UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO

FACULTAD DE POSGRADO

MAESTRÍA EN CIENCIA DE DATOS

Proyecto de Investigación previo a la obtención del Grado académico de Magíster en Ciencia de Datos

TEMA

"OPTIMIZACIÓN DE INVENTARIO EN LA GESTIÓN DE LA CADENA DE SUMINISTRO EN EMPRESAS DE CONSUMO MASIVO"

AUTOR

ING. ÁLVAREZ CARPIO GILBERTO GERMÁN

DIRECTOR

ING. DÍAZ MACÍAS EFRAÍN EVARISTO, M. SC.

QUEVEDO – ECUADOR

2024

CERTIFICACIÓN

Ing. Efraín Evaristo Díaz Macías, M. Sc. director del Proyecto de Investigación, previo a la obtención del Grado Académico de Magíster en Ciencia de Datos.

CERTIFICA:

Que el Ing. **Gilberto Germán Álvarez Carpio**, ha cumplido con la elaboración del Proyecto de Investigación titulado: **"Optimización de inventario en la gestión de la cadena de suministro en empresas de consumo masivo";** el mismo que se encuentra apto para la presentación y sustentación respectiva.

Quevedo, 06 de agosto del 2024

Firmado electrónicamente por:
EFRAIN EVARISTO
DIAZ MACIAS

Ing. Efraín Evaristo Díaz Macías, M. Sc.

DIRECTOR

AUTORÍA

El presente proyecto de investigación titulado: "OPTIMIZACIÓN DE INVENTARIO EN LA GESTIÓN DE LA CADENA DE SUMINISTRO EN EMPRESAS DE CONSUMO MASIVO" es un trabajo investigativo original, elaborado con esfuerzo y dedicación por parte del estudiante de la Universidad Técnica Estatal de Quevedo: ING. ÁLVAREZ CARPIO GILBERTO GERMÁN, con cédula de ciudadanía número 050405199-6, declaro que los criterios, marco contextual, marco teórico, metodología y propuesta de desarrollo son de mi exclusiva responsabilidad.

Ing. Gilberto Germán Álvarez Carpio

AUTOR

DEDICATORIA

A mi familia por brindarme la oportunidad

de formarme como profesional.

AGRADECIMIENTO

Mi más sincero agradecimiento a todas las instituciones que aportaron la información necesaria para la realización de este trabajo. A la Universidad Técnica Estatal de Quevedo y su cuerpo docente por los conocimientos impartidos. A mi director, el Ing. Efraín Evaristo Díaz Macías, MSc, por su guía en esta investigación. Y en especial, a la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) por su programa de "Becas Nacionales de Posgrado Fortalécete 2022", que me ha brindado una invaluable oportunidad para continuar con mis estudios de posgrado.

PRÓLOGO

La gestión de inventario en el entorno empresarial actual es el pilar de la sostenibilidad y éxito de las organizaciones, por lo que se requiere optimizar los procesos de almacenamiento y tratamiento del inventario. La optimización de inventarios abarca múltiples facetas, desde la previsión de la demanda y la gestión de proveedores, hasta la logística y la implementación de estrategias de mejora continua. Una gestión eficaz del inventario puede determinar la diferencia entre el éxito y el fracaso de una empresa, influyendo directamente en la eficiencia operativa y la rentabilidad.

El entorno empresarial es dinámico y competitivo, por lo que se requiere contar con herramientas que ayuden a las empresas a mejorar sus rendimientos financieros en base a la minimización de costos operativos. Una buena gestión de inventario determina la eficiencia del almacenamiento de los productos a través de la rotación del inventario, mostrando la forma en que como se mueven los productos y sirve para tomar decisiones sobre los productos con baja rotación y mejorar los costos operativos por almacenamiento.

En este trabajo de investigación se evalúa como optimizar la gestión de inventario en empresas de consumo masivo, utilizando herramientas que ayuden a mejorar la eficiencia y rentabilidad del inventario mediante técnicas de análisis de datos y modelado estadístico. En el trabajo se explora diversas formas de ayuda a las organizaciones a minimizar costos y maximizar la eficiencia operativa, analizando modelos básicos y avanzados para determinar mejoras en la gestión de inventario y poderlos aplicar en las empresas de consumo masivo.

El estudio se centra en realizar un análisis de segmentación para obtener los productos más significativos de la empresa, diseñar un modelo predictivo para determinar el egreso

de productos mediante técnicas de aprendizaje automático y estimar la existencia mínima y máxima de productos mediante técnicas de optimización para garantizar las necesidades de la empresa. La investigación utiliza un conjunto de datos de ventas históricas, el mismo que fue depurado mediante técnicas de preprocesamiento y transformación de los datos para aplicar modelos predictivos orientados a la gestión de inventario.

Con este trabajo de investigación se espera que las empresas puedan adoptar los resultados y utilizarlos como una herramienta estratégica para optimizar sus inventarios permitiéndoles operar de manera más eficiente, satisfacer mejor a sus clientes y mejorar su competitividad y rentabilidad a largo plazo.

EMILIO RODRIGO ZHUMA MERA

Firmado digitalmente por EMILIO RODRIGO ZHUMA MERA

Ing. Emilio Rodrigo Zhuma Mera, M. Sc.

Coordinador de Carrera de Software UTEQ

RESUMEN

En el contexto de la economía global, la gestión de inventarios y la cadena de suministro son esenciales para la competitividad y rentabilidad de las empresas, especialmente en el sector de productos de consumo masivo. Este trabajo se enfoca en la optimización del inventario mediante el uso de algoritmos de aprendizaje automático, buscando equilibrar los niveles de stock con la demanda y minimizar los costos de almacenamiento.

El objetivo principal de la investigación es determinar los niveles óptimos de inventario aplicando algoritmos predictivos para minimizar los costos asociados al almacenamiento y asegurar la disponibilidad adecuada de productos. Los objetivos específicos incluyen realizar un análisis de segmentación para identificar los productos más significativos, diseñar un modelo predictivo para la demanda de productos y estimar los niveles mínimos y máximos de inventario.

Los resultados obtenidos a través de métodos como K-means y Agglomerative Clustering revelaron una segmentación robusta de los productos en cuatro clusters, lo que facilitó una gestión de inventario más precisa. Además, el análisis de Pareto ayudó a identificar los productos que generan la mayor parte de las ventas y, por ende, requieren una gestión más cuidadosa.

En el análisis predictivo, se aplicaron modelos ARIMA, Prophet, Exponential Smoothing y Random Forest. Aunque ARIMA mostró un buen desempeño en uno de los productos, el modelo Exponential Smoothing se destacó como el más eficaz en general, proporcionando las predicciones más precisas para la mayoría de los productos

analizados. Finalmente, la predicción precisa de los niveles mínimos y máximos de inventario mediante Random Forest demostró ser la más efectiva, mejorando la eficiencia operativa y contribuyendo a la rentabilidad de la empresa.

Esta investigación no solo aporta al conocimiento teórico en la gestión de inventarios, sino que también proporciona herramientas prácticas para que las empresas de consumo masivo optimicen sus operaciones en un mercado dinámico y competitivo.

Palabras clave: Gestión de inventarios, Optimización, Aprendizaje Automático, Predicción de demanda, Ciclo de vida de productos.

ABSTRACT

In the global economy, inventory management and supply chain are crucial for the competitiveness and profitability of companies, especially in the fast-moving consumer goods sector. This study focuses on inventory optimization using machine learning algorithms, aiming to balance stock levels with demand and minimize storage costs.

The main objective of the research is to determine optimal inventory levels by applying predictive algorithms to minimize storage costs and ensure adequate product availability. Specific objectives include conducting segmentation analysis to identify the most significant products, designing a predictive model for product demand, and estimating minimum and maximum inventory levels.

Results obtained through methods such as K-means and Agglomerative Clustering revealed a robust segmentation of products into four clusters, facilitating more precise inventory management. Additionally, Pareto analysis helped identify the products that generate most of the sales and thus require more careful management.

In the predictive analysis, ARIMA, Prophet, Exponential Smoothing, and Random Forest models were applied. Although ARIMA showed good performance for one of the products, the Exponential Smoothing model stood out as the most effective overall, providing the most accurate predictions for the majority of the analyzed products. Finally, the precise prediction of minimum and maximum inventory levels using Random Forest proved to be the most effective, enhancing operational efficiency and contributing significantly to the company's profitability.

This research not only contributes to theoretical knowledge in inventory management but also provides practical tools for fast-moving consumer goods companies to optimize their operations in a dynamic and competitive market.

Keywords: Inventory management, Optimization, Machine Learning, Demand prediction, Product life cycle.

ÍNDICE

ÍNDICE DE ILUSTRACIONES

ÍNDICE DE TABLAS

INTRODUCCIÓN

En la economía global, la gestión de inventario y cadena de suministro es un pilar fundamental para la competitividad y rentabilidad de las empresas, especialmente en el sector de ventas de productos de consumo masivo. En este entorno altamente dinámico y competitivo, encontrar el equilibrio adecuado entre mantener inventarios suficientes para satisfacer la demanda del mercado y minimizar los costos asociados, es un desafío estratégico de primera magnitud.

La optimización de inventario se presenta como el proceso de armonizar los niveles de stock con las fluctuaciones en la demanda y los costos de almacenamiento, emerge como un componente esencial en la cadena de suministro de las organizaciones. Este equilibrio estratégico no solo impacta en la rentabilidad, sino también en la capacidad de respuesta ante cambios en el mercado y la satisfacción del cliente.

En este contexto, el aprendizaje automático (Machine Learning) emerge como una herramienta poderosa para mejorar la gestión de inventario y la toma de decisiones en la cadena de suministro. La capacidad de los algoritmos de aprendizaje automático brinda nuevas oportunidades para la optimización de inventario y la predicción de la demanda, apoyados en el análisis de grandes volúmenes de datos y el descubrimiento de patrones complejos.

En este trabajo se abordó la optimización de inventario en empresas de consumo masivo, explorando el papel del aprendizaje automático como una herramienta innovadora para mejorar la eficiencia y rentabilidad en este contexto. Para ello, se combinaron métodos de investigación descriptiva y predictiva, así como técnicas de análisis de datos y modelado estadístico.

La presente investigación tiene como objetivo contribuir al conocimiento teórico en el campo de la gestión de inventario, también proporciona a las empresas del sector de consumo masivo herramientas y enfoques prácticos para optimizar sus operaciones y mantenerse competitivas en un mercado en constante evolución.

Este proyecto de investigación se organiza en cinco capítulos esenciales. En el primer capítulo, se hace énfasis en las prácticas de gestión de inventario en el contexto empresarial en Ecuador. El segundo capítulo se dedica a revisar la literatura relevante sobre gestión de inventario, predicción de la demanda y la aplicación de modelos predictivos en el ámbito empresarial de consumo masivo. El tercer capítulo detalla la metodología utilizada en la investigación, abordando los métodos de recopilación y análisis de datos específicos para mejorar la gestión de inventario. En el cuarto capítulo se presentan los resultados obtenidos mediante la aplicación de modelos predictivos en la gestión de inventario, junto con su análisis y discusión. Finalmente, el quinto capítulo contiene las conclusiones derivadas de los resultados y ofrece recomendaciones prácticas para las empresas interesadas en implementar estrategias de gestión de inventario basadas en modelos predictivos.

