series temporales son oscilaciones que ocurren a intervalos irregulares y están generalmente asociadas con factores económicos o sociales más amplios, como ciclos económicos o cambios demográficos. A diferencia de las estacionalidades, los ciclos no tienen una periodicidad fija y pueden durar varios años. Por otro lado, la **aleatoriedad** se refiere a las fluctuaciones en los datos que no pueden ser explicadas ni predichas a partir de su propio pasado, ya que no guardan relación con tendencias o estacionalidades. Para abordar estos elementos, es común aplicar técnicas estadísticas como el análisis de Fourier o modelos ARIMA que permiten separar componentes cíclicos y aleatorios de la serie temporal. Comprender la naturaleza cíclica y aleatoria es esencial para mejorar la robustez de los modelos predictivos y minimizar el riesgo asociado con decisiones basadas en pronósticos inexactos.

4.6 Consideraciones sobre Overfitting y Costo Computacional

El **overfitting** es un fenómeno crítico en el modelado de series temporales que ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento, capturando tanto las tendencias generales como el ruido aleatorio presente en los datos. Esto puede resultar en un rendimiento deficiente al aplicar el modelo a nuevos conjuntos de datos, ya que la capacidad de generalización se ve comprometida. Para mitigar el overfitting, se pueden implementar diversas estrategias, como la validación cruzada, la regularización y la selección adecuada de hiperparámetros. Estas técnicas permiten evaluar la robustez del modelo frente a datos no vistos y asegurar que las predicciones sean más confiables.

El **costo computacional** es otra consideración esencial al seleccionar modelos para la predicción en series temporales. Diferentes algoritmos requieren distintos niveles de recursos computacionales, lo que puede impactar significativamente en la eficiencia operativa de una organización. Modelos más complejos, como los basados en machine learning, pueden ofrecer una mayor precisión pero a menudo conllevan un mayor costo en términos de tiempo de procesamiento y recursos computacionales. Por otro lado, modelos más simples como ARIMA pueden ser menos costosos computacionalmente, pero podrían no capturar adecuadamente patrones complejos en los datos.

Al abordar el problema del overfitting y el costo computacional, es fundamental encontrar un equilibrio entre la precisión del modelo y su viabilidad operativa. Esto implica no solo evaluar el rendimiento predictivo sino también considerar los recursos disponibles y la necesidad de interpretabilidad del modelo por parte de los usuarios finales. La elección del modelo debe alinearse con los objetivos estratégicos de la empresa, buscando maximizar la eficiencia y minimizar costos sin sacrificar la calidad de las predicciones.

Estas consideraciones son cruciales para garantizar que los modelos seleccionados no solo sean precisos en sus predicciones, sino también sostenibles y aplicables en entornos empresariales reales, donde los recursos son limitados y las decisiones deben basarse en análisis confiables.

5 Metodología

Para abordar el problema planteado en la tesis, se propone un enfoque metodológico basado en una combinación de técnicas cuantitativas y análisis comparativo, utilizando datos reales del sector retail y venta al por mayor. La metodología prevista es la siguiente:

1. **Recolección y Preprocesamiento de Datos:**

   Se trabajará con un conjunto de datos históricos de demanda, inventarios y ventas provenientes de una empresa mediana del sector retail. Estos datos serán preprocesados para eliminar inconsistencias, gestionar valores faltantes y escalar las variables cuando sea necesario. El análisis incluirá la incorporación de variables exógenas relevantes, como estacionalidad, eventos especiales y cambios en el comportamiento del consumidor, para enriquecer el modelo de predicción.

2. **Modelado Predictivo:**

   Se trabajará con diferentes modelos de predicción de demanda, incluyendo ARIMA, XGBoost, Prophet y TimeGPT. Los modelos serán ajustados con técnicas como validación cruzada y selección de hiperparámetros según se requiera, con el objetivo de optimizar su rendimiento en términos de precisión y capacidad de adaptación a cambios en el entorno de predicción.

3. **Comparación de Modelos:**

   Cada modelo será evaluado utilizando métricas de desempeño como MAE y RMSE. Es decir, dos funciones de pérdida diferentes: MAE mide el error absoluto promedio, mientras que RMSE enfatiza los errores más grandes al elevarlos al cuadrado. Ambas métricas tienen interpretación en términos de las unidades de medida de la demanda. A continuación se presentan tanto las fórmulas correspondientes como su interpretación e importancia:

   (a) **MAE (Mean Absolute Error)**

   - **Definición:** El **MAE** mide el promedio de los errores absolutos entre las predicciones ($\hat{y}_i$) y los valores reales (y_i). A diferencia del RMSE, no eleva los errores al cuadrado, lo que lo hace menos sensible a valores extremos.
   - **Fórmula:**
   $$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$
   - **Importancia:** El **MAE** proporciona una medida directa de la magnitud del error en las mismas unidades de la variable objetivo. No penaliza de manera desproporcionada los errores grandes, por lo que es útil cuando se busca una métrica más interpretable y menos influenciada por valores atípicos.
   - **Interpretación:**
     - Se expresa en las mismas unidades que la variable objetivo, lo que facilita su interpretación y comparación con los valores reales.
     - Es una métrica robusta y fácil de entender, ya que representa el error promedio sin dar más peso a los valores extremos.

- Por ejemplo, en un problema de predicción de demanda, un **MAE de 50** significa que, en promedio, las predicciones se desvían **50 unidades** de la demanda real, sin importar si estos errores son grandes o pequeños.

Adicionalmente, el uso de métricas como el MAE permitirá evaluar la calidad de las predicciones de manera intuitiva, complementando otros análisis como la importancia de características para entender qué variables impactan más en los resultados.

(b) **RMSE (Root Mean Squared Error):**

- **Definición:** El RMSE mide la raíz cuadrada del promedio de los errores al cuadrado entre las predicciones ($\hat{y}_i$) y los valores reales (y_i). Penaliza más los errores grandes debido a la elevación al cuadrado.
- **Fórmula:**

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

- **Importancia:** Penaliza errores más grandes debido a la elevación al cuadrado, siendo útil para detectar casos extremos.
- **Interpretación:**
  - Está expresado en las mismas unidades que la variable objetivo, lo que facilita la comparación directa con los valores reales.
  - Es útil para identificar modelos que tienden a realizar predicciones con grandes desviaciones, ya que los errores grandes tienen un impacto mayor en esta métrica.
  - Por ejemplo, en un problema de predicción de demanda, un RMSE de 50 significa que, en promedio, las predicciones están desviadas por 50 unidades de la demanda real, penalizando más aquellos casos con grandes errores.

Adicionalmente al uso de las métricas anteriormente mencionadas, se realizarán análisis de importancia de las características para comprender mejor qué variables influyen en las predicciones y facilitar la interpretabilidad.

4. **Análisis de Flexibilidad y Ahorro Económico:**

   Se evaluará la flexibilidad de los modelos en términos de su capacidad para adaptarse a nuevos escenarios, como cambios abruptos en la demanda o en el contexto del negocio. Además, se analizará el impacto económico de cada modelo, considerando factores como el costo computacional (horas de trabajo, uso de infraestructura en la nube) y el costo de oportunidad (optimización de inventarios, reducción de desperdicios y sobrecostos).

5. **Simulación de Escenarios:**

   Se realizarán simulaciones que permitan evaluar el desempeño de los modelos en diferentes escenarios de demanda, tanto en condiciones normales como en situaciones extremas (picos o caídas abruptas de la demanda). Este análisis contribuirá a

determinar qué modelo proporciona mejores resultados en términos de robustez y adaptación a cambios en el entorno.

6. **Análisis Cualitativo de Resultados:**

   Además de los análisis cuantitativos, se realizarán entrevistas o encuestas con actores clave del sector para recoger percepciones sobre la interpretabilidad y aplicabilidad de los modelos, con el fin de integrar una dimensión cualitativa en la evaluación. Al aplicar esta combinación de enfoques, la tesis buscará proporcionar una solución integral al problema de predicción de demanda, comparando de manera exhaustiva las ventajas y limitaciones de cada modelo, y brindando insights valiosos para su implementación en el sector retail y venta al por mayor.

5.1 Descripción de los Datasets

La ventana de tiempo del dataset inicia el 16/04/2022 (Abril 2022) hasta el 29/05/2024 (Mayo 2024), en frecuencia diaria (un total de 774 días y 275.100 transacciones). A continuación, se presentan los conjuntos de datos que serán utilizados para el análisis y la predicción de demanda. Los datos provienen de un sistema de gestión de ventas y productos, y están estructurados en tres componentes principales: clientes, productos y ventas.

5.1.1 Separación en Conjuntos de Entrenamiento y Testeo

Para llevar a cabo la separación, se definió una fecha de corte, TEST_DATE, establecida como 15-01-2024. Esta fecha fue seleccionada estratégicamente para que todos los registros anteriores a ella se utilizaran en el conjunto de entrenamiento, mientras que los registros correspondientes a esa fecha y posteriores se destinarían al conjunto de prueba.

Esta metodología asegura que el modelo no tenga acceso a información futura durante su fase de entrenamiento, lo cual es esencial para simular un entorno realista.

Una vez realizada la separación, se procedió a imprimir las estadísticas correspondientes a ambos conjuntos:

Los resultados obtenidos fueron[5]:

- Datos de training: 550 filas - 82.0 %
- Datos de testing: 119 filas - 18.0 %

Esto indica que aproximadamente el 82 % de los datos se utilizarán para entrenar el modelo, mientras que el 18 % restante servirá para evaluar su rendimiento. Esta proporción es comúnmente aceptada en prácticas de machine learning, ya que proporciona un equilibrio adecuado entre la cantidad de datos disponibles para el entrenamiento y la necesidad de contar con un conjunto significativo para la validación.

La correcta separación entre estos conjuntos es crucial no solo para evitar el sobreajuste, sino también para garantizar que las métricas obtenidas durante la evaluación reflejen

[5]La división 82-18 % se definió combinando criterios técnicos y decisiones de negocio, buscando reflejar un escenario realista donde el modelo predice sobre un futuro no observado.

con precisión la capacidad del modelo para generalizar a nuevos datos. Esto resulta especialmente relevante en el contexto del retail, donde la anticipación precisa de la demanda puede traducirse en importantes ahorros operativos y mejoras en la eficiencia.

5.1.2 Datos de Clientes

Este conjunto de datos contiene 1.670 clientes y 8 features con información sobre los clientes de la empresa, como su ID, razón social, nombre de fantasia, país, departamento, ciudad, estado de actividad y tipo de cliente. Este dataset es esencial para entender el perfil de los clientes y su distribución geográfica.

5.1.3 Datos de Productos

Contiene datos detallados con 1.320 productos y 11 features con información sobre los mismos ofrecidos por la empresa, incluyendo código de producto, descripción, costo, precio, stock actual, unidad de medida, proveedor, familia de productos, entre otros. Este conjunto de datos es fundamental para el análisis de la demanda de productos específicos.

5.1.4 Datos de Ventas

Registra 275.100 transacciones de ventas durante más de 2 años, con información de 6 distintos features, incluyendo el código de producto, moneda de venta, ID de cliente, fecha de venta, monto de venta y cantidad vendida. Este dataset es clave para modelar y predecir la demanda de productos en función del histórico de ventas.

La Tabla 1 resume los principales datasets utilizados en el estudio, incluyendo las variables más relevantes de cada uno. Esta información permite dimensionar y entender el enfoque estructurado adoptado para el modelado.

Tabla 1. Descripción de los datasets

Dataset	Columna	Descripción
Ventas	codigo_producto	Código identificador del producto vendido.
	moneda_venta	Moneda en la cual se realizó la venta (por ejemplo, pesos, dólares, etc.).
	id_cliente	Identificador único del cliente que realizó la compra.
	fecha_venta	Fecha en la que se realizó la venta.
	monto_ventas_producto	Monto total de la venta del producto.
	cantidad_ventas_producto	Cantidad de productos vendidos en la transacción.
Clientes	id	Identificador único del cliente.
	razon_soc	Razón social del cliente (nombre de la empresa).
	nombre_fantasia	Nombre de fantasía del cliente (nombre comercial, si lo tiene).
	pais	País donde está ubicado el cliente.
	departamento	Departamento o estado donde está ubicado el cliente.
	ciudad	Ciudad donde está ubicado el cliente.
	activo	Estado del cliente (N para inactivo, S para activo).
	tipo_cliente	Tipo de cliente (puede estar clasificado según ciertos criterios).
Productos	codigo_producto	Código único identificador del producto.
	descripcion_producto	Descripción detallada del producto.
	costo_producto	Costo del producto para la empresa.
	precio_producto	Precio de venta del producto.
	stock_actual_producto	Cantidad disponible en inventario del producto.
	unidad_de_medida_producto	Unidad de medida utilizada para el producto (por ejemplo, unidad, kg, etc.).
	proveedor_producto	Nombre del proveedor del producto.
	ramo_producto	Categoría o ramo al que pertenece el producto.
	familia_producto	Familia o subcategoría del producto.
	grupo_producto	Grupo o clasificación adicional del producto.
	marco_producto	Marca o fabricante del producto.

Fuente: Elaboración propia.

Esta síntesis facilita la comprensión del volumen y calidad de los datos disponibles, así como su estructura para el análisis posterior.

5.2 Preprocesamiento de Datos

5.2.1 Transformación y Limpieza de Datos

El preprocesamiento de los datos[6] fue un paso clave para garantizar la calidad y consistencia del análisis. Las principales tareas realizadas incluyen la **gestión de valores faltantes** en donde se identificaron y trataron los valores ausentes en las columnas críticas del dataset. Dependiendo del caso, estos valores se imputaron utilizando la mediana o media de las variables relevantes, o se descartaron registros irrelevantes para el análisis. Asimismo, se realizó una **normalización y escalado** para facilitar la convergencia de los modelos, las variables numéricas como montos de venta y cantidades fueron escaladas utilizando técnicas de normalización min-max, garantizando que los datos se ajusten al rango de los modelos.

Por otro lado, se realizó una **ingeniería de atributos** en donde se crearon nuevas variables basadas en el análisis exploratorio, como indicadores para días de la semana, fines de semana y eventos festivos. Estas transformaciones permitieron capturar patrones relevantes asociados a cambios en la demanda. Por consiguiente, se procedió a realizar

[6]Para mayor detalle sobre la transformación y limpieza de datos realizada consultar el Apéndice 9.1.

el **filtrado y limpieza de datos atípicos**. Se analizaron los datos en busca de valores atípicos en ventas y cantidades. Aquellos registros que superaban umbrales definidos con base en el análisis estadístico fueron revisados y, de ser necesario, eliminados o ajustados.