CAPÍTULO I

MARCO CONTEXTUAL DE LA INVESTIGACIÓN

1.1 UBICACIÓN Y CONTEXTUALIZACIÓN DE LA PROBLEMÁTICA

La empresa seleccionada para esta investigación se destaca a nivel nacional en la producción y distribución de productos de consumo masivo. Esta empresa ha logrado un reconocimiento significativo en la industria alimentaria gracias a la calidad de sus productos y su capacidad para satisfacer las demandas de un mercado diversificado en todo el país. Sin embargo, se plantea un desafío crítico en la eficiente gestión de inventario y la coordinación de la cadena de suministro en esta organización.

El departamento de Logística y Gestión de Inventario juega un papel esencial en la gestión de activos de la empresa. Este departamento, encargado de supervisar el flujo de productos, desde la producción hasta el almacenamiento y la distribución, enfrenta una serie de problemas complejos. Entre ellos se encuentra la identificación de productos clave, predicción de la demanda, y la determinación de niveles de inventario óptimos.

La optimización de la gestión de inventario en la cadena de suministro se configura como una necesidad apremiante para incrementar la competitividad y eficiencia de la empresa en un mercado constantemente cambiante. Una gestión incorrecta a nivel de inventario genera impactos directos en la rentabilidad y la competitividad de la empresa. Dentro de este contexto es común observar desequilibrios entre los volúmenes de producción y las capacidades de almacenamiento, lo que se traduce en costos de almacenamiento elevados, pérdida de ventas debido a la falta de disponibilidad de productos y posibles sanciones regulatorias.

1.2 SITUACIÓN ACTUAL DE LA PROBLEMÁTICA

En una empresa de consumo masivo, la gestión eficiente del inventario es crucial para maximizar las ventas y minimizar las pérdidas por productos caducados. Sin embargo, una gran diversidad de productos a comercializar puede complicar la identificación de aquellos que tienen un impacto significativo en las operaciones y finanzas de la empresa. Sin una metodología adecuada para identificar y priorizar los productos clave, se corre el riesgo de incurrir en sobrecostos y pérdidas innecesarias. La falta de un enfoque sistemático para determinar los productos más relevantes impide la implementación de estrategias efectivas de gestión del inventario.

Otro problema crítico que afectan directamente la gestión de inventario hace referencia a dificultad significativa de anticipar la demanda futura de productos, lo que impacta en la planificación de inventarios. Esta incertidumbre se origina en la complejidad inherente a la variabilidad de los patrones de consumo y la influencia de múltiples factores en la demanda, manifestándose en situaciones de exceso o escasez de inventario, lo que afecta directamente los costos operativos y la satisfacción del cliente.

De forma similar, la empresa no ha logrado determinar los niveles óptimos de inventario para sus productos perecederos, esto ha dado lugar a costos innecesarios de almacenamiento por mantener excesivo inventario, problemas de disponibilidad que pueden resultar en pérdida de ventas, deterioro de productos debido al almacenamiento prolongado, y dificultades en la planificación de la producción. Este problema impacta en los costos operativos, lo que motiva la necesidad de desarrollar estrategias efectivas de gestión de inventario para equilibrar disponibilidad y minimización de costos.

1.3 PROBLEMA DE INVESTIGACIÓN

1.3.1 Problema General

¿Cómo se podría optimizar el Inventario en la Cadena de Suministro en empresas de consumo masivo?

1.3.2 Problemas Derivados

- ¿De qué forma se podría identificar los productos más representativos de la empresa?
- ¿Cómo se podría pronosticar de manera más confiable la cantidad de producto egresada en la bodega de la empresa?
- ¿De qué manera se puede determinar el nivel óptimo de inventario para minimizar los costos asociados al almacenamiento y asegurar la disponibilidad adecuada de productos?

1.4 DELIMITACIÓN DEL PROBLEMA

CAMPO: Tecnología de la información y la comunicación

ÁREA: Tecnología de la información y la comunicación

LÍNEA: Ciencia de Datos

LUGAR: Quevedo

TIEMPO: De septiembre 2023 a enero 2024

1.5 OBJETIVOS

1.5.1 Objetivo General

Determinar los niveles óptimos de inventario mediante la aplicación algoritmos predictivos para minimizar los costos asociados al almacenamiento y asegurar la disponibilidad adecuada de productos.

1.5.2 Objetivos Específicos

- Realizar un análisis de segmentación para obtener los productos más significativos de la empresa.
- Diseñar un modelo predictivo para determinar el egreso de productos mediante técnicas de aprendizaje automático.
- Estimar la existencia mínima y máxima de productos mediante técnicas de optimización para garantizar las necesidades de la empresa.

1.6 JUSTIFICACIÓN

La gestión eficiente del inventario en la optimización de la cadena de suministro es un elemento fundamental para el éxito de las empresas de consumo masivo en un mercado altamente competitivo y dinámico. Estos aspectos no solo inciden en la rentabilidad y competitividad, sino que también impactan en la satisfacción del cliente y la sostenibilidad empresarial. La aplicación de tecnologías avanzadas, algoritmos predictivos y estrategias de gestión se ha convertido en una necesidad imperante para abordar estos desafíos y mejorar la eficiencia operativa en toda la cadena de suministro.

La correcta identificación y gestión de los productos clave en una empresa de consumo masivo es fundamental para optimizar los recursos y mejorar la eficiencia operativa. Al aplicar la ley de Pareto y técnicas de machine learning, se puede focalizar el esfuerzo en el grupo de productos que realmente impactan en la rentabilidad y operatividad del negocio.

La determinación de niveles óptimos de inventario mediante algoritmos predictivos basados en aprendizaje automático es esencial para una gestión precisa de recursos. Al minimizar costos de almacenamiento innecesarios y asegurar la disponibilidad de productos, se mejora significativamente la eficiencia de la cadena de suministro, generando beneficios económicos y competitivos.

La aplicación de técnicas de optimización en la estimación de niveles de inventario mínimos y máximos es esencial para garantizar una gestión eficiente de recursos. Esto permite a la empresa mantener un equilibrio preciso entre la disponibilidad de productos y la minimización de costos, optimizando así la cadena de suministro.

CAPÍTULO II

MARCO TEÓRICO DE LA INVESTIGACIÓN

2.1 FUNDAMENTACIÓN CONCEPTUAL

2.1.1 Optimización de inventario

La optimización de inventario es un proceso crucial en la gestión de cadenas de suministro, especialmente en el contexto de la economía global actual. Consiste en encontrar el equilibrio adecuado para mantener un inventario suficiente que satisfaga la demanda de los clientes, al tiempo que se reducen los costos asociados con el almacenamiento y mantenimiento de productos. Este equilibrio es fundamental, dado que un exceso de inventario puede llevar a costos innecesarios y riesgos de obsolescencia, mientras que un inventario insuficiente puede resultar en la pérdida de ventas y clientes insatisfechos. Una optimización efectiva del inventario implica considerar factores como la variabilidad de la demanda, plazos de entrega, costos de almacenamiento y producción, y la capacidad de respuesta de la cadena de suministro. Además, requiere una coordinación cuidadosa entre todos los actores involucrados en la cadena de suministro, ya que cada uno puede tener objetivos y metas individuales que deben estar alineados con la estrategia global de inventario (Vandeput, 2020).

2.1.2 Ciclo de vida de productos

De acuerdo con López Morales (2021), el ciclo de vida del producto es un modelo ampliamente aceptado en la gestión y mercadotecnia de productos que describe la trayectoria que un producto sigue a lo largo de su existencia en una industria. Este modelo divide la vida útil de un producto en cuatro etapas fundamentales, que constituyen un marco conceptual esencial para comprender la dinámica de los productos en el mercado. En la primera etapa, conocida como la "Introducción", el producto se crea, fabrica y se introduce en el mercado.

La segunda etapa, denominada "Crecimiento", se caracteriza por un aumento significativo en las ventas debido a la mayor penetración en el mercado y a las mejoras continuas en el producto. La tercera etapa, llamada "Madurez", representa el punto en el que las ventas alcanzan su punto máximo, y el costo de producción tiende a reducirse gracias a la escalabilidad y la eficiencia en la fabricación.

Finalmente, la cuarta etapa, conocida como el "Declive", señala el período en el que las ventas comienzan a disminuir, a menudo debido a la introducción de nuevos productos en el mercado o a la obsolescencia del producto original. Este concepto proporciona una base sólida para comprender cómo los productos evolucionan en el mercado y cómo las estrategias de gestión deben adaptarse en cada etapa para maximizar el éxito y la rentabilidad del producto.

2.1.3 Series de tiempo

Como mencionan Torres et al. (2021), una serie de tiempo se describe como una sucesión de valores dispuestos cronológicamente y observados a lo largo del tiempo. Aunque el tiempo es una variable que se mide de manera continua, los valores de una serie temporal se registran a intervalos regulares (frecuencia de muestreo fija). Las series temporales se distinguen por tres componentes principales: tendencia, estacionalidad y componentes irregulares, también llamados residuales. Las series temporales pueden visualizarse gráficamente. Específicamente, el eje x representa el tiempo, mientras que el eje y muestra los valores registrados en momentos específicos. Esta representación gráfica facilita la identificación visual de las características más relevantes de la serie, como la amplitud de las oscilaciones, las estaciones y ciclos presentes, o la presencia de datos anómalos o valores atípicos.

2.1.4 Aprendizaje Automático

El Aprendizaje Automático (ML), como mencionan Choi et al. (Choi et al., 2020), se rige como un pilar fundamental en la inteligencia artificial (IA), focalizándose en la capacidad de la IA para aprender y mejorar su rendimiento a partir de datos. En contraste con la programación tradicional, donde los algoritmos se construyen de manera explícita utilizando reglas y características conocidas, el ML adopta un enfoque más dinámico y adaptable. Utiliza subconjuntos de datos para iterativamente desarrollar algoritmos que tienen la capacidad de explorar combinaciones innovadoras y ponderaciones variables de características, permitiendo así que la IA derive su conocimiento desde principios fundamentales y experiencias previas.

El ML abarca un abanico diverso de métodos, siendo cuatro de los más comunes: el aprendizaje supervisado, que implica entrenar modelos con datos etiquetados; el no supervisado, que busca patrones y estructuras en datos no etiquetados; el semisupervisado, que combina elementos de ambos; y el aprendizaje por refuerzo, que se basa en la interacción con el entorno para aprender a tomar decisiones óptimas. Estos métodos se han convertido en pilares esenciales en la resolución de problemas en una variedad de dominios, y su aplicación se extiende desde la visión por computadora hasta la toma de decisiones autónomas y la mejora de procesos en numerosos campos.

2.1.5 Modelo autorregresivo integrado de media móvil (ARIMA)

Los modelos ARIMA (que abarcan los modelos ARMA, AR y MA) constituyen una clase general de modelos utilizados para pronosticar series temporales estacionarias (Yonar et al., 2020). Estos modelos se componen de tres partes principales:

- Una suma ponderada de valores rezagados de la serie (componente autorregresivo, AR).

- Una suma ponderada de errores de predicción rezagados de la serie (componente de media móvil, MA).
- Una diferencia de la serie temporal (componente integrado, I).

Un modelo ARIMA se designa comúnmente como ARIMA (p, d, q), donde p indica el orden de la parte AR, d representa el orden de diferenciación (componente "I") y q denota el orden del término MA.

2.1.6 Exponential smoothing

El método de suavizado exponencial emplea un promedio ponderado especial para suavizar las muestras de datos de series temporales. Este método descompone la serie temporal en tres componentes: la media general, la tendencia general y la tendencia estacional. Asigna ponderaciones de manera particular, de modo que cuanto más alejado esté un punto temporal en los datos históricos respecto al punto temporal actual, menos peso se le asigna a su valor real. A los valores anteriores al punto temporal actual se les otorga un peso que disminuye exponencialmente con la distancia, convergiendo gradualmente a cero. Esto asegura no solo la integridad de la información de la serie temporal, sino que también enfoca la atención en la información relevante en diferentes momentos del tiempo (Xie et al., 2022).

2.1.7 Random Forest

El bosque aleatorio es un método de aprendizaje automático ampliamente utilizado para crear modelos de predicción. Introducidos por Breiman (2001), los bosques aleatorios consisten en una colección de árboles de clasificación y regresión, los cuales son modelos simples que emplean divisiones binarias en variables predictoras para realizar predicciones de resultados. Los árboles de decisión son prácticos y ofrecen una forma intuitiva de predecir resultados al dividir los valores "altos" y "bajos" de un predictor

relacionado con el resultado. Aunque tienen muchas ventajas, los árboles de decisión a menudo muestran baja precisión cuando se aplican a conjuntos de datos complejos, como aquellos con gran tamaño o con interacciones complejas entre variables (Speiser et al., 2019).

2.1.8 Prophet

PROPHET es una herramienta para pronosticar datos de series temporales, desarrollada por el equipo de Core Data Science de Facebook. Su objetivo es facilitar la realización de pronósticos "a escala" mediante un enfoque automatizado que simplifica el ajuste de métodos de series de tiempo. Prophet está diseñado para ser accesible, permitiendo que analistas con distintos niveles de experiencia o incluso con conocimientos limitados en previsión puedan realizar pronósticos de manera efectiva. Para utilizar Prophet, se debe preparar el conjunto de datos con dos columnas: 'ds' (fecha) y 'y' (medida de pronóstico). Luego, se crea un objeto a partir de la clase Prophet(), se seleccionan los períodos a pronosticar y el resultado generado tendrá varias columnas, siendo las principales 'ds' y 'yhat'. La columna 'yhat' contiene los valores pronosticados de 'y' en el marco de datos históricos. Las columnas 'ds' y 'yhat' se pueden graficar para mostrar características como tendencias futuras o estacionalidad (Aditya Satrio et al., 2021).

2.1.9 K Means-Clustering

Esta metodología de agrupamiento es un tipo de aprendizaje no supervisado que se aplica en datos sin etiquetar. El objetivo de este algoritmo es identificar ciertos grupos en los datos, y K es una variable que representa el número de grupos. El algoritmo funciona de manera iterativa y asigna cada punto de datos a uno de los K grupos. Según los datos de similitud de características, los puntos se agrupan (Paul et al., 2020).

El algoritmo mencionado anteriormente se encuentra implementado en librería Open CV y se basa en aumentar k-means con una técnica de semilla aleatoria muy simple, se obtiene un algoritmo que es Θ (log k)- competitivo con la agrupación óptima. Los experimentos preliminares muestran que nuestro aumento mejora tanto la velocidad como la precisión de kmeans, a menudo bastante (Arthur & Vassilvitskii, 2007).

2.1.10 Criterio de información de Akaike (AIC)

El criterio de información de Akaike es un método matemático utilizado para evaluar la calidad del ajuste de un modelo respecto a los datos de los cuales se generó. Esta métrica es especialmente útil para evaluar modelos ARIMA, ya que ayuda a equilibrar el ajuste del modelo con su complejidad. Al penalizar el número de parámetros, el AIC facilita la selección de modelos que son tanto precisos como simples. Su aplicación en diversos campos, desde la predicción de la calidad del aire hasta la modelización epidemiológica, destaca su versatilidad y eficacia (Zhang & Meng, 2023).

2.2 FUNDAMENTACIÓN TEÓRICA

En el artículo de (Vijesh Chaudhary et al., 2023), se explora minuciosamente el uso del aprendizaje automático (ML) en la gestión de inventario con el objetivo de aumentar la rentabilidad. Comienza destacando la importancia crítica de la gestión de inventarios para la rentabilidad de una empresa y proporciona una breve explicación del papel del aprendizaje automático en la optimización de inventarios. Posteriormente, profundiza la aplicación de algoritmos de Regresión lineal, análisis de series de tiempo, Random forest, ANN, SVM y aprendizaje por refuerzo en la gestión de inventarios y realiza la comparación de su eficacia en comparación con los métodos tradicionales. Se subrayan los beneficios clave del uso del aprendizaje automático en la gestión de inventarios,

incluida la mejora en la precisión de las predicciones de demanda y la optimización más efectiva del inventario.

En el marco de la investigación sobre pronóstico de demanda en la cadena de suministro extendida, se ha identificado un artículo que examina la efectividad de técnicas avanzadas de aprendizaje automático no lineales. En (Carbonneau et al., 2008), se llevó a cabo una comparación exhaustiva de diversos modelos de pronóstico, incluyendo redes neuronales recurrentes, máquinas de vectores de soporte, redes neuronales y regresión lineal múltiple, junto con técnicas más simples como promedios móviles, métodos ingenuos y de tendencia. Los resultados obtenidos en el estudio indican que, aunque algunos métodos demostraron un mejor rendimiento que otros, no existe una diferencia estadísticamente significativa en cuanto a precisión entre el método de mejor rendimiento y el modelo de regresión. Estos hallazgos sugieren que mejorar la precisión en el pronóstico puede resultar en costos más bajos y una mayor satisfacción del cliente debido a entregas más puntuales, especialmente en situaciones donde las partes en la cadena de suministro no pueden colaborar estrechamente. A pesar de que los autores no mencionaron explícitamente el uso de redes LSTM multicapa en su método propuesto, sí utilizaron redes neuronales recurrentes (RNN) para el pronóstico de la demanda, las cuales pueden incluir LSTMs como un tipo de RNN. Además, destacaron que las RNN mostraron un mejor rendimiento que otros métodos, presumiblemente debido a su capacidad para capturar patrones temporales.

En un estudio adicional, se recopilan diversas aplicaciones comerciales de técnicas de aprendizaje automático (ML) en la gestión de la cadena de suministro. Este análisis exhaustivo de casos de estudio reales destaca cómo las organizaciones están aplicando con éxito técnicas de aprendizaje automático para mejorar la eficiencia y la toma de

decisiones en sus cadenas de suministro. Los casos de estudio abarcan desde la predicción de la demanda hasta la planificación de la producción y la optimización del transporte, demostrando el amplio espectro de aplicaciones que ofrece el aprendizaje automático en este contexto (Carbonneau et al., 2008).

En (Wan & Hong, 2022) se presenta un enfoque novedoso que combina herramientas computacionales eficientes de redes neuronales recurrentes (RNN) con la información estructural de las redes de producción. Este enfoque de simulación inspirado en RNN se ha destacado por su velocidad excepcional, siendo miles de veces más rápido que los enfoques tradicionales. Además de su velocidad, el enfoque también ha demostrado su capacidad para resolver problemas de optimización de inventario a gran escala en un período de tiempo razonable, lo que lo convierte en una contribución valiosa para la gestión de inventario en entornos de alta complejidad.

Además, en el estudio (Varkalys, 2021) se utilizaron algoritmos de aprendizaje automático para predecir los niveles mínimos y máximos de stock en una cadena de suministro. Se llevaron a cabo pruebas exhaustivas con diferentes algoritmos, como regresión de vectores de soporte, k-vecinos más cercanos, bosque aleatorio, red neuronal artificial, ARIMA y Prophet. Los resultados revelaron que el algoritmo de k vecinos más cercanos logró la mejor precisión en la predicción del nivel mínimo de existencias, con una medida sMAPE promedio del 7.0079%. Por otro lado, el algoritmo Random Forest destacó en la predicción del nivel máximo de existencias, con un sMAPE promedio del 15.0303%.

En la investigación de Abbasimehr et al. (2020), se propone un método de pronóstico de la demanda basado en redes LSTM multicapa. El método propuesto selecciona automáticamente el mejor modelo de pronóstico considerando diferentes combinaciones de hiperparámetros LSTM para una serie de tiempo determinada utilizando el método de

búsqueda de cuadrícula. El método propuesto se compara con algunas técnicas conocidas de predicción de series de tiempo a partir de métodos estadísticos y de inteligencia computacional que utilizan datos de demanda de una empresa de muebles. Estos métodos incluyen ARIMA, exponential smoothing, red neuronal artificial (ANN), red neuronal recurrente (RNN), KNN, SVM y LSTM de una sola capa. Los resultados experimentales indican que el método propuesto es superior entre los métodos probados en términos de medidas de rendimiento.

Por último, en (Kiuchi et al., 2020) se propone un enfoque de optimización bayesiano junto con un simulador de cadena de suministro basado en agentes para abordar un problema de optimización de inventario multiescalón restringido. Este enfoque se compara con el algoritmo genético (GA), ampliamente utilizado en la optimización de inventario. Los resultados experimentales demuestran que el método propuesto converge a la solución óptima significativamente más rápido que GA, lo que sugiere su potencial para mejorar la eficiencia en la gestión de inventarios multiescalón.

En términos generales, la demanda de los clientes se representa como datos secuenciales que reflejan las demandas de los clientes a lo largo del tiempo. Por consiguiente, el problema de prever la demanda puede plantearse como un problema de predicción de series temporales (Villegas et al., 2018). Por ello se utilizó los modelos ARIMA, ExponentialSmoothing, Prophet y Random forest.

2.3 FUNDAMENTACIÓN LEGAL

La Ley de Propiedad Intelectual de Ecuador, en su artículo 130, reconoce y protege los derechos de propiedad intelectual de las empresas, incluyendo la confidencialidad de sus datos comerciales y técnicos. Este marco legal establece que las empresas tienen el derecho exclusivo de controlar el acceso y el uso de sus datos y conocimientos no divulgados. Esto abarca información como fórmulas comerciales, procesos de producción, estrategias de negocio, y cualquier otro dato confidencial que pueda conferir una ventaja competitiva.

En consecuencia, la empresa que colabora en esta investigación está amparada por esta ley en su decisión de salvaguardar la confidencialidad de la información que comparte, asegurando así la protección de su propiedad intelectual y su posición en el mercado. La confidencialidad, respaldada por la legislación vigente, se erige como un pilar fundamental para la colaboración entre empresas y la investigación con datos sensibles, garantizando el respeto a los derechos de propiedad intelectual y la integridad de la información empresarial.