Finalmente, se obtuvo la **consolidación de datasets**. Los conjuntos de datos de clientes, productos y ventas se unieron utilizando identificadores comunes, asegurando consistencia en el manejo de las claves foráneas y eliminando duplicados.

Estas transformaciones proporcionaron un conjunto de datos limpio y estructurado, adecuado para los modelos de predicción y análisis subsiguientes.

5.2.2 Análisis Exploratorio

Los datos disponibles son adecuados en términos de calidad y cantidad para llevar a cabo el análisis y los experimentos necesarios. El siguiente paso en nuestro proceso será iniciar el análisis exploratorio de los mismos y la posterior implementación de las herramientas de forecasting previamente mencionadas.

Recordando que nuestro objetivo es armar un forecast de demanda de productos alimenticios, a continuación se detalla un primer análisis exploratorio del dataset correspondiente.

En primer lugar, contamos con un primer dataset parcial con una primera ingeniería de atributos que es la siguiente:

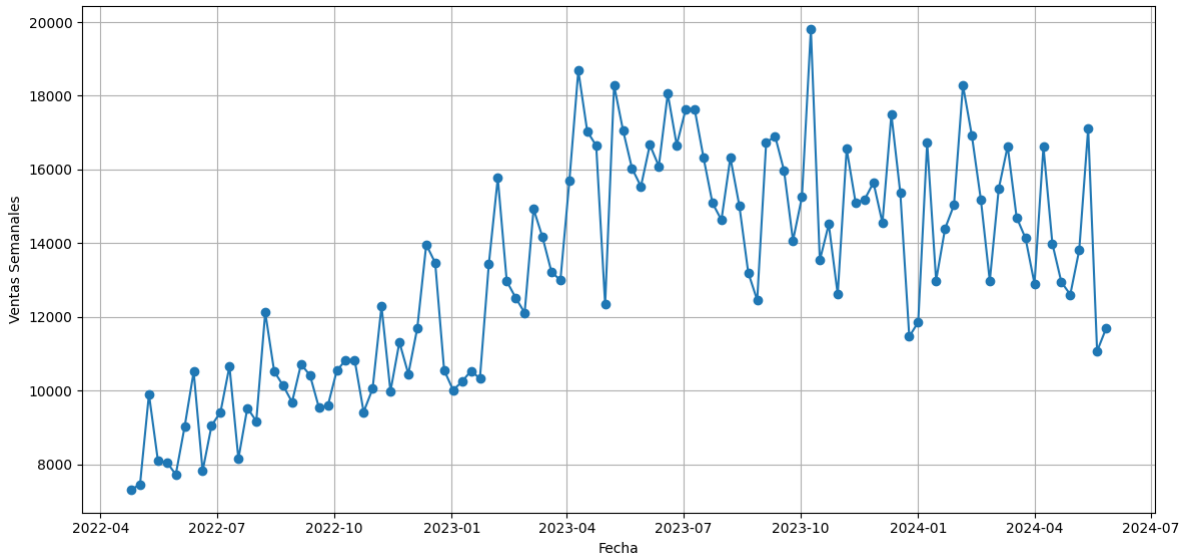
Gráfico 1. Consolidación de datasets

	ds	y	day_of_week	day_of_month	month	year	is_weekend	quarter	week_of_year	day_of_year	is_holiday	rolling_mean_2	rolling_mean_7	rolling_mean_30
0	2022-04-16	39.000	5	16	4	2022	1	2	15	106	0	39.0000	39.000000	39.000000
1	2022-04-17	165.640	6	17	4	2022	1	2	15	107	0	102.3200	102.320000	102.320000
2	2022-04-18	2235.130	0	18	4	2022	0	2	16	108	0	1200.3850	813.256667	813.256667
3	2022-04-19	1167.624	1	19	4	2022	0	2	16	109	0	1701.3770	901.848500	901.848500
4	2022-04-20	1734.857	2	20	4	2022	0	2	16	110	0	1451.2405	1068.450200	1068.450200
5	2022-04-21	1071.296	3	21	4	2022	0	2	16	111	0	1403.0765	1068.924500	1068.924500
6	2022-04-22	1497.522	4	22	4	2022	0	2	16	112	0	1284.4090	1130.152714	1130.152714
7	2022-04-23	857.400	5	23	4	2022	1	2	16	113	0	1177.4610	1247.067000	1096.058625
8	2022-04-24	8.000	6	24	4	2022	1	2	16	114	0	432.7000	1224.547000	975.163222
9	2022-04-25	977.481	0	25	4	2022	0	2	17	115	0	492.7405	1044.882857	975.395000

Fuente: Elaboración propia.

Como primer paso, analizaremos la serie de tiempo como medida agregada de todos los productos de las ventas semanales.

Gráfico 2. Serie de tiempo de ventas (semanal)



Fuente: Elaboración propia.

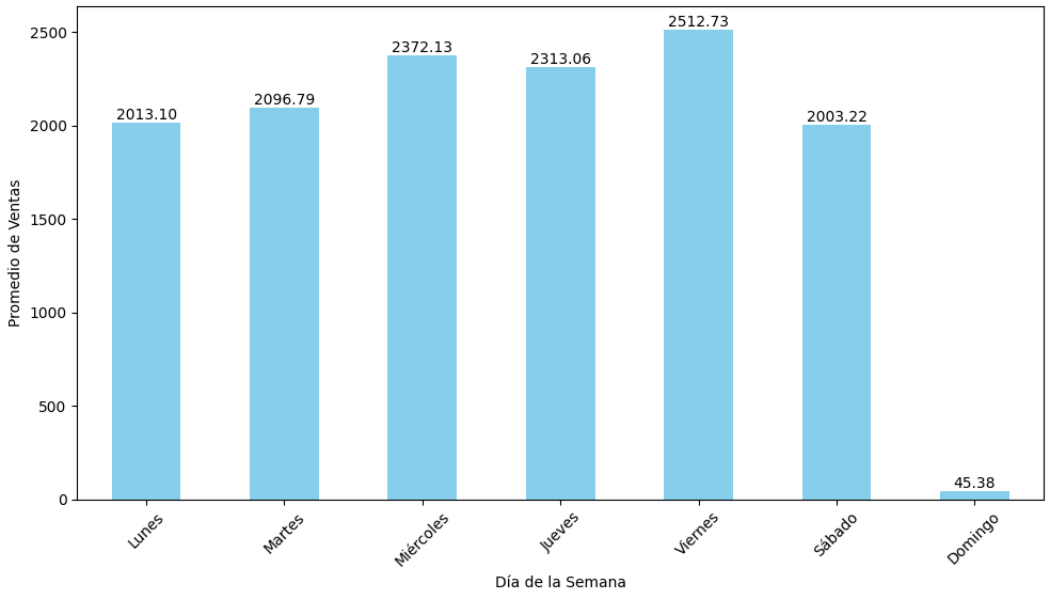
Se observa una tendencia general al alza en las ventas hasta abril de 2023, momento en el que las ventas parecen estabilizarse. Esto puede reflejar un periodo inicial de crecimiento en la empresa, seguido de una etapa de consolidación del mercado. En ese sentido, veremos más adelante que son períodos de campañas de marketing de crecimiento y luego de consolidación. Lo cual, hace sentido de negocio ese tipo de comportamiento.

Este mismo a nivel modelado puede indicar la necesidad de adaptar estrategias de marketing o considerar estacionalidades y eventos específicos que impactaron las ventas en este periodo.

Como segunda medida analizaremos el promedio de ventas según el día de la semana. Los días entre semana muestran consistentemente mayores ventas en comparación con los fines de semana, lo cuál también hace sentido con el negocio ya que los días domingos permanece cerrado al público, lo que vemos de movimiento dicho día corresponde a ajustes que se realizan por sistema.

Este patrón debe considerarse para ajustar estrategias promocionales, aumentando incentivos para compras en fines de semana (sábado particularmente) o planificando inventarios con base en la demanda diaria.

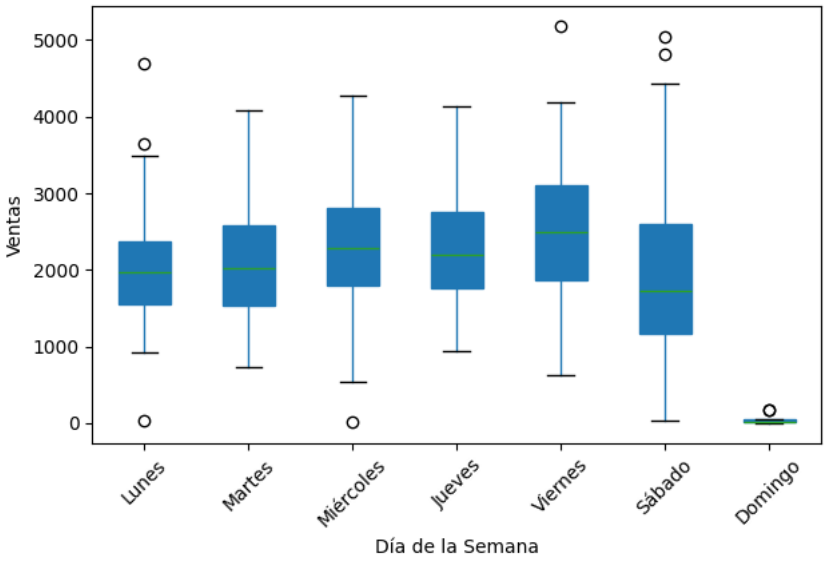
Gráfico 3. Promedio de ventas por día de la semana



Fuente: Elaboración propia.

En tercer lugar, analizaremos además del promedio de ventas según el día de la semana también su distribución mediante el siguiente gráfico de tipo boxplot.

Gráfico 4. Distribución de ventas por día de la semana



Fuente: Elaboración propia.

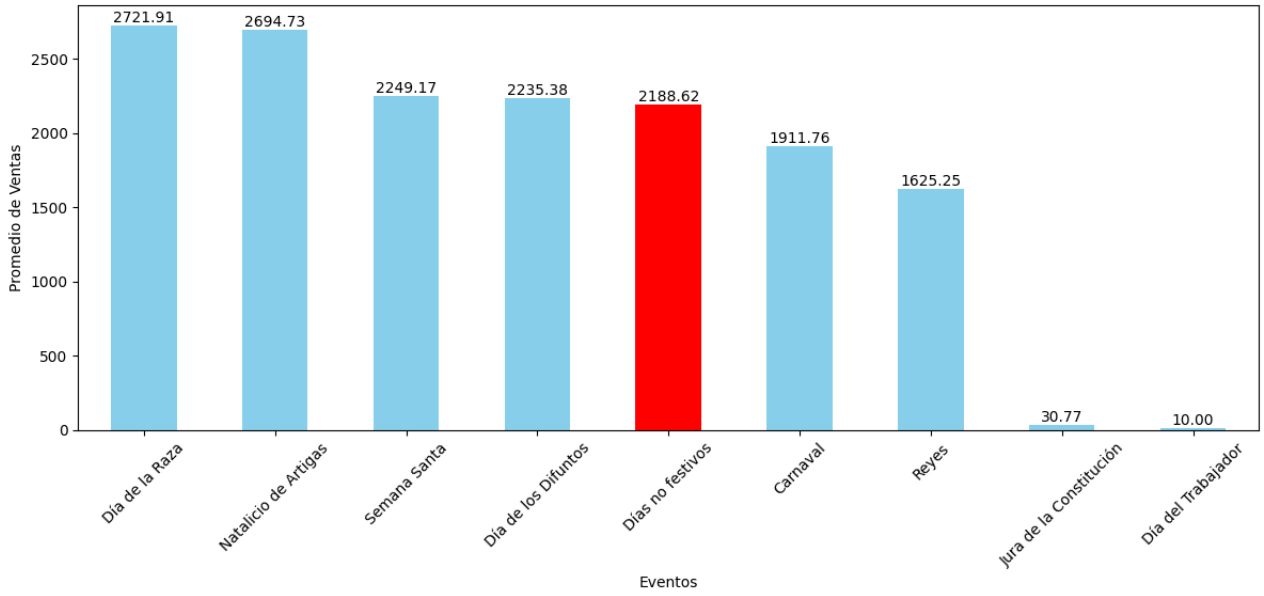
Las distribuciones más amplias en ciertos días reflejan una mayor variabilidad en la demanda. Esto podría estar influenciado por factores externos como promociones o even-

tos específicos en esos días.

Identificar las razones detrás de estas variaciones puede ayudar a implementar estrategias de planificación más efectivas.

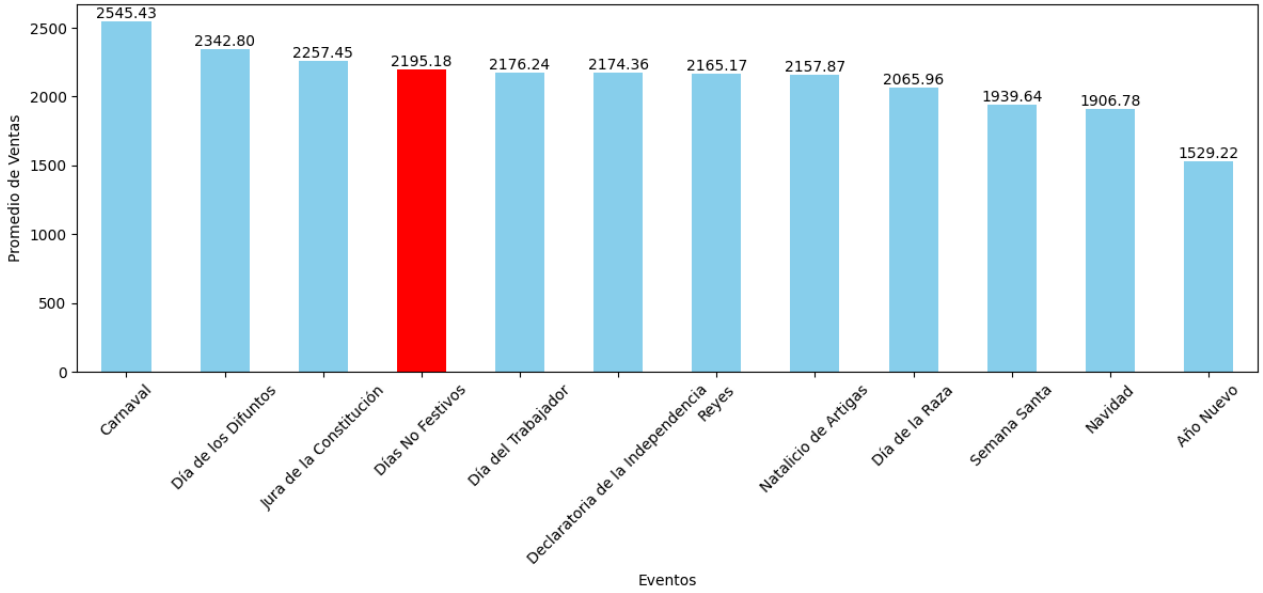
Debido a que trabajaremos con días festivos y eventos especiales, resulta importante poder analizar también cómo se distribuye el promedio de ventas según dichos eventos. Para ello, lo analizaremos de dos formas. En primer medida, tomando la fecha particular y en segundo lugar tomando una ventana de tres días hacia adelante y tres días hacia atrás de dicho evento. De esta forma, brindamos mayor flexibilidad para lograr capturar el comportamiento en las ventas.