Además, la Ley Orgánica de Datos Personales, aprobada en Ecuador en 2021, establece disposiciones específicas para la protección de los datos personales y, por extensión, de la información empresarial confidencial. Esta ley impone obligaciones a las empresas que manejan datos personales y refuerza la necesidad de implementar medidas de seguridad adecuadas para evitar el acceso no autorizado a esta información. En este sentido, la empresa que proporcionó los datos para esta investigación se encuentra en pleno cumplimiento de estas disposiciones legales, asegurando así la confidencialidad de la información suministrada, lo que garantiza la protección de sus intereses comerciales y el respeto a los derechos fundamentales de privacidad.

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1 TIPO DE INVESTIGACIÓN

En el presente estudio se emplearon dos tipos de investigación con el propósito de abordar de manera completa la problemática planteada en la gestión de inventario y la cadena de suministro de empresas de consumo masivo.

3.1.1 Investigación Descriptiva

La investigación descriptiva se utilizó como punto de partida para obtener una comprensión detallada de la gestión actual de inventario. Esta modalidad se centra en recopilar y analizar datos históricos, así como en identificar patrones y tendencias en el comportamiento de la demanda y la administración de inventarios.

3.1.2 Investigación Predictiva

Este enfoque de investigación se empleó para desarrollar modelos y algoritmos predictivos que permitan anticipar el consumo futuro de productos, así como determinar niveles óptimos de inventario. Además, contribuyó a establecer estrategias más eficaces para la gestión de inventarios y la cadena de suministro mediante la aplicación de técnicas de aprendizaje automático basado en datos históricos.

3.2 MÉTODOS UTILIZADOS EN LA INVESTIGACIÓN

Se aplicó una variedad de métodos de investigación con el propósito de alcanzar los objetivos específicos establecidos. A continuación, se detallan cada uno de ellos.

3.2.1 Método de análisis

El método analítico fue fundamental para comprender en profundidad la situación actual de la gestión de inventario en la cadena de suministro en la empresa de consumo masivo. A través de este método, se examinaron datos históricos, informes de inventario, y otros documentos clave para identificar patrones, tendencias y posibles áreas de mejora.

3.2.2 Método de Síntesis

Una vez realizado el análisis detallado, fue necesario sintetizar la información recopilada para obtener una visión global de la situación actual de la empresa. Este método permitió combinar e integrar los hallazgos para identificar patrones generales y tendencias relevantes para la toma de decisiones.

3.2.3 Método Deductivo

El método deductivo se utilizó para desarrollar y validar modelos predictivos basados en los datos recopilados. Partiendo de la teoría y conceptos establecidos en la literatura y la experiencia práctica, se generaron los modelos utilizando técnicas de aprendizaje automático para optimizar la gestión de inventario.

3.3 CONSTRUCCIÓN METODOLÓGICA DEL OBJETO DE INVESTIGACIÓN

3.3.1 Técnicas de investigación

Se utilizaron diversas técnicas de investigación que permitieron explorar, analizar y extraer conocimiento valioso de conjuntos de datos. A continuación, se detallan las técnicas utilizadas:

- **Estadística Descriptiva:** Se utilizaron medidas estadísticas como la media, mediana, desviación estándar, junto con gráficos como histogramas y diagramas de dispersión para describir y resumir los datos.
- **Análisis Exploratorio de Datos (EDA):** Se aplicó este enfoque para analizar el conjunto de datos y resumir sus características principales. Esto se logró mediante métodos visuales e incluyó la identificación de patrones, valores atípicos y relaciones entre variables.

- **Aprendizaje Automático (Machine Learning):** Se utilizaron algoritmos y modelos de aprendizaje automático para predecir o clasificar datos nuevos. Esto incluirá técnicas como regresión, clasificación, agrupación y redes neuronales.
- **Análisis Predictivo y Modelado Estadístico:** Se utilizaron modelos estadísticos y matemáticos para prever eventos o tendencias futuras basándose en datos históricos.
- **Visualización de Datos:** Se emplearon gráficos y representaciones visuales para comunicar patrones y tendencias en los datos, facilitando la comprensión de los resultados del análisis.
- **Optimización:** Se utilizaron técnicas para encontrar la mejor solución a un problema, como la optimización de recursos limitados para maximizar beneficios.

3.3.2 Instrumentos de la Investigación

- **Python:** Lenguaje de programación ampliamente reconocido en ciencia de datos. Incorpora bibliotecas como Pandas, que es especialmente útil para la manipulación de datos, NumPy para cálculos numéricos, y scikit-learn para el aprendizaje automático.
- **TensorFlow:** Librería de Python esencial para el desarrollo de modelos de aprendizaje profundo. TensorFlow es una potente biblioteca de código abierto que facilita la creación y entrenamiento de modelos de aprendizaje automático, incluyendo redes neuronales. Ofrece un entorno flexible y escalable que permite trabajar con conjuntos de datos de gran tamaño.
- **Keras:** Interfaz de alto nivel que se ejecuta sobre TensorFlow, simplificando la creación y entrenamiento de modelos de aprendizaje profundo. Su diseño intuitivo y su capacidad para crear modelos de manera rápida lo convierten en una

herramienta esencial para desarrolladores e investigadores en el campo del aprendizaje automático.

- **Matplotlib y Seaborn:** Bibliotecas de Python que permiten la creación de gráficos y visualizaciones de datos. Facilitan la representación gráfica de resultados y patrones.
- Microsoft SQL Server: Es esencial para la extracción y manipulación de datos almacenados en bases de datos estructuradas.
- GitHub: Plataforma esencial para el almacenamiento y gestión de código en la nube. Facilita la colaboración en proyectos y el control de versiones, lo que resulta crucial en el desarrollo de soluciones en ciencia de datos.

3.4 ELABORACIÓN DEL MARCO TEÓRICO

Para la elaboración del marco teórico, se siguió una metodología rigurosa y secuencial que aseguró la inclusión de información relevante y actualizada. El proceso se detalló de la siguiente manera:

- **Definición de Temas y Conceptos Clave:** Se identificaron los conceptos fundamentales y temas relevantes relacionados con la gestión de inventario y cadena de suministro en empresas de consumo masivo. Esto proporcionó una base sólida para la investigación.
- **Búsqueda bibliográfica:** Se realizó una búsqueda sistemática en Google Scholar para localizar artículos científicos pertinentes. Se aplicó un filtro para que los artículos seleccionados no tuvieran más de 5 años de antigüedad y pertenecieran a revistas de renombre como SCOPUS, IEEE y Springer. Este paso garantizó la inclusión de la información más actualizada y confiable disponible en la literatura académica.

- **Revisión y Selección de Artículos**: Se revisaron minuciosamente los artículos encontrados, evaluando su relevancia y contribución al tema de investigación. Se prestó especial atención a los enfoques y metodologías utilizados en estudios similares, así como a los resultados y conclusiones obtenidos.
- **Análisis y Síntesis de la Información**: Se llevó a cabo un proceso de análisis y síntesis de la información recopilada. Esto implicó la identificación de patrones, tendencias y hallazgos comunes entre los artículos seleccionados. Se destacaron los conceptos clave, modelos teóricos y enfoques metodológicos que emergieron de la literatura.
- **Creación del Marco Teórico Conceptual**: Se construyó un marco teórico conceptual que integró los conceptos clave identificados. Este marco proporcionó una estructura sólida para comprender y abordar la gestión de inventario y cadena de suministro en empresas de consumo masivo.
- **Exploración de Aspectos Legales:** Además de la literatura académica, se investigaron aspectos legales relacionados con el tratamiento de datos personales.

3.5 RECOLECCIÓN DE LA INFORMACIÓN

Este proceso de recolección de información se llevó a cabo con el objetivo de obtener datos primarios relevantes y confiables, necesarios para el análisis y desarrollo de la investigación sobre la gestión de inventario en empresas de consumo masivo.

La **Tabla 1** contiene las variables utilizadas, este conjunto de datos está construido por los registros de movimientos de inventario y comprenden información detallada sobre productos, ubicaciones de almacenamiento, fechas de transacción, cantidades involucradas, valores monetarios y otros atributos relevantes. En total, el número de registros es de aproximadamente 10 millones de registros.

Tabla 1

Diccionario de datos de los movimientos de inventario

Nombre del Campo	**Descripción**	**Escala de Medición**
CodProducto	Código de Stock	Categórica (Nominal)
Bodega	Almacén	Categórica (Nominal)
Anio	Año de Transacción	Numérica (Decimal)
Mes	Mes de Transacción	Numérica (Decimal)
Fecha	Fecha de Entrada	Numérica (Fecha)
Tiempo	Hora de Transacción	Numérica (Decimal)
TipoMov	Tipo de Movimiento	Categórica (Nominal)
Cantidad	Cantidad de Transacción	Numérica (Decimal)
Valor	Valor de Transacción	Numérica (Decimal)
Fuente	Fuente	Categórica (Nominal)
Referencia	Referencia	Categórica (Nominal)
CostoUnit	Costo Unitario	Numérica (Decimal)
Proveedor	Proveedor	Categórica (Nominal)
Factura	Factura	Categórica (Nominal)
TipoDoc	Tipo de Documento	Categórica (Nominal)
CodCliente	Cliente	Categórica (Nominal)
ValorCosto	Valor de Costo	Numérica (Decimal)
Sucursal	Sucursal	Categórica (Nominal)
Vendedor	Vendedor	Categórica (Nominal)
ClaseProducto	Clase de Producto	Categórica (Nominal)
Lote	Lote o Serial	Categórica (Nominal)

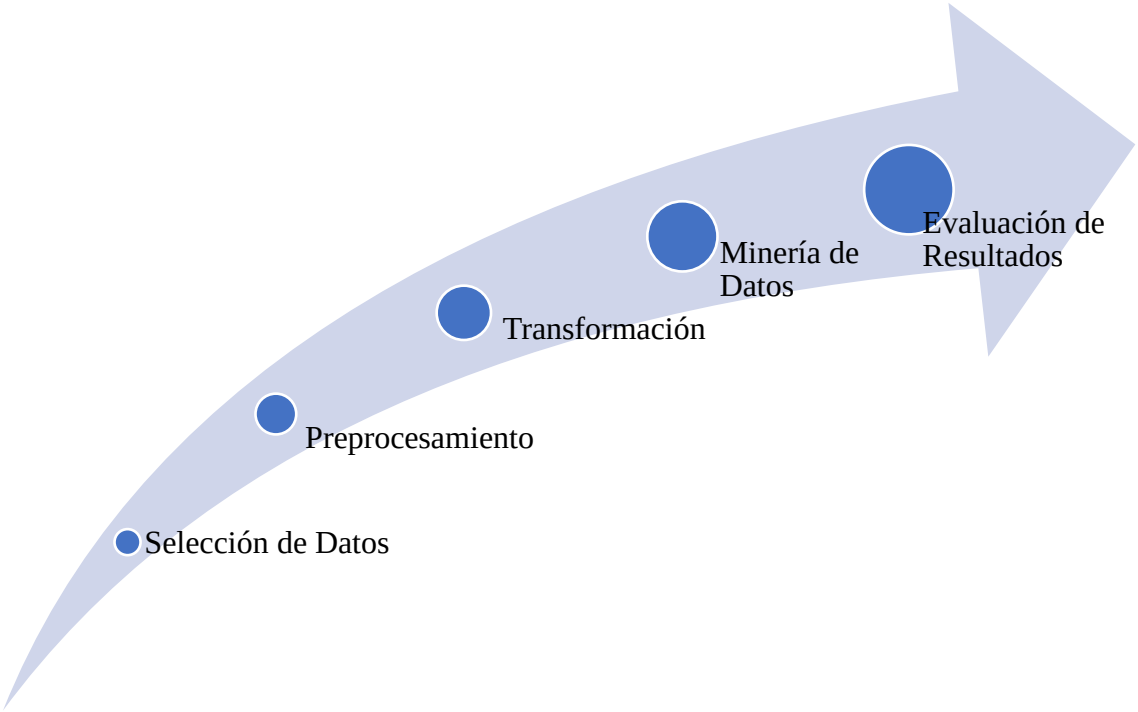
Nota. Fuente: Autor

3.6 **PROCESAMIENTO Y ANÁLISIS**

Se llevó a cabo el procesamiento y análisis de la información recopilada implementando la metodología KDD (*Knowledge Discovery in Databases*). Este enfoque sistemático proporciona una estructura robusta para descubrir conocimientos valiosos a partir de los conjuntos de datos (Plotnikova et al., 2020). En la **Ilustración 1** se puede observar los cinco pasos esenciales para aplicar esta metodología.

Ilustración 1

Pasos de la metodología KDD



Nota. Fuente: Autor

3.6.1 Selección de Datos

En este primer paso, se identificó el conjunto de datos utilizados en la investigación. Esto incluyó la elección de tablas y campos específicos de la base de datos, centrándose en información relacionada con la gestión de inventario en la cadena de suministro.

Objetivo 1: Segmentación de productos relevantes

La selección de datos se llevó a cabo mediante una consulta SQL diseñada para agrupar y extraer información relevante sobre la cantidad, el valor de venta y la utilidad de cada producto. Esta consulta resultó en un archivo CSV que contenía un registro único por producto, facilitando así el análisis posterior.

El conjunto de datos final se puede observar en la **Ilustración 2**, el cual está compuesto por 269 registros y cinco columnas: Producto, Cantidad, Venta, Utilidad y el código original del producto mismo que por motivos de confidencialidad de a excluido. Cada fila representaba un producto diferente, y las columnas proporcionaban información cuantitativa sobre la cantidad vendida, el valor total de las ventas y la utilidad generada en los años 2022 y 2023.

Ilustración 2

Conjunto de datos utilizado para la segmentación de productos

	Producto	Cantidad	Venta	Utilidad
0	Producto0	2.0	0.53	0.06
1	Producto1	1.0	0.58	0.26
2	Producto2	1.0	0.90	0.33
3	Producto3	2.0	1.97	0.62
4	Producto4	35.0	5.68	0.97

Shape: (269, 5)

Nota. Fuente: Autor

Objetivo 2: Modelo predictivo del consumo de productos

Se utilizaron los productos de la **Ilustración 10** correspondiente al top 10 de productos que generaron mayor utilidad en el año 2022 a 2023, estos resultados fueron obtenidos en el objetivo anterior. Se verificaron los registros de ventas mensuales de cada uno de los productos a través de una consulta en la base de datos de SQL Server, esto para

identificar la amplitud de la serie temporal a analizar. Los productos que tenían registros de venta en la mayor cantidad de los meses son lo que fueron considerados para el análisis. Se seleccionó el producto Producto259 debido que mostro tener ventas continuas desde enero de 2016 hasta diciembre de 2023, totalizando 96 meses en los 8 años. También se analizó los productos Producto267 y Producto266 ubicados en el top 3 de la lista, estos tienen ventas desde abril del 2017 hasta diciembre del 2023, totalizando 81 meses en los 7 años. El producto Producto268 ubicado en el top 1 de la lista no se lo consideró para el análisis debido a que presenta ventas desde diciembre del 2020 hasta diciembre del 2023, con esto solo se obtiene un total de 37 meses para análisis de serie temporal.

Objetivo 3: Niveles mínimos y máximos de inventario.

Se realizó una consulta a la base de datos para obtener el detalle de los movimientos de inventario relacionados con las ventas del producto de interés. El conjunto de datos inicial incluía las siguientes columnas: Producto, Año, Mes, Fecha, y Cantidad. Estos datos se usaron para construir el conjunto de datos que contenía los niveles mínimos y máximos del producto en cada uno de los 96 periodos. El conjunto de datos resultante está conformado con las columnas: Producto, Fecha, Cantidad, nivel Mínimo y nivel Máximo.

3.6.2 Preprocesamiento de Datos

Se aplicaron técnicas de preprocesamiento para garantizar la calidad y consistencia de los datos. Esto incluyó la limpieza de registros, la detección y manejo de valores atípicos, así como la imputación de datos faltantes.

Objetivo 1: Segmentación de productos relevantes

En primer lugar, se realizó una verificación de la existencia de valores faltantes, este proceso aseguró que no había datos ausentes en ninguna de las columnas del conjunto de datos, confirmando que todos los registros estaban completos. A continuación, se

verificaron posibles valores anómalos, específicamente valores negativos, en las columnas de Cantidad, Venta y Utilidad. Este paso fue crucial para garantizar que todos los datos eran válidos y representaban correctamente las transacciones y utilidades de los productos.

Además, se examinó la estadística descriptiva del conjunto de datos para entender mejor la distribución y la variabilidad de los datos. Este análisis reveló una alta variabilidad en las variables, con diferencias significativas entre los valores mínimos y máximos, lo que motivó la necesidad de normalizar las variables para mejorar la precisión de los modelos de segmentación.

Objetivo 2: Modelo predictivo del consumo de productos

El conjunto de datos inicial contenía las columnas Fecha, Producto, Anio, Mes y Cantidad. Como se va a realizar el análisis de series temporales de un producto, solo se consideró las columnas Fecha y Cantidad. La columna Fecha se transformó a datetime y se estableció como el índice del DataFrame. Posteriormente, se realizó una visualización de la serie temporal para observar su comportamiento y se graficó la distribución de la densidad de la serie.

Objetivo 3: Niveles mínimos y máximos de inventario.

El conjunto de datos original contenía múltiples registros de ventas para cada mes. Por lo tanto, se realizó una agrupación por año y mes para obtener la suma total de las cantidades vendidas en cada uno de los periodos. Este paso consolidó la información en un conjunto de datos mensual que contenía 96 registros, correspondientes a 8 años de ventas mensuales. No se encontraron valores faltantes en la serie temporal. Posteriormente, se añadió una columna de LeadTime que representa los días que tarda en llegar el producto desde que se crea la orden de compra.

3.6.3 Transformación de Datos

En esta etapa se llevaron a cabo transformaciones en los datos para prepararlos adecuadamente para el análisis. Este paso incluye la normalización de variables, creación de nuevas características derivadas y codificación de variables categóricas.

Objetivo 1: Segmentación de productos relevantes

Para abordar la alta variabilidad observada en las variables Cantidad, Venta y Utilidad, se llevó a cabo un proceso de normalización. Este proceso se realizó utilizando la técnica de escalado estándar, que transforma los datos para que tengan una media de cero y una desviación estándar de uno. La normalización es crucial en el análisis de datos porque garantiza que todas las variables contribuyan de manera equitativa al análisis, evitando que aquellas con rangos más amplios dominen los resultados.

Objetivo 2: Modelo predictivo del consumo de productos

Para el modelo ARIMA, se verificó si la serie era estacionaria utilizando la prueba de Dickey-Fuller aumentado (adfuller). Los resultados iniciales de la prueba indicaron que la serie no era estacionaria (p-value = 0.595928). Por lo tanto, se aplicó la primera diferencia a la serie para lograr la estacionariedad, lo cual se confirmó con una segunda prueba de adfuller que mostró un p-value de 4.280958e-10. Esta transformación fue crucial para la correcta implementación del modelo ARIMA.

Para el modelo RandomForestRegressor, se realizaron varias transformaciones adicionales para generar características relevantes. Primero, se renombraron las columnas Fecha y Cantidad a 'ds' y 'y' respectivamente, y se crearon dos nuevas columnas: 'Year' (año) y 'Mes' (mes), derivadas de la columna 'ds'. Además, se añadieron retrasos (lags) de 1 a 12 meses como nuevas columnas, que representan los valores de ventas de los meses

anteriores. Estas columnas son fundamentales para capturar la dependencia temporal en el modelo de regresión

Objetivo 3: Niveles mínimos y máximos de inventario.

Se agregaron columnas adicionales útiles para el cálculo de los niveles mínimos y máximos de inventario, a continuación, se describe cada una de ellas:

- LeadTime: Número de días de tiempo de entrega.
- L1, L2, L3: Cantidades vendidas en los tres meses anteriores.
- ServiceLevel: Nivel de servicio objetivo del 99%.
- Total: Suma de las ventas de los tres meses anteriores.
- Promedio_mensual: Promedio mensual de ventas de los tres meses anteriores.
- Promedio_diario: Promedio diario de ventas basado en el promedio mensual.
- Desviacion_estandar_ms: Desviación estándar de las ventas de los tres meses anteriores.
- LeadTime_mes: LeadTime en meses.
- ServiceLevel_Z: Valor Z correspondiente al nivel de servicio del 99% (2.33).
- Inventario_seguridad: Inventario de seguridad calculado usando la desviación estándar y el valor Z.
- PuntoReOrden: Punto de reorden calculado como el promedio diario multiplicado por el LeadTime más el inventario de seguridad.
- Minimo: Nivel mínimo de inventario calculado como el promedio diario multiplicado por el LeadTime.
- Maximo: Nivel máximo de inventario calculado como el mínimo más el promedio diario multiplicado por el LeadTime.

El resultado de estos cálculos fue almacenado en un archivo CSV para realizar el análisis de series temporales del nivel mínimo y máximo de inventario.

3.6.4 Minería de Datos

Se aplicaron algoritmos de minería de datos para descubrir patrones, tendencias y relaciones en los datos. Se usaron técnicas de agrupación, clasificación y regresión, así como análisis de series temporales.

Objetivo 1: Segmentación de productos relevantes

En esta etapa, se aplicaron técnicas avanzadas de minería de datos para realizar la segmentación de los productos de la empresa. Se utilizaron dos métodos de clustering: K-means y Agglomerative Clustering. El método K-means fue el primer enfoque utilizado para la segmentación. Se probó con diferentes números de clusters, desde uno hasta siete, para determinar el número óptimo de clusters. La evaluación se realizó mediante el análisis de la inercia promedio, que mide la suma de las distancias cuadradas entre cada punto y el centroide del cluster al que pertenece. Una menor inercia indica una mayor cohesión interna dentro de los clusters. La elección final de cuatro clusters se basó en la observación de una disminución significativa en la inercia, lo que sugiere que este número de clusters ofrece un buen equilibrio entre cohesión y separación de los datos.