Gráfico 5. Promedio de ventas por eventos



Fuente: Elaboración propia.

Gráfico 6. Promedio de ventas por eventos (con ventana de 3 días)



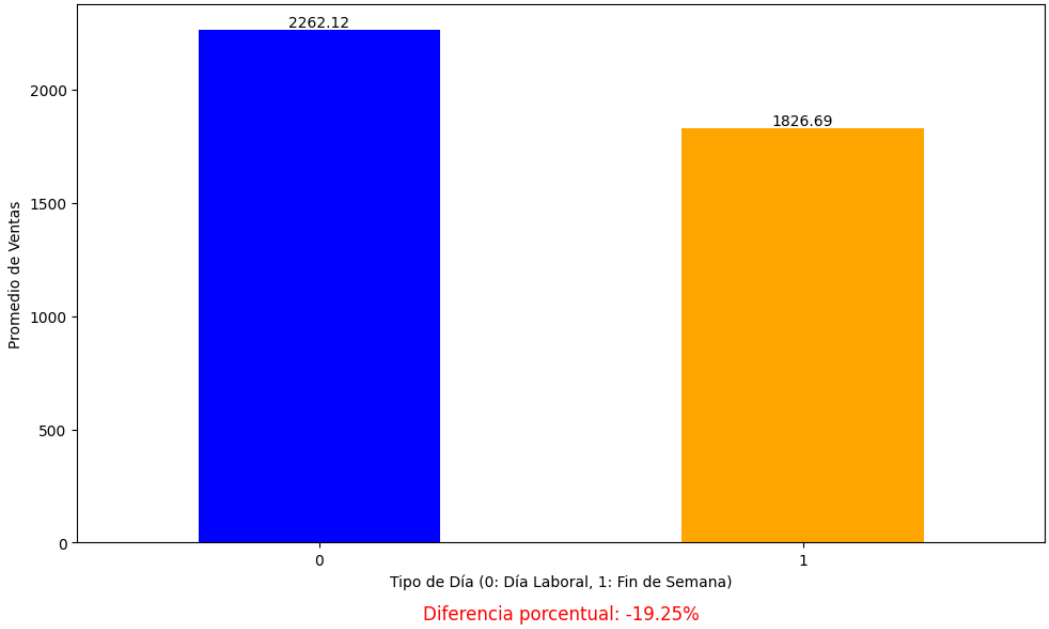
Fuente: Elaboración propia.

Se evidencia un incremento de ventas alrededor de ciertos eventos, especialmente cuando se considera una ventana de 3 días antes y después. Esto muestra que los eventos especiales tienen un efecto anticipado y rezagado en la demanda.

Incorporar estos efectos en los modelos de predicción puede mejorar la precisión del forecast, especialmente en periodos cercanos a eventos importantes.

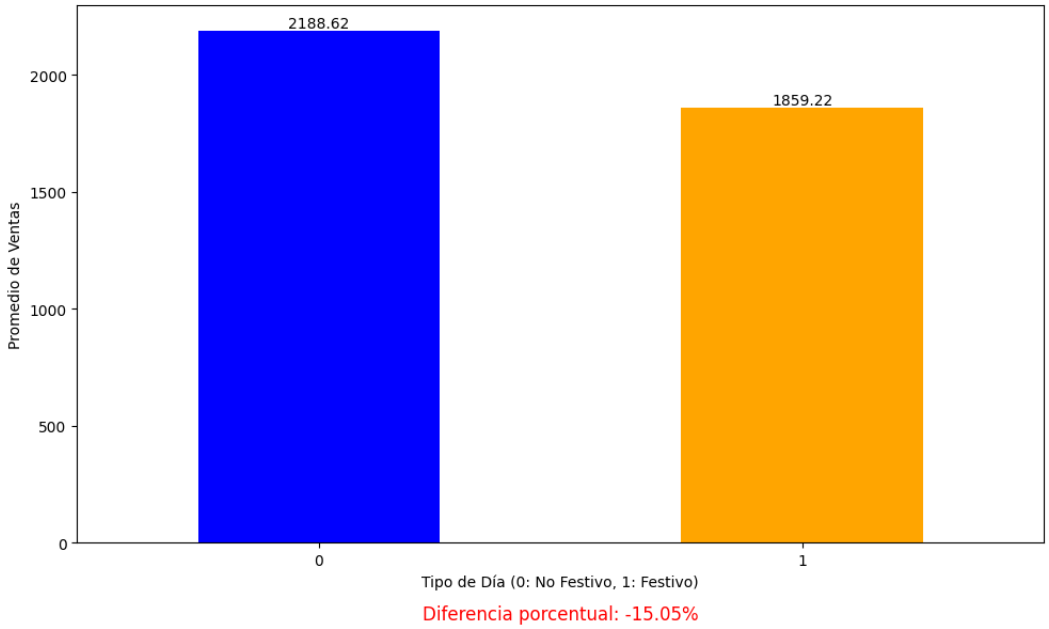
A continuación analizaremos el promedio de ventas según el tipo de día, ya sea día de semana o fin de semana. Y además, si nos encontramos frente a un evento o no.

Gráfico 7. Promedio de ventas por tipo de día (fin de semana vs día laboral)



Fuente: Elaboración propia.

Gráfico 8. Promedio de ventas por tipo de día (festivo vs no festivo)



Fuente: Elaboración propia.

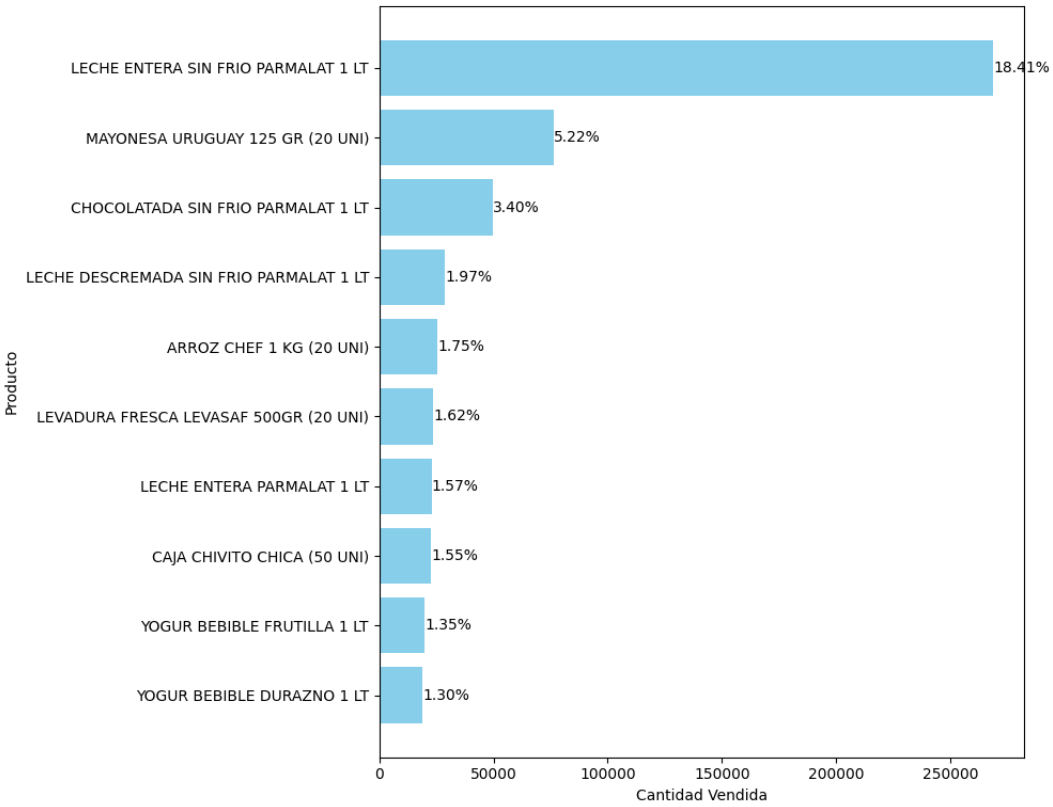
Lo que podemos observar es que, en promedio, los fines de semana la venta baja un 19,25 % respecto a un día de semana. Mientras que, en un evento o día festivo, las ventas

son un 15,05 % menor en comparación a un día no festivo.

Esto sugiere que los hábitos de compra disminuyen en periodos no rutinarios. Este hallazgo puede informar la planificación de inventarios, priorizando abastecimiento en días laborables y no festivos.

Finalmente, analizando el top 10 productos más vendidos de la empresa nos encontramos con el siguiente ranking tanto su porcentual respecto del total de ventas como en unidades vendidas[7].

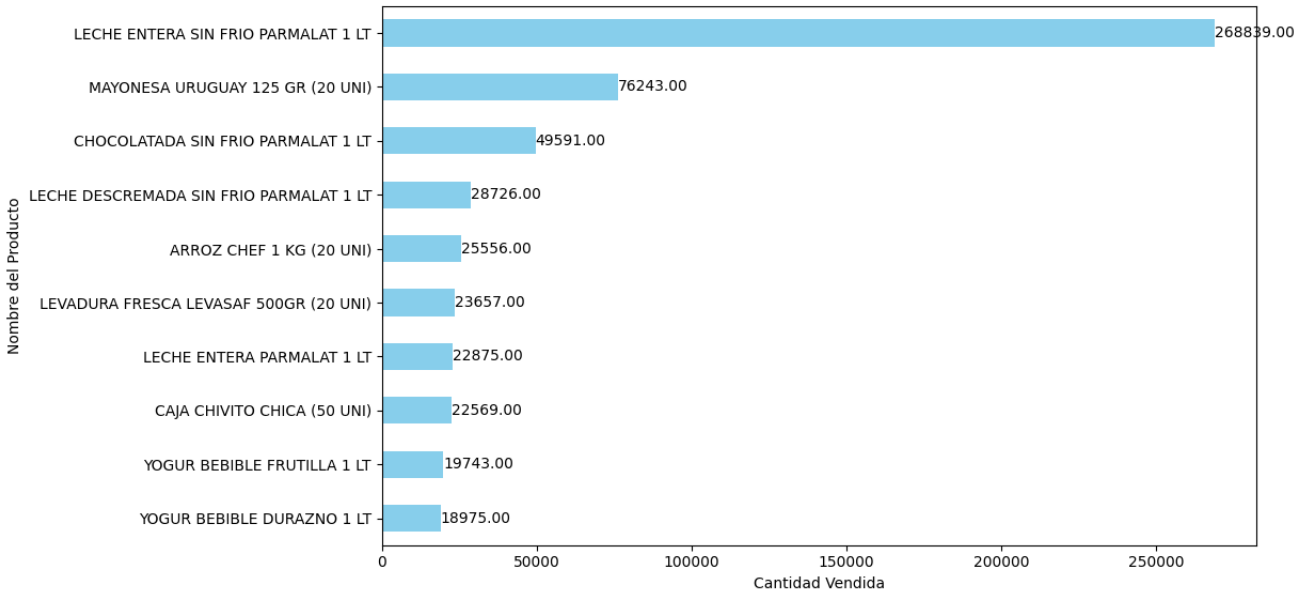
Gráfico 9. Top 10 productos más vendidos, %



Fuente: Elaboración propia.

[7]El orden para alcanzar el top 10 tanto en el gráfico 9 como en el 10 se realizó tomando en cuenta la cantidad de ventas en unidades del producto.

Gráfico 10. Top 10 productos más vendidos, valor absoluto



Fuente: Elaboración propia.

Como se puede observar, el top 10 productos representan el 38.13 % de las ventas totales, destacando una alta concentración de demanda en un subconjunto limitado del portafolio. Posiblemente nuestro forecast deba ser más preciso en dicha canasta. Es decir que la precisión del forecast debe priorizar este subconjunto de productos, dado su impacto significativo en los ingresos totales. Esto es algo que necesitaremos evaluar durante el transcurso de la tesis.

5.3 Modelos de Predicción

La tesis examina cuatro modelos principales: ARIMA, XGBoost, Prophet y TimeGPT. Cada uno de estos modelos presenta ventajas y desventajas que se evaluarán a lo largo del estudio.

- **ARIMA:** Este modelo tradicional es eficaz para datos con patrones lineales y estacionales, aunque su capacidad para manejar relaciones complejas es limitada.
- **XGBoost:** Un enfoque más moderno que permite la inclusión de múltiples variables, ofreciendo mayor precisión en las predicciones, pero con un costo en términos de interpretabilidad.
- **Prophet:** Desarrollado por Facebook, es especialmente útil para datos irregulares y eventos especiales, aunque su efectividad ante cambios abruptos en la demanda sigue siendo un área a explorar.
- **TimeGPT**[8]**:** Un modelo emergente que utiliza técnicas generativas, mostrando un potencial significativo para capturar patrones no lineales en grandes volúmenes de datos.

[8] Algunas de las características que la propia documentación destaca (https://www.nixtla.io/timegpt) son:

El objetivo principal es identificar cuál de estos modelos proporciona la mejor combinación de precisión, interpretabilidad y flexibilidad, además de generar ahorros económicos significativos para el caso de estudio. A través de un análisis comparativo que incluirá métricas clave y simulaciones, se buscará determinar el modelo más adecuado para optimizar la gestión de inventarios y mejorar la eficiencia operativa en el sector retail.

Este enfoque integral permitirá no solo evaluar el rendimiento predictivo de cada modelo, sino también su aplicabilidad práctica en un entorno empresarial real.

5.4 Diseño de los Experimentos

En esta sección se describirá el diseño de los experimentos realizados para evaluar la efectividad de los modelos de predicción de demanda seleccionados. Se establecerán las condiciones bajo las cuales se llevarán a cabo las pruebas, asegurando que los resultados sean válidos y representativos del rendimiento real de cada modelo. La metodología incluye la definición de conjuntos de datos previamente mencionada, la selección de métricas de evaluación y el proceso de ajuste de hiperparámetros.