El segundo método utilizado fue el Agglomerative Clustering, un enfoque jerárquico que construye un árbol de agrupamiento (dendrograma) para visualizar las distancias euclidianas entre los productos. Se empleó el método de enlace de Ward para minimizar la varianza dentro de cada cluster. Además del dendrograma, se calculó el coeficiente de silueta promedio para diferentes números de clusters. Este coeficiente mide la calidad de la agrupación al comparar la similitud de cada punto con su propio cluster y con otros clusters. La elección de cuatro clusters también se respaldó con este método, debido a que

ofreció una buena combinación de cohesión y separación, según lo indicado por el coeficiente de silueta.

Objetivo 2: Modelo predictivo del consumo de productos

Para determinar el mejor modelo ARIMA, se utilizó la librería pmdarima, que permitió identificar el modelo ARIMA(1,1,1)(0,0,1)[12] como el más adecuado para los datos. Este modelo fue entrenado utilizando los datos de enero de 2016 a diciembre de 2022, reservando los 12 meses del 2023 para pruebas.

Además de ARIMA, se implementaron otros tres modelos de predicción: Prophet, Exponential Smoothing y Random Forest Regressor. Para Prophet, se renombraron las columnas a 'ds' y 'y', y se realizó un entrenamiento similar al de ARIMA. El modelo Exponential Smoothing se configuró con tendencias y estacionalidad aditivas, y se entrenó. Finalmente, para Random Forest Regressor, se generaron características adicionales incluyendo retrasos de 1 a 12 meses, se ajustó el modelo y se realizaron las predicciones.

Objetivo 3: Niveles mínimos y máximos de inventario.

Realizar la estimación de los niveles mínimos y máximos de inventario se convirtió en un análisis de series de tiempo. Por lo tanto, se aplicaron las mismas técnicas y validaciones husadas en el objetivo anterior:

- ARIMA: Se utilizó la función auto_arima para identificar el mejor modelo ARIMA para la serie temporal de niveles mínimos y máximos de inventario. Este modelo se ajusta automáticamente a los datos, considerando factores como la estacionalidad y las diferencias necesarias para hacer la serie estacionaria.

- Prophet: Este modelo de Facebook es adecuado para series temporales con componentes de estacionalidad y tendencia. Prophet permite descomponer la serie temporal en estas componentes y predecir futuros valores.
- Exponential Smoothing: Este método suaviza los datos mediante la aplicación de medias exponenciales, siendo útil para datos con tendencia y estacionalidad. Se configuró con tendencias y estacionalidad aditivas y se entrenó el modelo.
- RandomForestRegressor: Este algoritmo de aprendizaje automático basado en árboles de decisión se utilizó para predecir los niveles mínimos y máximos de inventario. Los modelos de Random Forest son robustos y pueden manejar datos complejos con múltiples características.

3.6.5 Evaluación de Resultados

Se evaluaron los resultados obtenidos de la minería de datos para determinar la relevancia y utilidad de los conocimientos descubiertos. Esto implicó la validación de los patrones identificados y su interpretación en el contexto de la gestión de inventario en la cadena de suministro.

Para la evaluación del objetivo 1, se utilizó la técnica del codo y el coeficiente de silueta para determinar el número óptimo de clústeres. Estas técnicas ayudaron a identificar la segmentación más adecuada para los datos, asegurando una clasificación eficiente y precisa.

Para los objetivos 2 y 3, se utilizaron las métricas MAE, MSE, RMSE y MAPE para evaluar la precisión de las predicciones realizadas por los modelos ARIMA, Prophet, Exponential Smoothing y RandomForestRegressor. Además, se realizaron gráficos comparativos para validar visualmente los resultados y asegurar que los modelos proporcionaran estimaciones confiables y útiles para la gestión de inventarios. Estas

evaluaciones permitieron seleccionar el mejor modelo para cada objetivo, optimizando así la estrategia de inventarios de la empresa.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. ANÁLISIS DE SEGMENTACIÓN PARA OBTENER LOS PRODUCTOS MÁS SIGNIFICATIVOS DE LA EMPRESA

El conjunto de datos utilizado en este análisis contiene la información de 269 productos agrupando el total de la cantidad de producto vendida, valor de venta y utilidad generada en los años 2022 y 2023. En la **Ilustración 2** de la sección anterior se puede apreciar el conjunto de datos utilizado, el cual está compuesto por 269 filas (productos) y cinco columnas: Producto, Cantidad, Venta, Utilidad y el código original del producto.

En la **Ilustración 3** se observan las medidas de tendencia central y dispersión observados, el promedio los productos muestran una cantidad de aproximadamente 431,007 unidades, ventas de alrededor de 296,752 dólares, y utilidades de 188,330 dólares. La variabilidad entre los valores mínimos y máximos es notable, destacándose cantidades desde 1 hasta 16,298,670 unidades, ventas desde 0.53 hasta 10,116,025.17 dólares, y utilidades desde 0.06 hasta 7,054,929.43 dólares. El análisis por percentiles revela que el 25 % de los productos presentan cantidades inferiores a 3,567 unidades, ventas inferiores a 6,265.92 unidades monetarias, y utilidades inferiores a 3,483.50 unidades monetarias. Además, el 50% de los productos muestran cantidades por debajo de 21,250 unidades, ventas inferiores a 28,763.26 dólares, y utilidades menores a 18,285.22 dólares. Finalmente, el 75% de los productos exhiben cantidades inferiores a 137,601 unidades, ventas por debajo de 166,359.10 dólares, y utilidades menores a 96,658.94 dólares.

Ilustración 3

Descripción de las columnas del conjunto de datos utilizado para la segmentación

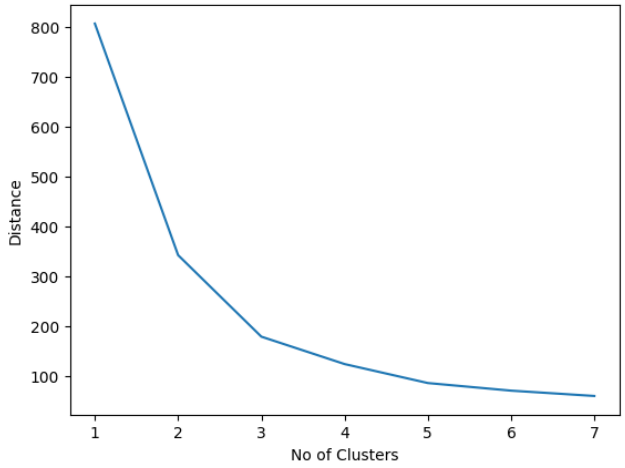
	Cantidad	Venta	Utilidad
count	269.00	269.00	269.00
mean	431007.51	296751.62	188329.94
std	1579467.53	964173.14	646468.51
min	1.00	0.53	0.06
25%	3567.00	6265.92	3483.50
50%	21250.00	28763.26	18285.22
75%	137601.00	166359.12	96658.94
max	16298672.00	10116025.17	7054929.43

Nota. Fuente: Autor

Se aplicaron dos técnicas principales de agrupación, K-means y Agglomerative Clustering, para segmentar los productos de la empresa basándose en métricas clave como cantidad, venta y utilidad. El algoritmo K-means requiere que se especifique el valor para su parámetro K que hace referencia al número de grupos que deseamos formar. Para encontrar el valor óptimo de este parámetro se usó el *Método del codo* donde se estableció un rango de K-clúster de 1 a 7. La selección final de cuatro clústeres se basó en la evaluación de la inercia promedio, como se puede observar en la **Ilustración 4** se mostró una disminución significativa al aumentar el número de clúster hasta cuatro, indicando una agrupación coherente y distintiva de los productos según las métricas de cantidad, venta y utilidad.

Ilustración 4

Método del codo para identificar el número de clúster




```
[806.9999999999999, 342.21759446267026, 178.5779933177249, 123.57838895456902, 85.47965445361208, 70.2066038534174,
```

Nota. Fuente: Autor

Por otro lado, el método de Agglomerative Clustering, que construye una estructura jerárquica de clústeres mediante el enlace de Ward, también se aplicó a los datos normalizados. Este método proporcionó una visualización clara de las relaciones entre los productos según las distancias euclidianas, respaldada por un análisis del coeficiente de silueta promedio. La **Ilustración 5** muestra cada uno de los valores obtenidos en cada n_clusters, la elección de cuatro clústeres en este método se justificó por un coeficiente de silueta que indicó una segmentación significativa y bien definida.

Ilustración 5

Coeficiente de silueta promedio para determinar cantidad de clústeres jerárquicos

```
Para n_clusters = 2 El coeficiente de silueta promedio es : 0.49300671607490637
Para n_clusters = 3 El coeficiente de silueta promedio es : 0.4611268273204411
Para n_clusters = 4 El coeficiente de silueta promedio es : 0.451641412953162
Para n_clusters = 5 El coeficiente de silueta promedio es : 0.45273536269919756
Para n_clusters = 6 El coeficiente de silueta promedio es : 0.38186254743545445
Para n_clusters = 7 El coeficiente de silueta promedio es : 0.3734741031344378
Para n_clusters = 8 El coeficiente de silueta promedio es : 0.3826908051913126
Para n_clusters = 9 El coeficiente de silueta promedio es : 0.385752432820866
Para n_clusters = 10 El coeficiente de silueta promedio es : 0.39458674837020646
```

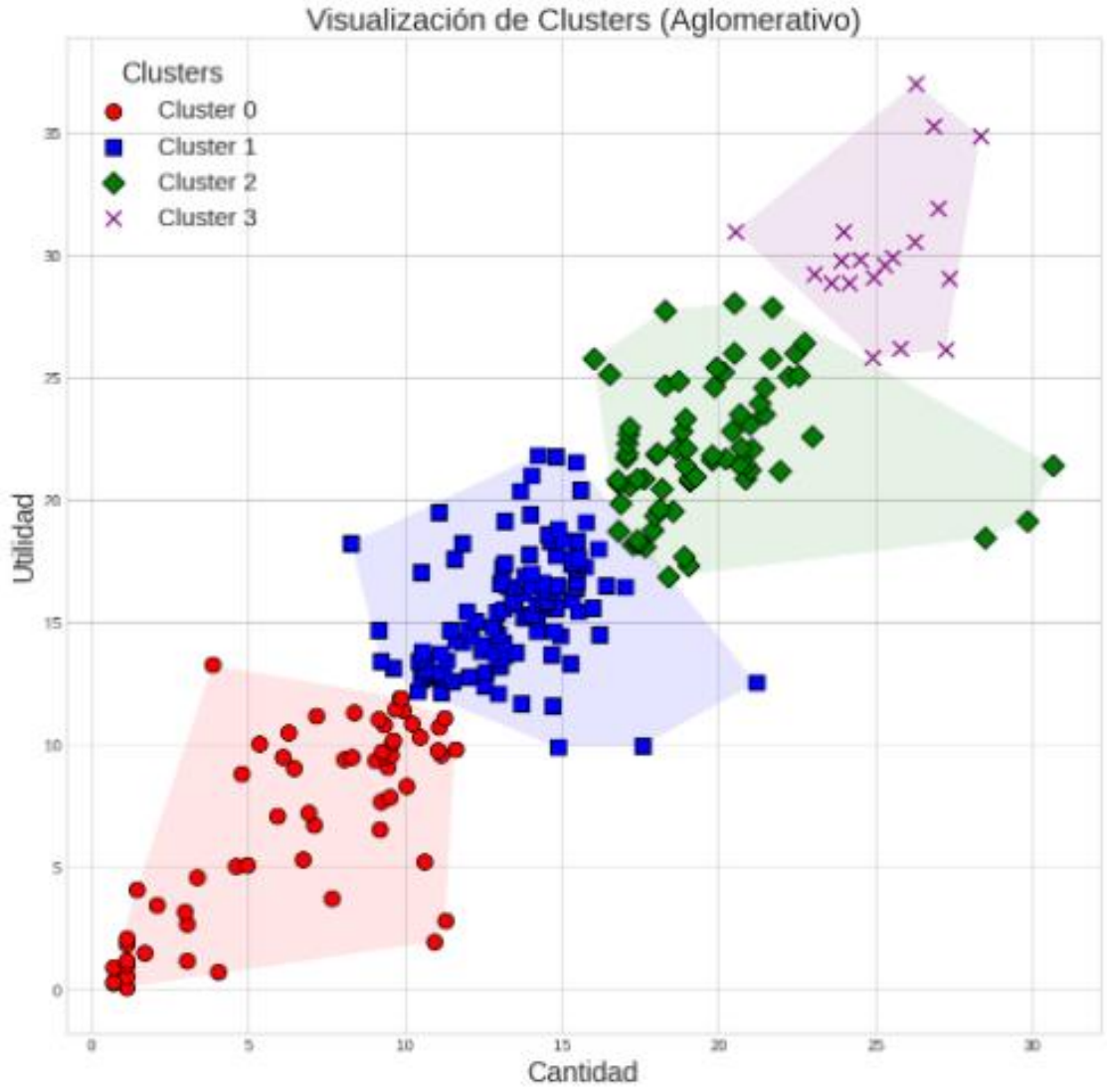
Nota. Fuente: Autor

Para comparar y evaluar los resultados obtenidos en las dos técnicas de clustering, se utilizaron gráficos de dispersión, el eje X representa las cantidades vendidas de los productos y el eje Y los valores de utilidad generada. En la **Ilustración 6** muestra los resultados de la segmentación de productos empleando la técnica de Clúster aglomerativo, la disposición de los clústeres en el gráfico revela una clara separación entre los grupos de productos, lo que sugiere que el algoritmo de clustering utilizado ha sido efectivo en identificar y agrupar productos con características similares. A continuación, se detallan los hallazgos de cada clúster:

- Clúster 0 (Rojo, círculo): Este clúster se encuentra ubicado en el cuadrante inferior izquierdo y presenta valores relativamente bajos tanto en cantidad como en utilidad. Esta situación podría indicar la necesidad de una revisión exhaustiva de las estrategias de marketing y producción asociadas a estos productos.
- Cluster 1 (Azul, cuadrado): Este clúster está localizado en la parte central izquierda del gráfico y sus productos tienen valores moderados de cantidad y utilidad, esto nos indica un rendimiento moderado de los productos.
- Cluster 2 (Verde, diamante): Los productos de este clúster se agrupan en la parte central derecha del gráfico. Estos productos tienen valores de cantidad y utilidad ligeramente superiores en comparación con los del clúster 1.
- Cluster 3 (Púrpura, cruz): Este clúster se encuentra en el cuadrante superior derecho del gráfico. Los productos en este grupo poseen los valores más altos tanto en cantidad como en utilidad. lo que podría implicar un enfoque de ventas exitoso o una demanda alta en el mercado.

Ilustración 6

Visualización de Clústeres obtenidos con la técnica de Clúster Aglomerativo

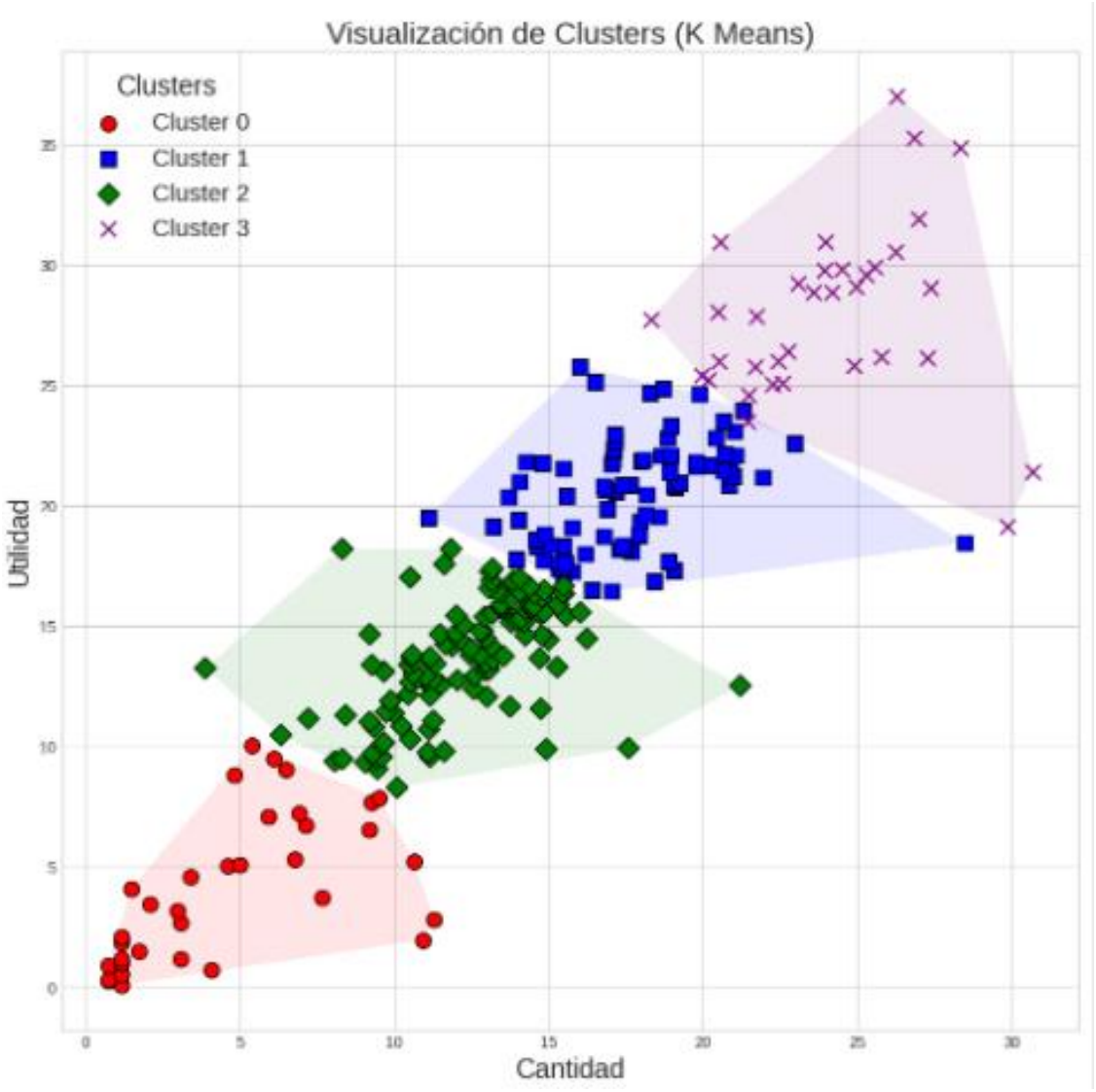


Nota. Fuente: Autor

Por otro lado, en la **Ilustración 7** se observan la segmentación generada por K-Means, esta reveló patrones similares, pero con ligeras variaciones en los rangos de las columnas utilizadas, aquí se destaca especialmente el clúster 3 por sus altos valores en todas las métricas evaluadas.

Ilustración 7

Visualización de Clusteres obtenidos con K-Means



Nota. Fuente: Autor

Se llevó a cabo un análisis de Pareto para evaluar el impacto de la segmentación de productos en la utilidad total de la empresa. Para esta evaluación, se emplearon dos técnicas de segmentación: el método aglomerativo y el algoritmo K-means. Con el fin de seleccionar la técnica más adecuada, se utilizó la métrica Silhouette Score de la biblioteca scikit-learn, la cual mide la cohesión interna de los puntos dentro de un mismo clúster en comparación con su separación respecto a otros clústeres. Los resultados obtenidos fueron

los siguientes: para el método aglomerativo, el Silhouette Score fue de 0.4499, mientras que para K-means fue de 0.4592. En base a estos resultados, se ha decidido optar por los clústeres generados mediante la técnica aglomerativo.

Una vez establecidos los clústeres, se procedió a calcular la contribución de cada clúster a la utilidad total obteniendo el promedio de las utilidades individuales de los productos dentro de cada clúster. Utilizando los datos transformados y etiquetados por clúster, se obtuvo una lista detallada de la utilidad generada, utilidad acumulada y cantidad de productos de cada grupo de productos, esto se puede observar en la **Ilustración *8***.

Ilustración 8

Utilidad promedio, utilidad acumulada y número de productos de cada clúster Aglomerativo

Cluster	Utilidad	Utilidad_Acumulada	Cant Items	Label
3	30.190103	40.427791	19	Cluster3 (19)
2	22.083945	70.000565	70	Cluster2 (70)
1	15.526272	90.791911	119	Cluster1 (119)
0	6.876289	100.000000	61	Cluster0 (61)