5.4.1 Métricas de Evaluación

Para medir el rendimiento de los modelos, se utilizarán diversas métricas que permitirán cuantificar su precisión y efectividad. Las métricas seleccionadas incluyen:

- **MAE (Mean Absolute Error):** Mide el error absoluto promedio entre las predicciones y los valores reales, expresado en las mismas unidades de la variable objetivo. Es fácil de interpretar y no penaliza de manera desproporcionada los errores grandes, lo que lo hace útil para evaluar la precisión de un modelo sin sesgo hacia valores atípicos.
- **RMSE (Root Mean Squared Error):** Evalúa la raíz del error cuadrático medio, siendo sensible a grandes errores.

Las métricas MAE y RMSE se computaron utilizando las predicciones generadas por cada modelo sobre el conjunto de test, comparándolas contra las ventas reales observadas en dicho período. Este enfoque out-of-sample permite simular un escenario realista, en el que los modelos deben predecir datos futuros no vistos durante el entrenamiento. Evaluar el desempeño con datos fuera de muestra es clave para garantizar la robustez y generalización del modelo, y evita resultados artificialmente optimistas que pueden surgir al evaluar sobre el mismo set de entrenamiento (in-sample).

Estas métricas permitirán comparar el rendimiento predictivo de cada modelo en función de su capacidad para ajustarse a los datos históricos y realizar pronósticos precisos.

– Fácil: Solo tienes que cargar y hacer predicciones. No se requiere formación.

– Grande: TimeGPT se entrenó con la colección de datos más grande de la historia: más de 100 mil millones de filas de datos financieros, meteorológicos, energéticos y web.

– Revolucionario: Democratiza el poder del análisis de series de tiempo haciéndolo accesible para todos a través de una simple llamada API.

– Rápido: Predecir una o millones de series en segundos.

5.4.2 Consistencia de regresores entre modelos

Para asegurar una comparación justa entre los modelos de pronóstico, se buscó mantener constante el conjunto de variables predictoras utilizadas.

De esta manera, se garantiza que las diferencias de performance entre modelos no se deban a diferencias en la información utilizada, sino a la capacidad predictiva de cada técnica.

5.4.3 Búsqueda de Hiperparámetros

La búsqueda de hiperparámetros es un paso crucial en el diseño experimental, ya que permite optimizar el rendimiento de cada modelo. Se implementarán técnicas como: **búsqueda en cuadrícula (Grid Search)** en donde se explorarán combinaciones exhaustivas de hiperparámetros para encontrar la configuración óptima. Por otro lado, la **búsqueda aleatoria (Random Search)**. Con ella se seleccionarán aleatoriamente combinaciones de hiperparámetros dentro de un rango definido, lo que puede ser más eficiente en términos de tiempo.

Este proceso garantizará que cada modelo opere en su máxima capacidad, adaptándose a las peculiaridades del conjunto de datos.

5.4.4 Validación Cruzada

La validación cruzada es una técnica fundamental para evaluar la robustez y generalización de los modelos. En este estudio, se aplicará la validación cruzada $k - fold$, donde el conjunto de datos se dividirá en k subconjuntos. El proceso incluirá la división del conjunto en donde los datos se dividirán aleatoriamente en k partes iguales. El entrenamiento y prueba, en donde cada modelo se entrenará k veces, utilizando $k - 1$ subconjuntos para el entrenamiento y 1 para la prueba en cada iteración. Y finalmente el promedio de los resultados. Con ello se calcularán las métricas promedio sobre todas las iteraciones para obtener una estimación más precisa del rendimiento del modelo.

Este enfoque asegura que los modelos sean evaluados en diferentes segmentos del conjunto de datos, proporcionando una visión completa de su capacidad predictiva y reduciendo el riesgo de sobreajuste.

El diseño experimental propuesto permitirá una evaluación rigurosa y sistemática de los modelos seleccionados, facilitando la identificación del enfoque más efectivo para la predicción de demanda en el sector retail.

5.4.5 Validación Cruzada en Series de Tiempo

Dado que las series temporales presentan una estructura cronológica que no debe romperse, la validación cruzada tradicional (k-fold aleatoria) no es adecuada, ya que puede introducir fugas de información del futuro al pasado. Por ello, en este estudio se optó por un enfoque de validación cruzada específica para series de tiempo, conocido como **Time Series Cross-Validation** o **Expanding Window**. Este método consiste en entrenar el modelo sobre una ventana inicial de datos y evaluar su rendimiento en una ventana futura, incrementando progresivamente el tamaño del conjunto de entrenamiento. Esta estrategia

preserva el orden temporal de los datos y simula escenarios reales de pronóstico, donde las predicciones deben hacerse hacia adelante en el tiempo. En nuestro caso, se utilizaron múltiples ventanas de entrenamiento y validación, respetando siempre la secuencia cronológica de la demanda.

6 Implementación de Modelos

6.1 ARIMA

6.1.1 Configuración y Ajuste del Modelo

1. **Preparación de los Datos:**
   - La serie temporal fue transformada para garantizar estacionariedad mediante diferenciaciones sucesivas.
   - Se analizaron los correlogramas de autocorrelación (ACF) y autocorrelación parcial (PACF) para identificar los parámetros iniciales p, d y q.
2. **Selección de Parámetros:**
   - Mediante una búsqueda exhaustiva (Grid Search), se seleccionaron los valores óptimos de los parámetros del modelo, buscando minimizar el AIC (Akaike Information Criterion).
3. **Validación y Evaluación:**
   - Se aplicó validación cruzada con ventanas deslizantes para medir el rendimiento del modelo en diferentes segmentos temporales.
   - Las métricas principales utilizadas fueron MAE y RMSE, enfocándose en la capacidad del modelo para predecir la demanda con precisión.

6.1.2 Resultados Obtenidos

Tras la implementación y ajuste del modelo ARIMA, se obtuvieron los siguientes resultados:

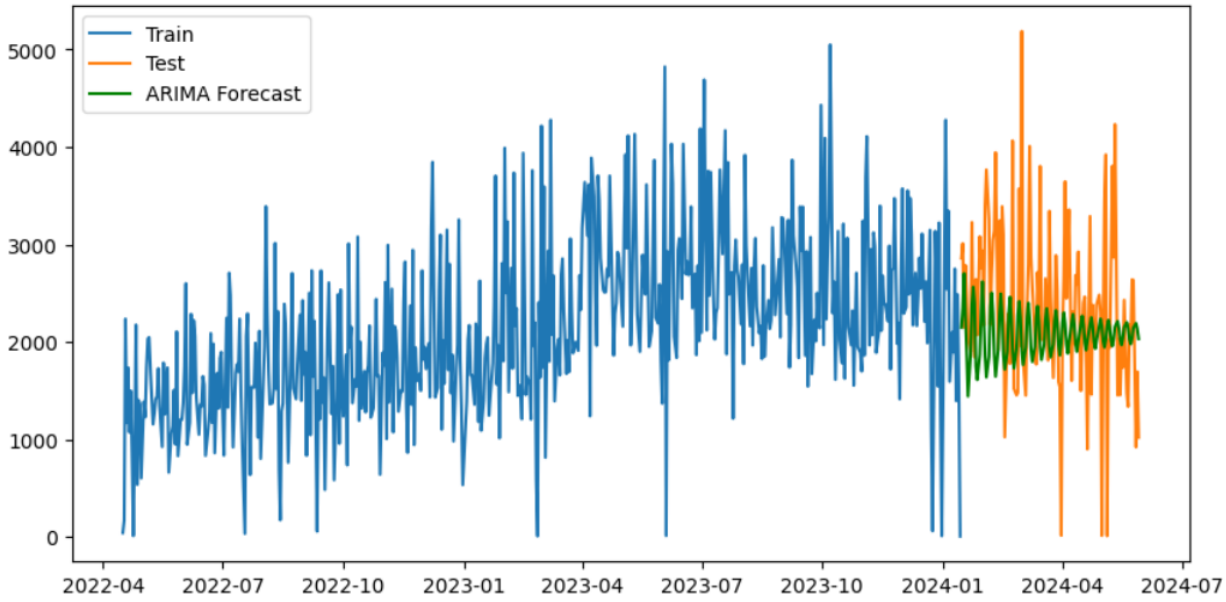
1. **Validación de Estacionariedad:**
   - Se aplicó la prueba Augmented Dickey-Fuller (ADF) a la serie original, arrojando un **p-value = 0,157**, indicando que la serie no era estacionaria.
   - Luego de aplicar una diferenciación, el **p-value = 3.17e-15**, confirmando que la serie se volvió estacionaria.
2. **Parámetros del Modelo:** Utilizando auto_arima, se seleccionaron los parámetros óptimos para el modelo: **ARIMA(5,1,5)**, minimizando criterio AIC.
   - AIC: 8.770,295
   - BIC: 8.817,684
   - HQIC: 8.788,815

3. **Evaluación del Modelo:** El modelo fue evaluado en el conjunto de prueba y las métricas de error obtenidas son:

   - MAE (Error Absoluto Medio): 691,14
   - RMSE (Raíz del Error Cuadrático Medio): 932,64

4. **Comparación Visual:** A continuación, se presenta el gráfico comparativo entre los valores reales y las predicciones del modelo ARIMA:

Gráfico 11. Train, Test y ARIMA Forecast



Fuente: Elaboración propia.

6.1.3 Discusión y Análisis

- **Fortalezas del Modelo:** El modelo ARIMA mostró un buen ajuste a los patrones generales de la serie temporal, especialmente en periodos con comportamientos regulares. La combinación de parámetros $p = 5$, $d = 1$, $q = 5$ permitió capturar tanto la autocorrelación como el ruido presente en los datos.

- **Limitaciones Observadas:** Aunque el modelo logró una buena precisión general, presenta dificultades para capturar picos de demanda o caídas abruptas, lo que se evidencia en el error relativamente alto en la métrica MAE. Esto sugiere que el modelo ARIMA puede no ser la solución más robusta en contextos con alta volatilidad.

- **Recomendación:** ARIMA es útil como línea base por su simplicidad y bajo costo computacional, pero resulta importante compararlo con modelos más avanzados como XGBoost, Prophet o TimeGPT, que tienen mayor capacidad para manejar relaciones no lineales y eventos exógenos.

6.2 XGBoost

6.2.1 Configuración y Ajuste del Modelo

1. **Preparación de los Datos:**

   - Se consolidaron las variables de entrada mediante técnicas de ingeniería de atributos, incluyendo indicadores de eventos festivos, fines de semana y tendencias de ventas.
   - Los datos fueron divididos en conjuntos de entrenamiento (82 %) y prueba (18 %), con un corte temporal en enero de 2024 para simular un entorno de predicción realista.

2. **Optimización de Hiperparámetros:**

   - Se utilizó una combinación de búsqueda en cuadrícula (Grid Search) y aleatoria (Random Search) para encontrar la configuración óptima del modelo. Los mejores parámetros encontrados fueron:
     - α (regulación L1 para reducir overfitting): 3.7557
     - col sample by tree (fracción de columnas usadas en cada árbol): 0.4713
     - γ (reducción mínima en la función de pérdida para dividir un nodo): 0.1025
     - λ (regulación L2 para reducir overfitting): 2.9812
     - Learning rate (tasa de aprendizaje para ajustar pesos): 0.0385
     - Max depth (profundidad máxima de los árboles): 3
     - Min child weight (peso mínimo requerido para dividir un nodo): 5
     - N estimators (cantidad de árboles en el modelo): 301
     - Subsample (fracción de datos utilizada en cada iteración): 0.7865

3. **Validación y Evaluación:**

   - Se aplicó validación cruzada k-fold para garantizar la robustez del modelo frente a datos no vistos.
   - Las métricas de evaluación incluyeron MAE y RMSE para medir la precisión en diferentes horizontes de tiempo.

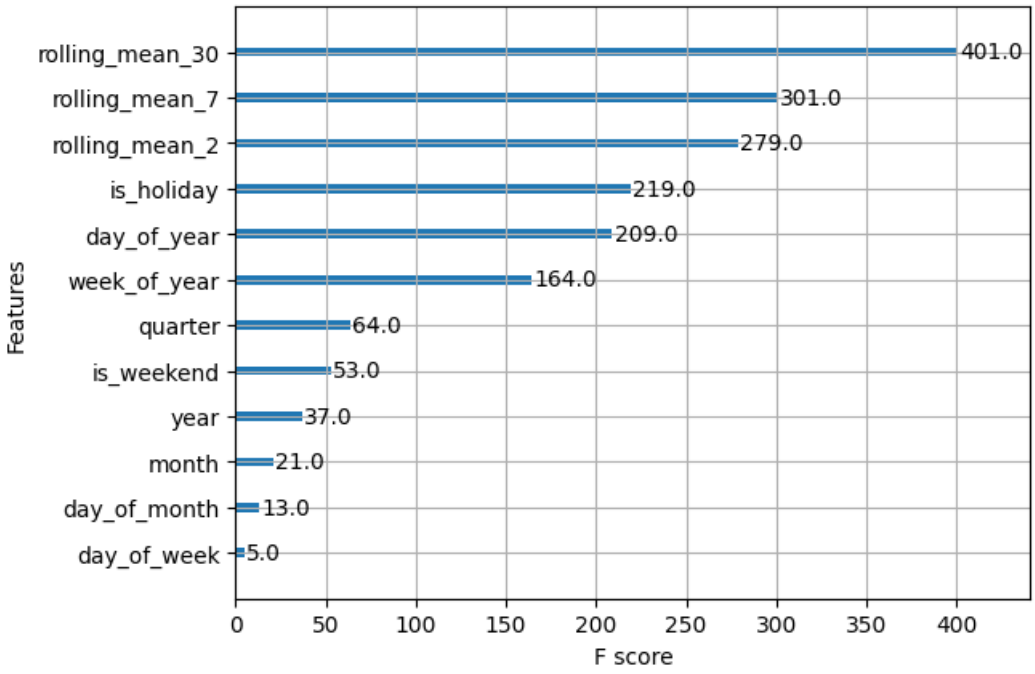
6.2.2 Resultados Obtenidos

El modelo mostró un desempeño sólido en el conjunto de prueba, con los siguientes resultados clave:

- MAE: 439,69
- RMSE: 568,89

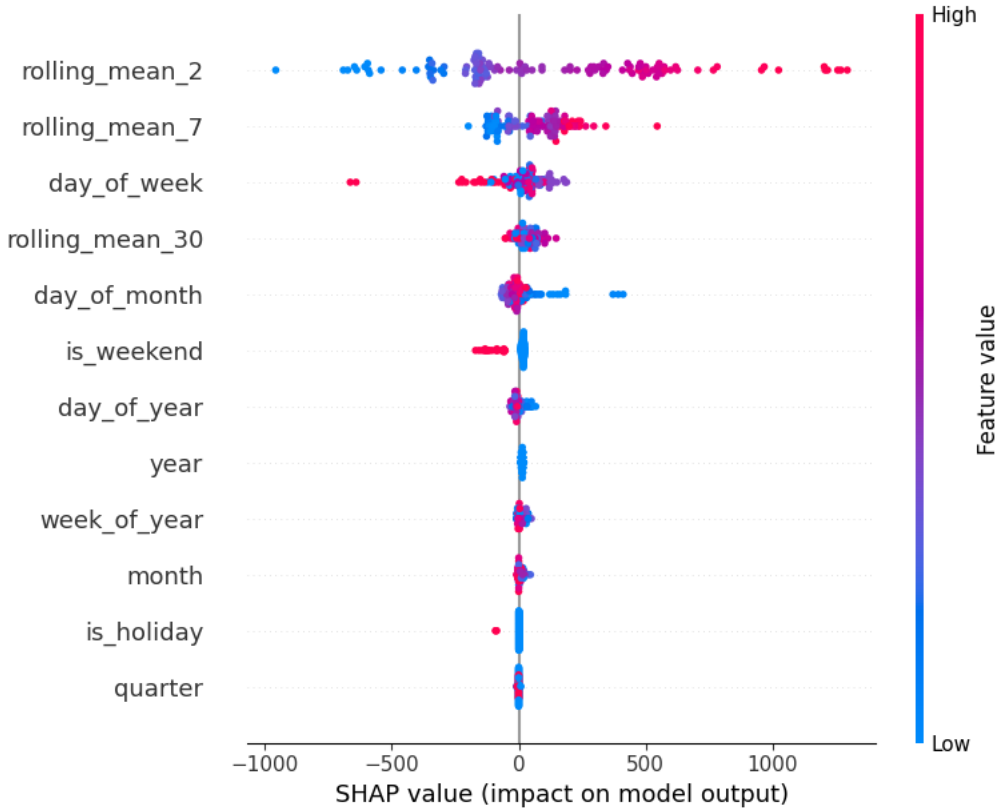
Además, se analizó el impacto de las predicciones utilizando gráficos de importancia de características y valores SHAP para interpretar los resultados.

Gráfico 12. Feature importance según XGBoost



Fuente: Elaboración propia.

Gráfico 13. Análisis con Shap values



Fuente: Elaboración propia.

Los Gráficos 12 y 13 presentan dos enfoques complementarios para interpretar la importancia de las variables en el modelo XGBoost.

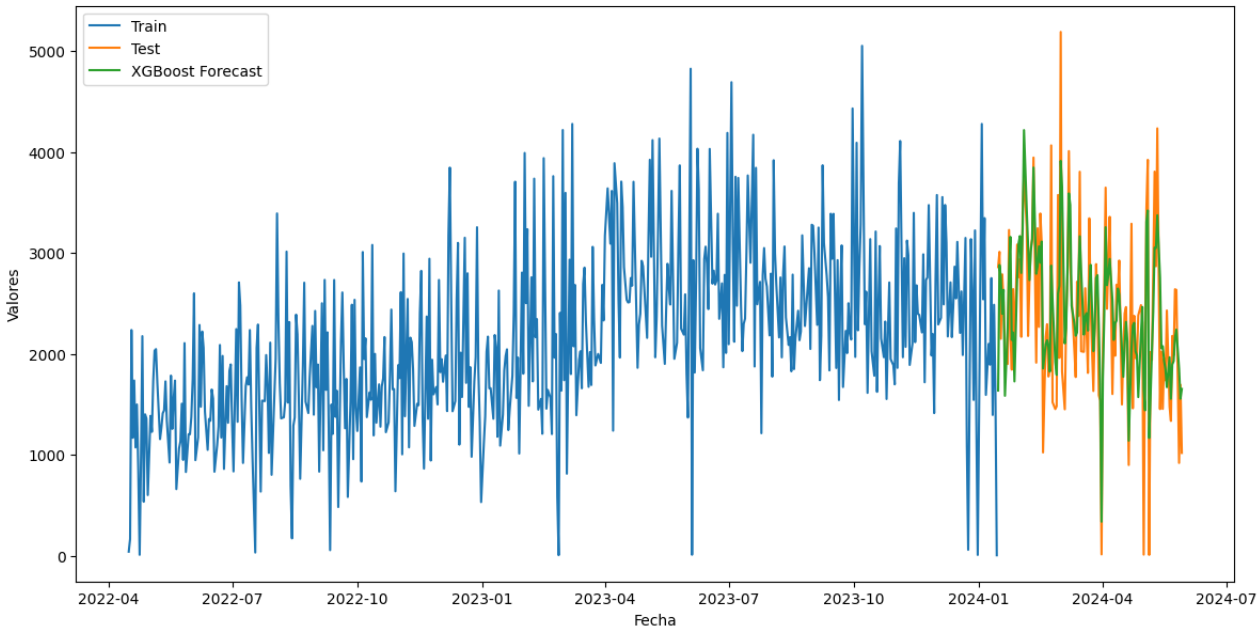
El Gráfico 12 utiliza la métrica nativa de XGBoost basada en **ganancia acumulada**, la cual indica qué variables contribuyen más a reducir el error en los árboles de decisión. Esta representación es útil para identificar qué características fueron más utilizadas en la construcción del modelo.

Por otro lado, el Gráfico 13 emplea **valores SHAP** (SHapley Additive exPlanations), una técnica de interpretabilidad basada en teoría de juegos que no solo cuantifica la importancia de cada variable, sino también su **dirección e impacto** sobre las predicciones individuales. Este enfoque permite comprender mejor cómo cada feature influye en la predicción hacia arriba o hacia abajo, y es especialmente útil para explicar decisiones en entornos reales.

Ambas técnicas coinciden en destacar ciertas variables como clave (por ejemplo, *semana del año* o *promoción activa*), lo cual refuerza su relevancia para el problema de forecasting abordado.

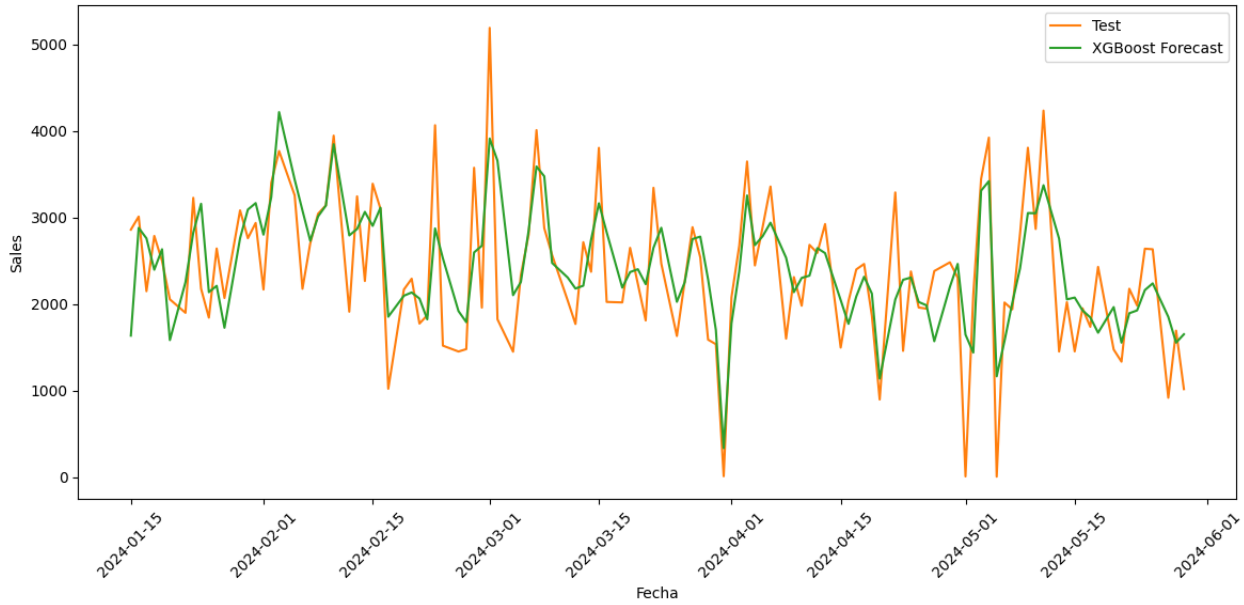
Comparación Visual. A continuación, se presenta el gráfico comparativo entre los valores reales y las predicciones del modelo XGBoost:

Gráfico 14. Train, Test y XGBoost Forecast



Fuente: Elaboración propia.

Gráfico 15. XGBoost Forecast vs Test (zoom)



Fuente: Elaboración propia.

6.2.3 Discusión y Análisis

- **Fortalezas del Modelo:**
  - Alta capacidad para capturar relaciones complejas entre variables.
  - Flexibilidad para incorporar datos exógenos como eventos festivos y tendencias.
  - Resultados interpretables mediante análisis de importancia de características y valores SHAP.
- **Limitaciones Observadas:**
  - A pesar de su precisión, el modelo requiere un mayor costo computacional en comparación con enfoques tradicionales como ARIMA.
  - La interpretación de algunos patrones complejos sigue siendo un desafío para usuarios no técnicos.
- **Recomendación:**
  - XGBoost es una opción sólida para contextos con datos ricos y complejos, especialmente cuando la precisión es prioritaria. Sin embargo, su aplicabilidad práctica debe considerar recursos computacionales y necesidades de interpretabilidad.

6.3 Prophet

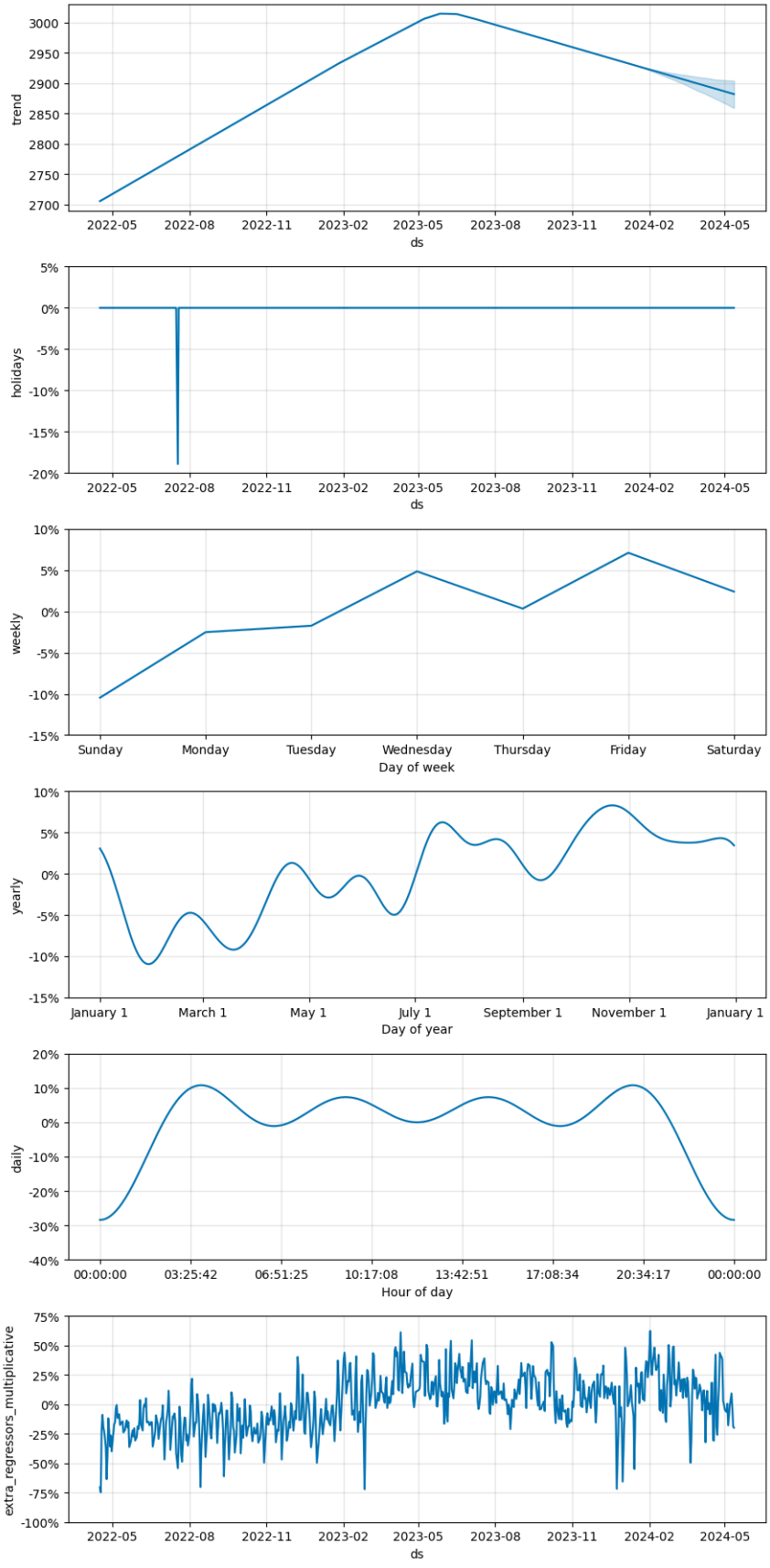
6.3.1 Configuración y Ajuste del Modelo

1. **Preparación de Datos:**
   - Se transformó el dataset de ventas diarias en una serie temporal estructurada con columnas `ds` (fecha) y `y` (demanda observada).
   - Se incluyeron días festivos relevantes para enriquecer el modelo mediante la funcionalidad `add_country_holidays` de Prophet.
2. **Definición de Componentes:**
   - Se configuraron estacionalidades anuales y semanales de forma automática.
   - Se incorporaron regresores adicionales, como eventos especiales y datos climáticos, para capturar patrones exógenos.
3. **Ajuste del Modelo:**
   - El modelo fue entrenado utilizando los datos hasta el 15 de enero de 2024, mientras que el conjunto de prueba comprende datos posteriores a esa fecha.
   - Se optimizaron hiperparámetros como el intervalo de incertidumbre y el ajuste de estacionalidades multiplicativas.

6.3.2 Resultados Obtenidos

1. **Métricas de Evaluación:**
    - MAE: 450,57
    - RMSE: 577,41
2. **Descomposición de Componentes:**
    - El modelo descompone las predicciones en tendencia, estacionalidades y efectos de días festivos.
    - Se identificó que la tendencia general es estable, mientras que los picos de demanda están altamente influenciados por los eventos festivos.

Gráfico 16. Decomposición de componentes

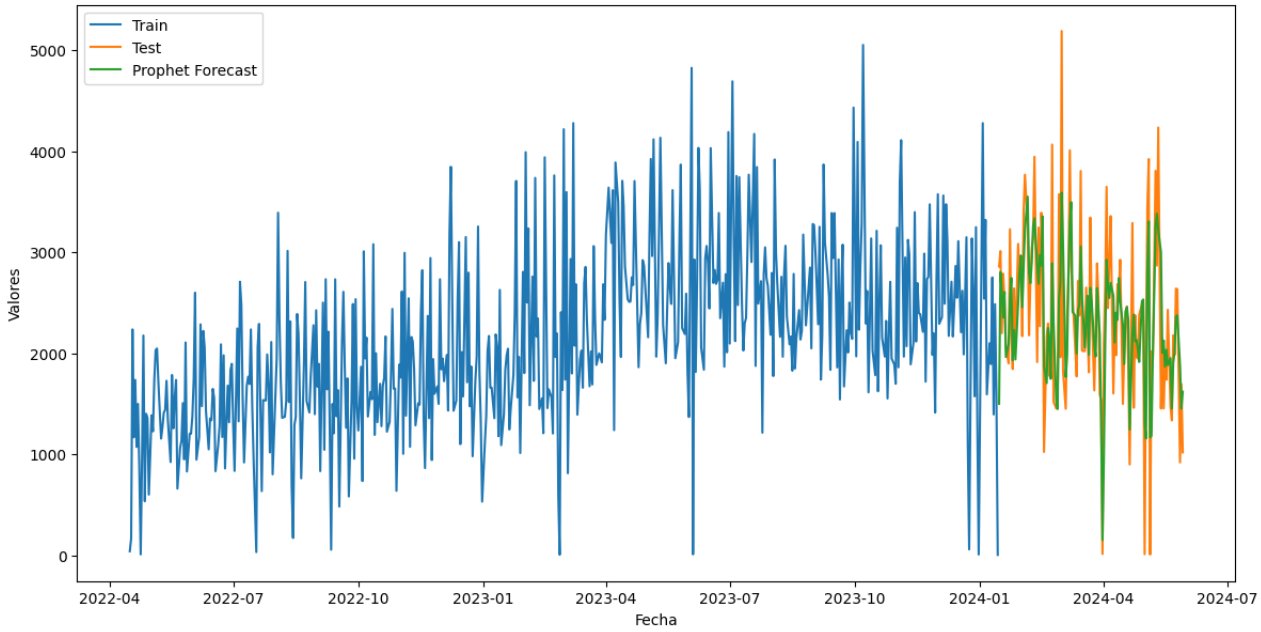


Fuente: Elaboración propia.

3. **Comparación Visual:**

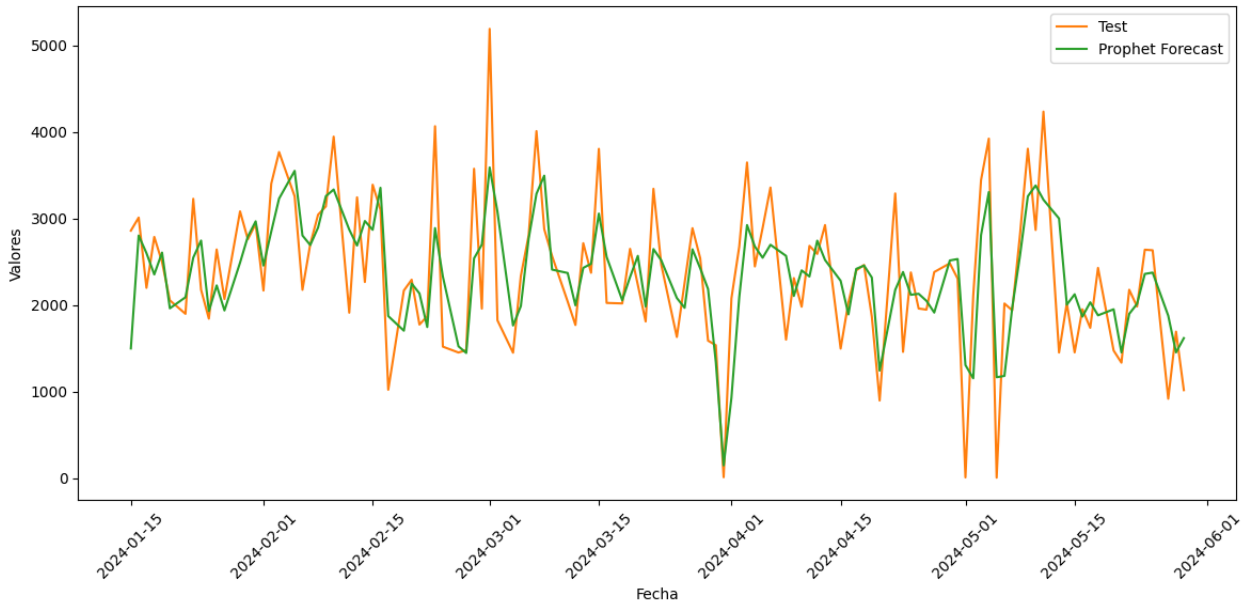
   - En el siguiente gráfico, se presentan las predicciones del modelo Prophet frente a los valores reales del conjunto de prueba.

Gráfico 17. Prophet Forecast vs Test



Fuente: Elaboración propia.

Gráfico 18. Prophet Forecast vs Test (zoom)

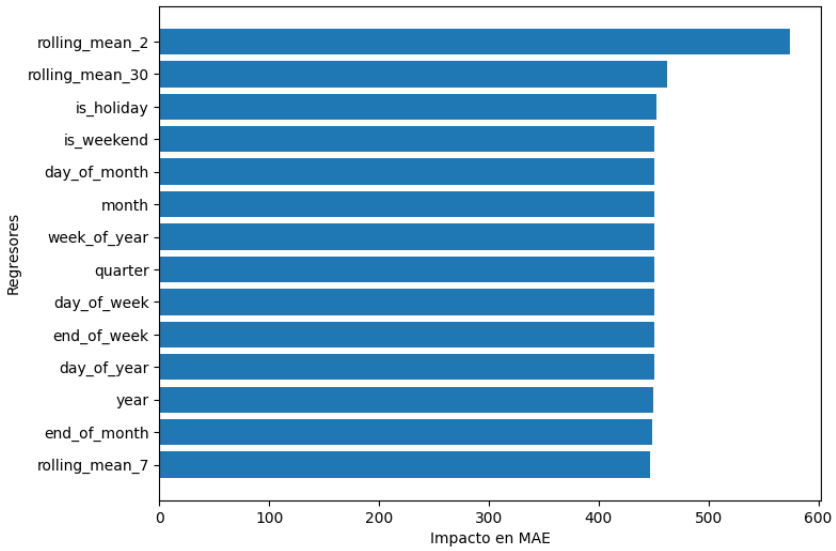


Fuente: Elaboración propia.

4. **Importancia de los Componentes:**

   - Se analizó la contribución relativa de cada componente (tendencia, estacionalidades y regresores) utilizando gráficos específicos de importancia.

Gráfico 19. Feature importance según Prophet



Fuente: Elaboración propia.

6.3.3 Discusión y Análisis

- **Fortalezas:**
  - Prophet logró capturar patrones estacionales complejos y efectos de días festivos, lo que se reflejó en un RMSE competitivo frente a otros modelos.
  - Su capacidad para incorporar regresores adicionales permitió enriquecer las predicciones, mostrando alta flexibilidad.
- **Limitaciones:**
  - Aunque Prophet manejó adecuadamente los patrones generales, tuvo dificultades para capturar ciertos picos abruptos en la demanda.
  - El ajuste de hiperparámetros y la inclusión de regresores aumentaron significativamente el costo computacional.
- **Recomendaciones:**
  - Considerar una combinación de Prophet con modelos más robustos, como XGBoost o TimeGPT, para mejorar la precisión en escenarios de alta volatilidad.
  - Seguir refinando los regresores y evaluar su impacto en futuras iteraciones del modelo.

6.4 TimeGPT

6.4.1 Configuración y Ajuste del Modelo

1. **Preparación de los Datos:**

   - Se consolidaron las variables de entrada mediante técnicas de ingeniería de atributos, incluyendo indicadores de eventos festivos, fines de semana y tendencias de ventas.
   - Los datos fueron divididos en conjuntos de entrenamiento (82 %) y prueba (18 %), con un corte temporal en enero de 2024 para simular un entorno de predicción realista.

2. **Implementación:**

   - El modelo TimeGPT fue configurado utilizando su API oficial, lo que simplifica la integración y elimina la necesidad de un ajuste manual de hiperparámetros.
   - Las series temporales se cargaron directamente en el sistema, y el modelo generó pronósticos semanales en cuestión de segundos.

3. **Ajuste de Hiperparámetros:**

   - Aunque el modelo no requiere ajuste manual, se realizó un análisis exploratorio para determinar la granularidad más adecuada en la predicción.

6.4.2 Resultados Obtenidos

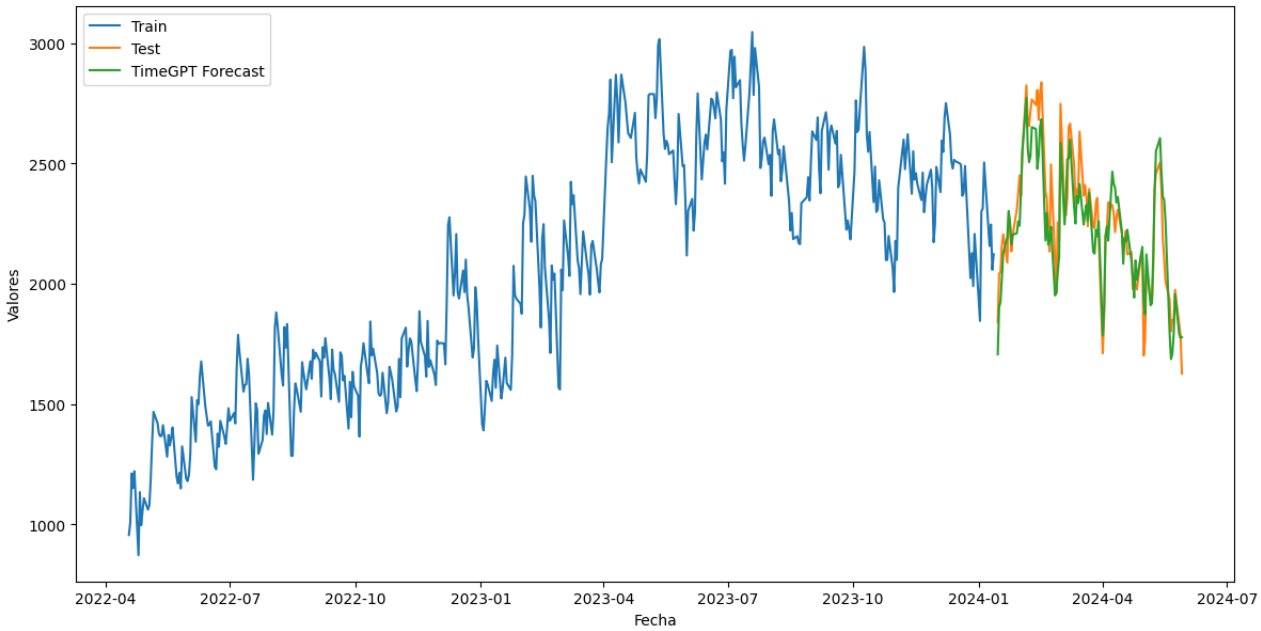
1. **Desempeño del Modelo:**

   - MAE: 94,83
   - RMSE: 117,47

2. **Visualización:**

   - Se generaron gráficos comparativos entre los valores reales y las predicciones del modelo, así como análisis de los errores en distintos horizontes de tiempo. En este caso hemos utilizado un suavizado exponencial, ya que a diferencia de los modelos previos se notó aquí una mejora significativa en lás métricas.

Gráfico 20. TimeGPT Forecast vs Test



Fuente: Elaboración propia.

Gráfico 21. TimeGPT Forecast vs Test (zoom)

Fuente: Elaboración propia.

3. **Eficiencia Computacional:**

   - TimeGPT procesó las predicciones en 7,55 segundos, lo que confirma su idoneidad para escenarios con altos volúmenes de datos.

6.4.3 Discusión y Análisis

1. **Fortalezas del Modelo:**
   - Facilidad de uso mediante API.
   - Alta precisión en contextos con datos no lineales y estacionales.
   - Rapidez en la generación de pronósticos, incluso con múltiples series temporales.
2. **Limitaciones Observadas:**
   - Dependencia de la conexión y latencia de la API.
   - Costos asociados al uso de la herramienta en proyectos de gran escala.
3. **Recomendaciones:**
   - Integrar TimeGPT como parte de un sistema híbrido, complementándolo con modelos interpretables como Prophet o XGBoost para justificar los drivers de los pronósticos.
   - Evaluar escenarios de escalabilidad y costos para asegurar la viabilidad del modelo en implementaciones prácticas.

7 Resultados y Análisis

7.1 Evaluación de la Precisión de los Modelos

Se realizó una comparación entre los modelos ARIMA, XGBoost, Prophet y TimeGPT utilizando métricas estándar como MAE (Mean Absolute Error) y RMSE (Root Mean Squared Error). Los resultados obtenidos se presentan en la siguiente tabla:

Tabla 2. Resumen general de métricas

Modelo	Tiempo de ejecución (seg)	MAE	RMSE
ARIMA	69,46	691,14	932,64
XGBoost	8.707,60	439,69	568,89
Prophet	22,62	450,57	577,41
TimeGPT	**7,55**	**94,83**	**117,47**

Fuente: Elaboración propia.

- **ARIMA**:
  - **MAE = 691,14** → En promedio, las predicciones de ARIMA se desvían **691 unidades** del valor real de la demanda.
  - **RMSE = 932,64** → Este valor indica que los errores más grandes afectan significativamente la precisión del modelo. ARIMA tiende a cometer errores más elevados en ciertos puntos, lo que sugiere que no se adapta bien a cambios abruptos en la demanda.

Conclusión: ARIMA es un modelo útil cuando la demanda sigue patrones simples y estables, pero su alto error indica que no es la mejor opción para entornos con fluctuaciones o tendencias cambiantes.

- **XGBoost**:
  - **MAE = 439,69** → En promedio, las predicciones de XGBoost se desvían **439 unidades** del valor real, una mejora significativa frente a ARIMA.
  - **RMSE = 568,89** → Aunque los errores más grandes siguen presentes, su menor valor en comparación con ARIMA indica que XGBoost logra un ajuste más preciso.

  Conclusión: XGBoost es capaz de capturar relaciones más complejas en los datos y reduce los errores absolutos. Es una mejor opción para datos con patrones no lineales y cambios en la demanda.

- **Prophet**:
  - **MAE = 450,57** → Las predicciones de Prophet se desvían **450 unidades** en promedio, similar a XGBoost.
  - **RMSE = 577,41** → Este error es un poco mayor que el de XGBoost, lo que indica que algunos errores grandes pueden afectar sus predicciones.

  Conclusión: Prophet es especialmente útil para detectar estacionalidades y patrones cíclicos en la demanda. Aunque su error es un poco mayor que XGBoost en valores extremos, puede ser la mejor opción en escenarios donde la demanda sigue ciclos repetitivos.

- **TimeGPT**:
  - **MAE = 94,83** → En promedio, las predicciones de TimeGPT se desvían **94 unidades**, lo que indica un alto error absoluto.
  - **RMSE = 117,47** → El error cuadrático medio es el más bajo, lo que significa que en ciertos momentos los errores más grandes afectan menos a la precisión del modelo.

  Conclusión: TimeGPT se adapta muy bien a entornos con alta volatilidad. Su enfoque es detectar cambios abruptos y los datos demuestran que para este trabajo es el mejor en base a las métricas obtenidas.

7.2 Evaluación de Significancia: Test de Diebold-Mariano

Para evaluar si las diferencias en el rendimiento de los modelos de predicción son estadísticamente significativas, se realizó una aproximación del **Test de Diebold-Mariano**. En este contexto, se contrastaron las predicciones de cada modelo frente al benchmark (ARIMA), utilizando como métrica el error cuadrático (squared error).

La hipótesis nula del test (H_0) establece que no hay diferencia en la capacidad predictiva entre ambos modelos, es decir, que el valor esperado de la diferencia de errores es cero. La hipótesis alternativa (H_1), en este caso bilateral, plantea que existe una diferencia

significativa en la precisión de los modelos comparados.

Formalmente:

- H_0: $E[d_t] = 0$ (los modelos tienen igual precisión)
- H_1: $E[d_t] \neq 0$ (uno de los modelos tiene mejor precisión)

Donde d_t representa la diferencia entre los errores de predicción al tiempo t. Un valor de $p < 0{,}05$ se interpretó como evidencia suficiente para rechazar la hipótesis nula y concluir que existe una diferencia estadísticamente significativa entre los modelos comparados.

7.3 Resultados del Test

A continuación, se presentan los resultados del test de Diebold-Mariano aproximado, donde se compararon los modelos en función de sus valores de MAE y RMSE.

Tabla 3. Resultados del Test de Diebold-Mariano

Modelo 1	Modelo 2	t-stat	p-value
ARIMA	XGBoost	4.19	0.00003
ARIMA	Prophet	4.01	0.00006
ARIMA	TimeGPT	9.94	< 0.00001
XGBoost	Prophet	-0.17	0.864
XGBoost	TimeGPT	5.75	< 0.00001
Prophet	TimeGPT	5.93	< 0.00001

Fuente: Elaboración propia.

7.4 Análisis de Resultados

A partir de los resultados obtenidos, se pueden extraer las siguientes conclusiones:

- **Diferencias significativas** (p-value <0.05):
  - XGBoost tiene un MAE significativamente menor que ARIMA ($p = 0{,}00003$).
  - Prophet también supera significativamente a ARIMA ($p = 0{,}00006$).
  - TimeGPT es significativamente mejor que ARIMA ($p < 0{,}00001$).
  - TimeGPT es significativamente mejor que XGBoost ($p < 0{,}00001$).
  - TimeGPT también supera significativamente a Prophet ($p < 0{,}00001$).
- **Diferencias NO significativas** (p-value >0.05):
  - XGBoost y Prophet tienen desempeños similares ($p = 0{,}864$).

7.5 Conclusión General

A partir de estos resultados, se concluye que tanto XGBoost como Prophet superan significativamente a ARIMA. Ambos muestran desempeños similares, por lo que cualquiera de los dos es una opción viable en términos de precisión. TimeGPT presenta mejoras significativas sobre todos los modelos.

Dado esto, si el objetivo es maximizar la precisión del pronóstico, el modelo **TimeGPT** es la mejor opción, aunque Prophet o XGBoost también son competitivos dependiendo de las necesidades del negocio.

7.6 Análisis de Resultados según el Contexto del Retail

En el contexto del retail, la precisión de la predicción de demanda impacta directamente en la gestión de inventarios, la planificación de la cadena de suministro y la eficiencia operativa. XGBoost demostró ser particularmente eficaz en entornos con múltiples variables exógenas, como promociones y eventos especiales. Prophet destacó por su capacidad de capturar patrones estacionales y eventos festivos, lo que lo hace ideal para negocios con ciclos de demanda predecibles. Finalmente TimeGPT, mostró un menor error absoluto y fue muy útil en escenarios de alta volatilidad debido a su capacidad para adaptarse rápidamente a cambios abruptos en la demanda.

7.7 Implementación Práctica en la Empresa

7.7.1 Puesta en producción

La implementación de los modelos en entornos productivos requiere un enfoque escalable y flexible. Se recomienda un enfoque híbrido que combine la robustez de XGBoost con la capacidad de adaptación de TimeGPT para escenarios dinámicos.

7.7.2 Herramientas a utilizar

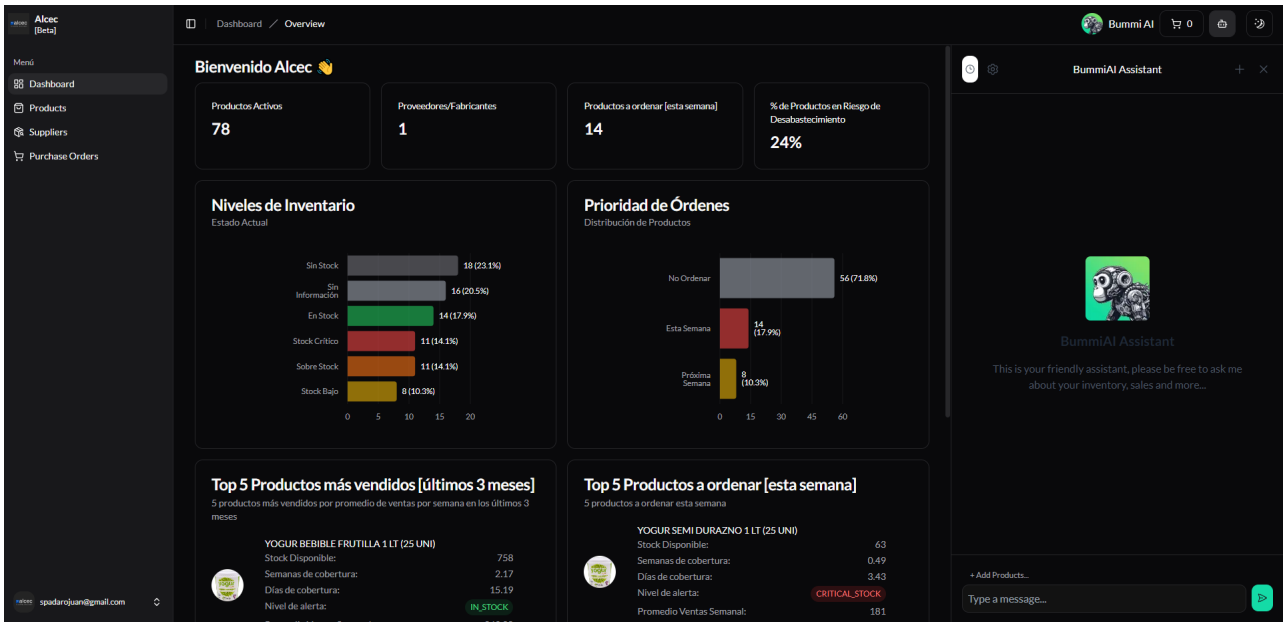
Se propone el uso de herramientas como Docker para la contenedorización de los modelos, Azure Functions para la orquestación de las predicciones en la nub, y un front end propio desarrollado para la visualización de resultados y análisis de datos en tiempo real.

7.7.3 Primer MVP

Como resultado del proceso de investigación y modelado, se desarrolló un MVP (Producto Mínimo Viable) que integra las predicciones generadas por los modelos en una interfaz visual interactiva, orientada a facilitar la toma de decisiones operativas en la empresa. Esta primera versión fue desarrollada con el objetivo de validar rápidamente el valor práctico del sistema en un entorno real.

El gráfico 22 muestra la vista general del dashboard principal. En ella se puede observar la demanda estimada, la comparación con el stock disponible y un resumen de alertas de reabastecimiento. Este panel actúa como un centro de control para el usuario.

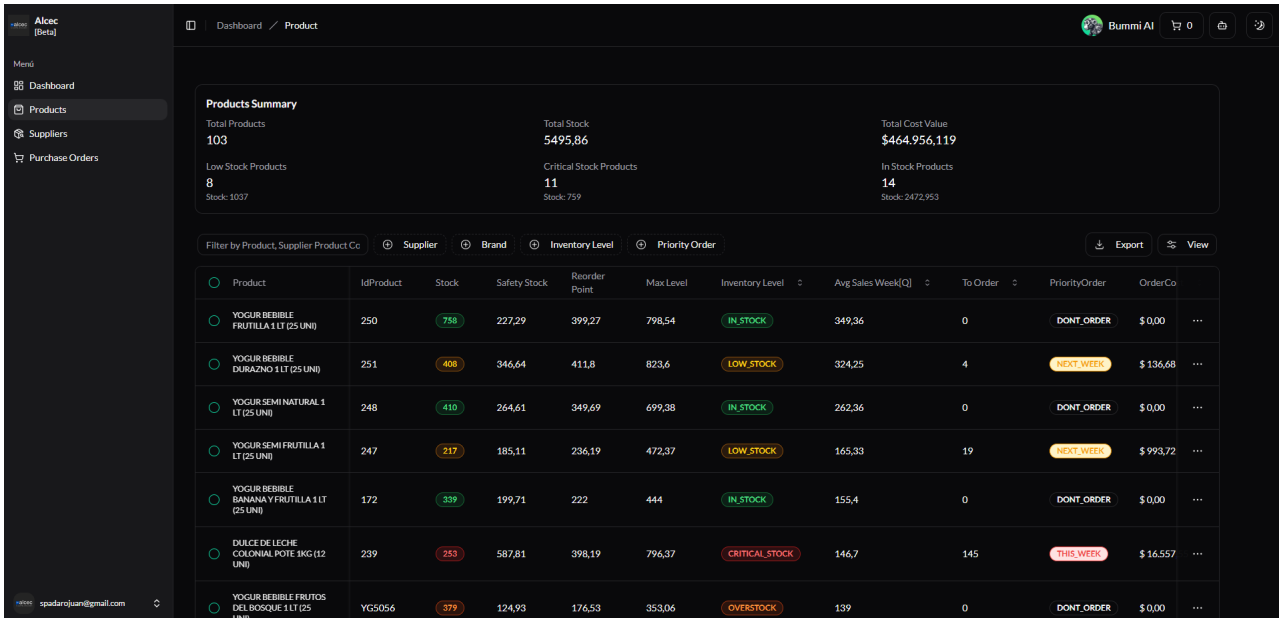
Gráfico 22. MVP versión 1 - Vista del dashboard general y copiloto de IA



Fuente: Elaboración propia.

A continuación, se presenta una vista más detallada por producto. Esta funcionalidad permite desagregar la información y analizar la evolución de la demanda esperada, las compras realizadas y las decisiones recomendadas a nivel individual, facilitando una mejor gestión del inventario.

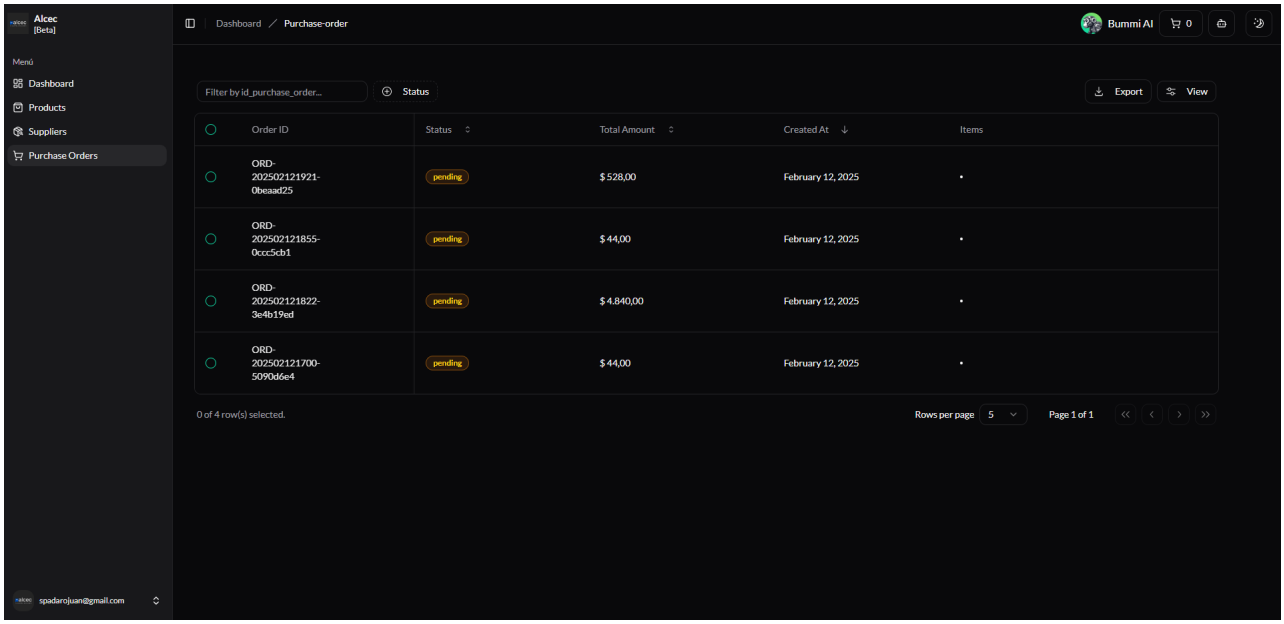
Gráfico 23. MVP versión 1 - Vista por productos



Fuente: Elaboración propia.

En el gráfico 24 se muestra la sección de órdenes de compra, donde se detallan los productos sugeridos para reabastecimiento, su prioridad, cantidad sugerida, y el estado actual del pedido. Esta automatización busca reducir el trabajo manual del equipo de compras y minimizar errores humanos.

Gráfico 24. MVP versión 1 - Vista de las órdenes de compra y su estado



Fuente: Elaboración propia.

Finalmente, el gráfico 25 muestra la integración con WhatsApp mediante un asistente conversacional, que permite consultar en lenguaje natural las recomendaciones del sistema. Este copiloto de IA responde consultas como "¿Qué productos tengo que pedir esta semana?" o "¿Cuánto debería pedir del producto X?", acercando la tecnología al equipo operativo.

Gráfico 25. MVP versión 1 - Vista del Agente de IA integrado a WhatsApp

Fuente: Elaboración propia.

Esta primera versión del MVP fue presentada al equipo directivo de la empresa y validada como base para una futura integración completa en el sistema operativo interno.

7.8 Cálculo de Costos/Beneficios Económicos

7.8.1 Introducción

En esta sección se analizarán los costos y beneficios económicos de cada modelo de predicción de demanda (ARIMA, XGBoost, Prophet y TimeGPT) con el objetivo de determinar cuál ofrece la mejor relación costo-beneficio para la empresa del estudio de caso. Se evaluarán aspectos como los costos computacionales, costos operativos, ahorro en inventarios y eficiencia en la toma de decisiones.

7.8.2 Costos Asociados a la Implementación de Modelos

Cada modelo de predicción implica distintos costos de implementación y operación, los cuales pueden clasificarse en:

- **Costo computacional**: tiempo de cómputo, uso de infraestructura en la nube, necesidad de GPU o servidores dedicados.
- **Costo de desarrollo**: tiempo de trabajo requerido para ajuste de hiperparámetros, preprocesamiento de datos y mantenimiento del modelo.
- **Dificultad de integración**: esfuerzo necesario para implementar el modelo en la infraestructura tecnológica existente.

Para definir los niveles de costo (*bajo, medio, alto*) en cada categoría, se utilizaron los siguientes criterios:

- **Costo computacional:**
  - **Bajo**: Tiempo de ejecución < 1 minuto, sin necesidad de GPU ni servidores avanzados. Puede correr en una laptop común.
  - **Medio**: Tiempo de ejecución entre 1 y 10 minutos, con uso moderado de cloud computing o infraestructura optimizada.
  - **Alto**: Tiempo de ejecución > 10 minutos, requiere infraestructura en la nube, GPU o cómputo paralelo, con costos en AWS, Azure o GCP.
- **Costo de desarrollo:**
  - **Bajo**: < 10 horas de trabajo. Implementación sencilla con pocos ajustes manuales.
  - **Medio**: Entre 10 y 40 horas. Requiere optimización de hiperparámetros y pruebas en distintos escenarios.
  - **Alto**: > 40 horas. Necesita ajuste complejo, validación cruzada avanzada y pruebas exhaustivas.
- **Dificultad de integración:**
  - **Baja**: El modelo genera resultados en formatos estándar (CSV, JSON) fácilmente integrables en sistemas existentes.
  - **Media**: Requiere algunos ajustes en la infraestructura actual, pero no demanda una reestructuración profunda.
  - **Alta**: Necesita infraestructura adicional (nuevas bases de datos, APIs externas, reentrenamiento frecuente, servidores dedicados).

La siguiente tabla resume los costos asociados a cada modelo en función de estos criterios:

Tabla 4. Costos estimados por modelo

Modelo	Costo Computacional	Costo de Desarrollo	Dificultad de Integración
ARIMA	Bajo	Medio	Baja
XGBoost	Medio	Alto	Media
Prophet	Medio	Bajo	Baja
TimeGPT	Alto	Bajo	Alta

Fuente: Elaboración propia.

Se observa que: ARIMA es el modelo con menor costo computacional e integración más sencilla, pero con requerimientos medios de ajuste. XGBoost presenta costos computacionales y de desarrollo elevados, debido a su necesidad de optimización de hiperparámetros y ajuste de variables. Prophet es una alternativa de menor complejidad, con costos computacionales y de desarrollo moderados. TimeGPT tiene el mayor costo computacional y dificultad de integración, debido a su dependencia de una API externa y costos asociados a su uso.

La elección del modelo dependerá de la disponibilidad de recursos computacionales y la facilidad de implementación dentro de la infraestructura de la empresa.

7.8.3 Beneficios Económicos de la Implementación

Los beneficios de un buen modelo de predicción de demanda pueden medirse en términos de:

- **Reducción del stock excedente**: Menos costos de almacenamiento y menos desperdicio de productos perecederos.
- **Menos quiebres de stock**: Aumento en la disponibilidad de productos sin generar sobrestock.
- **Optimización de costos de compra**: Posibilidad de negociar mejores precios con proveedores gracias a una planificación más precisa.
- **Reducción de costos operativos**: Menos urgencias logísticas, reducción en costos de distribución por pedidos imprevistos.

Tabla 5. Impacto económico estimado por modelo

Modelo	Reducción de Stock (%)	Menos Quiebres de Stock (%)	Ahorro Operativo (%)
ARIMA	10 %	5 %	7 %
XGBoost	20 %	15 %	12 %
Prophet	18 %	12 %	10 %
TimeGPT	25 %	18 %	15 %

Fuente: Elaboración propia.

7.8.4 Comparación del Retorno de Inversión (ROI)

Para evaluar qué modelo representa una mejor inversión, se calculará el ROI utilizando la siguiente fórmula:

$$ROI = \frac{\text{Beneficio Neto}}{\text{Costo Total de Implementación}} \times 100 \quad (1)$$

Donde el **Beneficio Neto** es la suma de los ahorros en inventario, costos operativos y mejoras en ventas.

7.8.5 Conclusión y Recomendaciones

A partir del análisis de costos y beneficios económicos, **si la empresa prioriza bajos costos computacionales y facilidad de implementación**, ARIMA o Prophet serían opciones recomendadas. **Si se busca un mayor ahorro y optimización operativa a largo plazo**, XGBoost ofrece un mejor balance entre inversión y beneficio. **Si se dispone de mayor presupuesto y se necesita alta precisión en escenarios complejos**, TimeGPT puede ser la mejor alternativa.

Finalmente, se recomienda evaluar la combinación de modelos en función del tipo de productos y la estructura operativa de la empresa, considerando el trade-off entre inversión inicial y beneficios económicos obtenidos.

8 Conclusiones

8.1 Resumen de Hallazgos

Esta investigación comparó cuatro modelos de predicción de demanda: ARIMA, XGBoost, Prophet y TimeGPT. Los resultados muestran que ARIMA sirvió como modelo base, mostrando limitaciones en la captura de dinámicas complejas, pero destacando por su simplicidad y bajo costo computacional. XGBoost obtuvo el mejor rendimiento en términos de precisión absoluta (RMSE), gracias a su capacidad para integrar múltiples variables y capturar relaciones no lineales. Prophet sobresalió en la identificación de patrones estacionales y la incorporación de eventos especiales, lo que lo hace ideal para negocios con ciclos de demanda definidos. TimeGPT, aunque presentó mayor error absoluto, demostró una notable adaptabilidad en escenarios de alta volatilidad.

8.2 Recomendaciones para la Industria

Con base en los hallazgos, se proponen las siguientes recomendaciones para empresas del sector retail y venta al por mayor. En **entornos estables**, se recomienda el uso de Prophet, aprovechando su capacidad para capturar estacionalidades y eventos recurrentes. Para **escenarios volátiles o con múltiples variables exógenas**, XGBoost ofrece mayor precisión y flexibilidad, siendo adecuado para optimizar inventarios y reducir costos operativos. **TimeGPT** es recomendable para empresas con grandes volúmenes de datos y necesidades de rápida adaptación a cambios abruptos en la demanda. Implementar métodos de cuantificación de incertidumbre, como **Conformal Prediction**, para mejorar la gestión de riesgos asociados a la planificación de la demanda.

8.3 Líneas Futuras de Investigación y Aplicación

Esta investigación sienta las bases para una serie de desarrollos futuros que pueden enriquecer significativamente la precisión y utilidad de los modelos de predicción de demanda.

Algunos ejemplos son: **Integración de modelos híbridos**, combinando la robustez estructural de algoritmos como XGBoost con la capacidad de generalización de modelos generativos como TimeGPT, con el objetivo de aprovechar lo mejor de ambos mundos. **Incorporación de regresores dinámicos**, como señales provenientes de redes sociales, movilidad, búsquedas online o comportamiento de compra en tiempo real, que podrían anticipar cambios abruptos en la demanda antes de que se vean reflejados en los datos históricos. **Optimización de costos computacionales y sostenibilidad**, mediante el análisis del consumo de recursos en la nube y la búsqueda de soluciones que balanceen eficiencia y escalabilidad, especialmente para PyMEs con recursos limitados. **Exploración de métodos de predicción probabilística como Conformal Prediction**, que permiten generar intervalos de confianza calibrados para cada forecast. Esto brindaría mayor transparencia y seguridad en la toma de decisiones operativas bajo incertidumbre, especialmente en productos de alta volatilidad. **Transferencia a otros sectores industriales**, como logística, salud o energía, donde los desafíos de predicción de demanda presentan características similares, permitiendo validar la generalización de los hallazgos obtenidos en el retail alimentario.

En síntesis, la predicción de demanda requiere enfoques cada vez más inteligentes, interpretables y adaptables. Los avances en modelos y técnicas complementarias como Conformal Prediction representan oportunidades concretas para aumentar la confiabilidad y el impacto de estas herramientas en entornos productivos reales.

Finalmente, otra línea futura relevante consiste en desagregar las predicciones a nivel de familia de productos. Si bien el MVP desarrollado para la empresa se centró en una predicción agregada de la demanda total, se identificó que una segmentación por grupos permitiría afinar la planificación de inventarios y compras, adaptándola a las dinámicas específicas de cada categoría del catálogo. Esta funcionalidad se proyecta para una segunda iteración del sistema, con el objetivo de aumentar la precisión operativa y el valor estratégico del modelo.

9 Bibliografía

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.

Cohen, M. C., Gras, P.-E., Pentecoste, A., & Zhang, R. (2022). *Demand prediction in retail: A practical guide to leverage data and predictive analytics*. Springer.

García, M., & Pérez, L. (2022). Análisis comparativo de modelos tradicionales y modernos para pronósticos [Consultado en mayo de 2025]. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo (RIDE)*. https://www.ride.org.mx/index.php/RIDE/article/download/1203/3556/

Garza, A., Challu, C., & Mergenthaler-Canseco, M. (2023). TimeGPT-1. *arXiv preprint arXiv:2310.03589*.

Haque, M. S., Amin, M. S., & Miah, J. (2023). Retail demand forecasting: a comparative study for multivariate time series. *arXiv preprint arXiv:2308.11939*.

Hasan, M. R., Kabir, M. A., Shuvro, R. A., & Das, P. (2022). A comparative study on forecasting of retail sales. *arXiv preprint arXiv:2203.06848*.

Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction.

Hochreiter, S. (1997). Long Short-term Memory. *Neural Computation MIT-Press*.

Hyndman, R. (2018). *Forecasting: principles and practice*. OTexts.

Liu, C., & Sustik, M. A. (2021). Elasticity based demand forecasting and price optimization for online retail. *arXiv preprint arXiv:2106.08274*.

López-Machado, H. A. (2024). Modelos econométricos para predecir el crecimiento económico de América Latina [Consultado en mayo de 2025]. *Revista de Ciencias Económicas*. https://dialnet.unirioja.es/descarga/articulo/9282009.pdf

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International journal of forecasting*, *16*(4), 451-476.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, *13*(3), e0194889.

Simmhan, Y., & Ranganathan, A. (2023). TimeGPT: Generative Time Series Forecasting for Large-Scale Demand. Proceedings of the ACM on Measurement and Analysis of Computing Systems. *7(1), 1-22*.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37-45.

Vashishtha, R. K., Burman, V., Kumar, R., Sethuraman, S., Sekar, A. R., & Ramanan, S. (2020). Product age based demand forecast model for fashion retail. *arXiv preprint arXiv:2007.05278*.

10 Apéndices

10.1 Transformación y Limpieza de datos

En esta sección se documentan las transformaciones y el proceso de limpieza aplicado a los datos antes de su análisis y modelado. Las principales tareas incluyen:

10.1.1 Conversión de tipos de datos

Se aseguró que las columnas de fechas estuvieran en el formato adecuado.
Esto permitió generar variables adicionales derivadas de las fechas:

- Día de la semana: Clasificación de los días de lunes (0) a domingo (6).
- Mes y año: Identificación del periodo temporal.
- Trimestre y semana del año.

10.1.2 Filtrado según tipo de moneda

Por una cuestión de simpleza, se decidió filtrar aquellos datos en U$S y sólo trabajar con la moneda local.

Gráfico 26. Monto de venta por tipo de moneda

moneda_venta	Monto Total	Porcentaje del Total
$	1,460,236.00	99.99%
U$S	140.00	0.01%

Fuente: Elaboración propia.

10.1.3 Identificación y tratamiento de valores nulos

Si bien la estrategia inicial fue la eliminación de filas cuando el porcentaje de datos faltantes superaba el 30 %, no se han identificado columnas con valores faltantes por lo que no se realizaron acciones correspondientes. La única excepción fue al utilizar el modelo de TimeGPT, en donde fue necesario agregar con 0 los días domingos por como trabaja el modelo en sí.

10.1.4 Detección de outliers

Se utilizó el rango intercuartílico (IQR) para identificar valores atípicos.

- Umbral utilizado: 1.5 veces el rango intercuartílico.
- Porcentaje de observaciones eliminadas: Aproximadamente el 0.6 % de los datos.

10.1.5 Incorporación de eventos especiales

Se incluyeron eventos especiales y feriados como factores externos que podrían influir en las ventas. Los eventos se generaron para un rango de años (2022-2026) y se definió una ventana de impacto de 3 días antes y 3 días después del evento.

- **Impacto en los datos:** Estos eventos se fusionaron con el dataset principal y se codificaron como variables categóricas para evaluar su influencia en las ventas:
  - **Feriados identificados:** 11 eventos principales repetidos anualmente.
  - **Ventana de impacto:** 7 días (3 días antes, el día del evento y 3 días después).

Gráfico 27. Eventos especiales utilizados

	holiday	ds	lower_window	upper_window
0	Año Nuevo	2022-01-01	-3	3
1	Reyes	2022-01-06	-3	3
2	Carnaval	2022-02-12	-3	3
3	Semana Santa	2022-03-28	-3	3
4	Día del Trabajador	2022-05-01	-3	3
5	Natalicio de Artigas	2022-06-19	-3	3
6	Jura de la Constitución	2022-07-18	-3	3
7	Declaratoria de la Independencia	2022-08-25	-3	3
8	Día de la Raza	2022-10-12	-3	3
9	Día de los Difuntos	2022-11-02	-3	3
10	Navidad	2022-12-25	-3	3

Fuente: Elaboración propia.

10.1.6 Normalización de variables

Se aplicó escalado min-max a variables continuas como ingresos y precios para mejorar la estabilidad del modelo.

Este proceso garantizó datos limpios y preparados para el análisis, reduciendo el ruido y mejorando la calidad de los modelos.

10.2 Código Fuente

Todo el código fuente con los análisis correspondiente mencionados en esta tesis se encuentran documentados en el siguiente repositorio público de GitHub:
https://github.com/juanspadaro/forecasting_for_planning