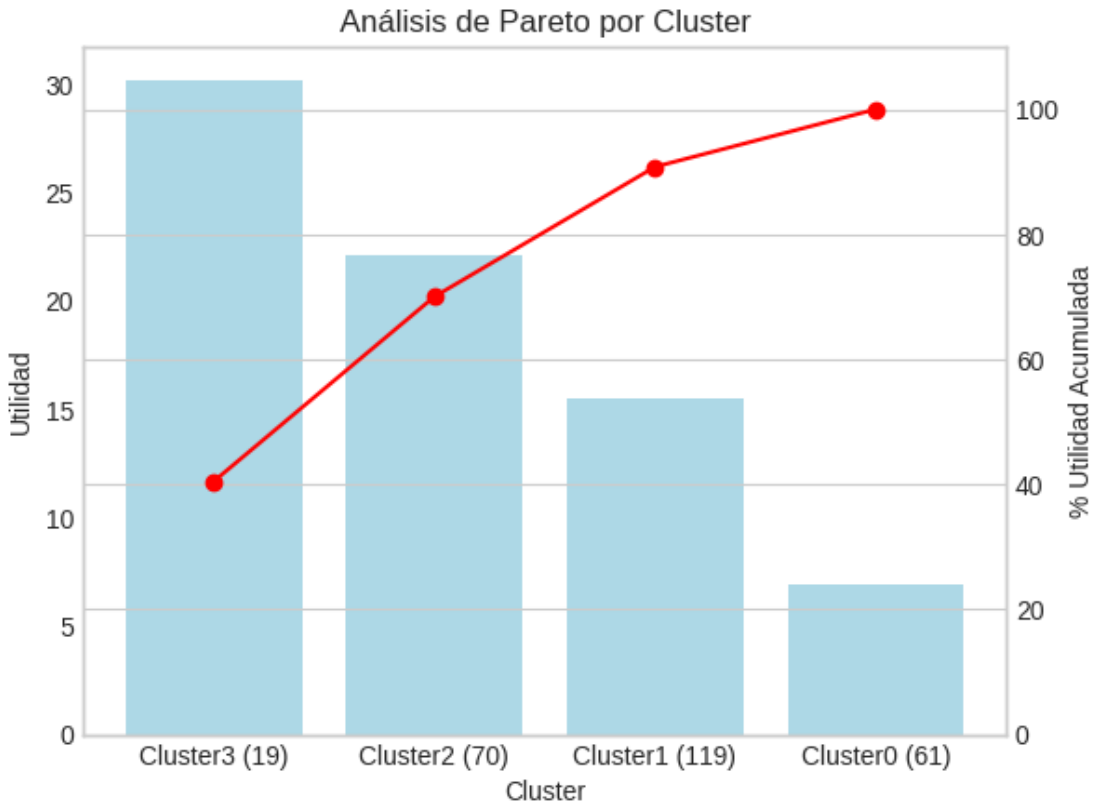
Nota. Fuente: Autor

Posteriormente, se procedió a ordenar los clústeres según su contribución a la utilidad total de la empresa. Esto se logró mediante la generación de un gráfico de barras que mostraba la utilidad acumulada de cada clúster en orden descendente, lo cual facilitó la identificación rápida de los clústeres más relevantes desde una perspectiva económica. De acuerdo con la **Ilustración 9**, los clústeres 3 y 2 destacan como los principales contribuyentes a la utilidad total de la empresa (70% de la utilidad acumulada), esto se logra con una cantidad de 89 de los 269 productos agrupados lo cual corresponde al 33 % de los productos. Estos clústeres no solo mostraron altos niveles de utilidad individual

por producto, sino que también agruparon productos cuyas ventas y utilidades se alinearon de manera favorable con los objetivos financieros de la empresa.

Ilustración 9

Gráfico de Pareto por clúster aglomerativo



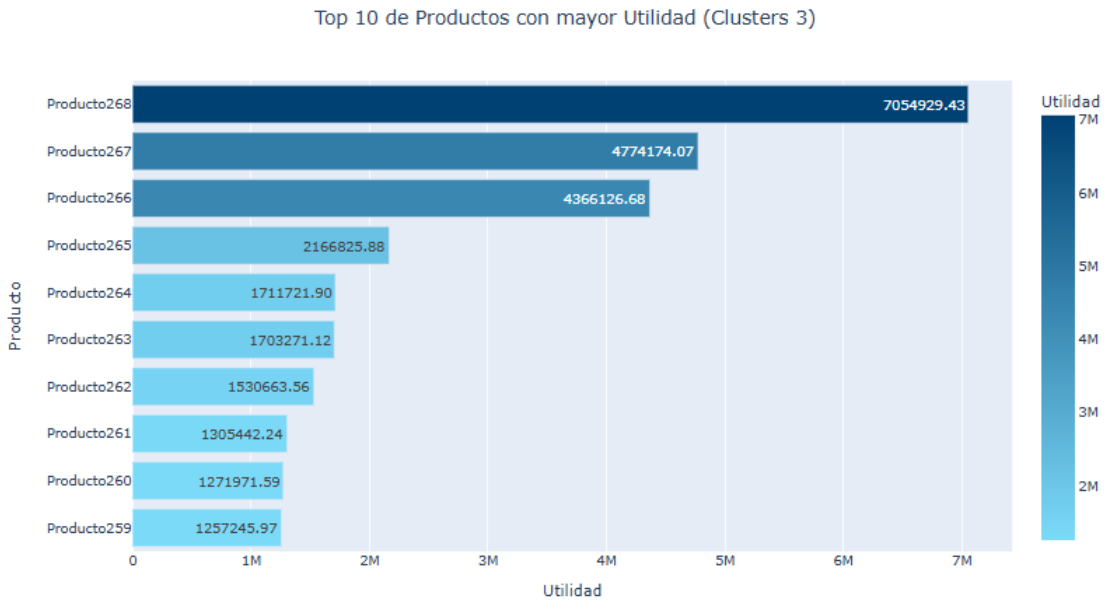
Nota. Fuente: Autor

En la **Ilustración 10** se muestra de manera ordenada los diez productos más rentables dentro del tercer clúster, basándose en la utilidad. El eje horizontal representa la utilidad de cada producto, mientras que el eje vertical enumera los productos. El Producto268 encabeza la lista con una utilidad de 7,054,929.43 dólares, seguido por Producto267 y Producto266, con utilidades de 4,774,174.07 y 4,366,126.68 dólares, respectivamente. Producto259 cierra el top 10 con una utilidad de 1,257,245.97 dólares. Este análisis destaca la importancia de centrarse en los productos más rentables, como Producto268, y sugiere la necesidad de examinar más detenidamente los productos menos rentables para

identificar oportunidades de mejora. Para los siguientes objetivos se tomaron estos productos para focalizar el análisis en aquellos que sean rentables para la empresa.

Ilustración 10

Top 10 de productos del clúster 3 con mayor utilidad



Nota. Fuente: Autor

Discusión

Los resultados de la segmentación mediante K-means y Agglomerative Clustering muestran una clara agrupación de los productos en cuatro clusters distintos. Ambos métodos de clustering identificaron patrones similares en los datos, lo que sugiere que los clusters formados son robustos y consistentes. La elección de cuatro clusters se justificó tanto por la disminución significativa en la inercia promedio en K-means como por los coeficientes de silueta promedio en Agglomerative Clustering, indicando una buena cohesión interna y separación entre los clusters. En las investigaciones de (Onumanyi et al., 2022; Patel et al., 2022) se han validado la robustez de estas técnicas para determinar el número de clústeres adecuados, a la vez que se proporciona un equilibrio entre la precisión del modelo y su complejidad

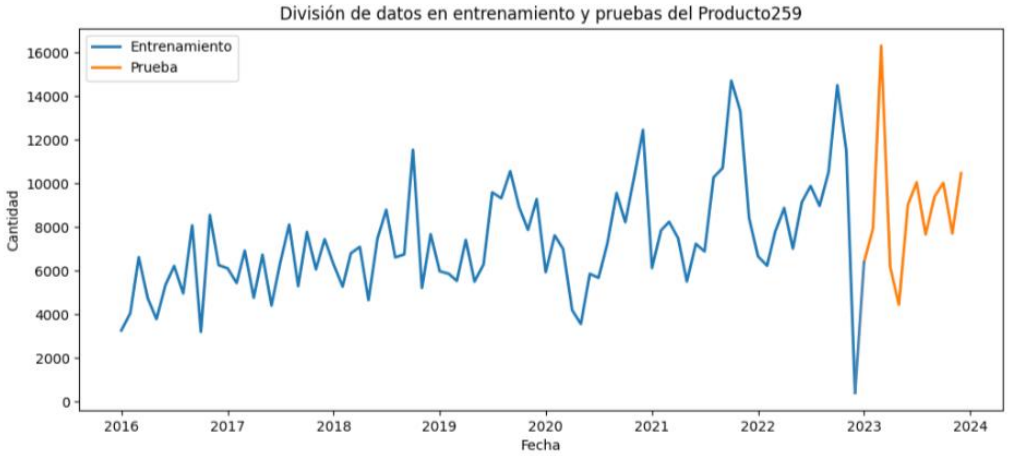
La gráfica de coordenadas paralelas proporcionó una herramienta visual poderosa para entender las diferencias y similitudes entre los clusters. Los productos en el mismo cluster comparten características comunes en términos de cantidad, venta y utilidad, lo que permite una segmentación efectiva y útil para la toma de decisiones. Esta segmentación facilita la identificación de productos que son críticos para la empresa, permitiendo estrategias más precisas en la gestión de inventarios y optimización de recursos (Chakraborty et al., 2023). Ademas, al priorizar los recursos y esfuerzos en los productos identificados en los clústeres 3 y 2, la empresa puede maximizar su retorno financiero y optimizar la disponibilidad de inventario para satisfacer la demanda del mercado de manera eficiente.

4.2. MODELO PREDICTIVO PARA DETERMINAR EL EGRESO DE PRODUCTOS MEDIANTE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

La serie temporal utilizada comprende los datos de ventas mensuales del Producto259. En la **Ilustración 11** se observa la serie temporal correspondiente cuyos datos van desde enero de 2016 hasta diciembre de 2023, con un total de 96 observaciones. De estas observaciones se tomarán los 12 meses del año 2023 para validar las predicciones de cada modelo.

Ilustración 11

Serie de tiempo del Producto259 con su respectiva división en datos de entrenamiento y pruebas



Nota. Fuente: Autor

Se implementaron y evaluaron cuatro modelos predictivos: ARIMA, Prophet, Exponential Smoothing y RandomForestRegressor. A continuación, se describen los resultados de cada modelo.

4.2.1. ARIMA

Se realizó la prueba de Dickey-Fuller aumentada (ADF) para verificar la estacionariedad de la serie de tiempo, este es un punto importante al momento de generar un modelo ARIMA. La **Ilustración *12*** muestra los resultados de la prueba ADF aplicada a la serie

de tiempo original. El valor obtenido (p-valor = 0.595928) supera el 0.05 con lo cual no se puede rechazar la hipótesis nula, esto indica que la serie no era estacionaria.

Ilustración 12

Prueba de Dickey-Fuller aumentada (ADF) en la serie de tiempo original del Producto259

```
Resultados de la prueba de Dickey-Fuller para columna: Cantidad
Test Statistic                          -1.371358
p-value                                  0.595928
No Lags Used                            11.000000
Número de observaciones utilizadas      84.000000
Critical Value (1%)                     -3.510712
Critical Value (5%)                     -2.896616
Critical Value (10%)                    -2.585482
dtype: float64
Conclusion:====>
No se puede rechazar la hipótesis nula
Los datos no son estacionarios
```

Nota. Fuente: Autor

Para transformar la serie de tiempo en estacionaria se aplicó la primera diferencia y se volvió a realizar la prueba ADF. La ***Ilustración 13*** muestra los resultados de la prueba ADF aplicada a la serie de tiempo con la primera diferencia. El valor obtenido (p-valor = 4.280958e-10) es menor a 0.05 con lo cual se rechaza la hipótesis nula, esto indica que la serie es estacionaria.

Ilustración 13

Prueba de Dickey-Fuller aumentada (ADF) en la serie de tiempo del Producto259 con 1ra diferencia

```
Resultados de la prueba de Dickey-Fuller para columna: Cantidad_diff
Test Statistic                        -7.096285e+00
p-value                                4.280958e-10
No Lags Used                           1.000000e+01
Número de observaciones utilizadas     8.400000e+01
Critical Value (1%)                   -3.510712e+00
Critical Value (5%)                   -2.896616e+00
Critical Value (10%)                  -2.585482e+00
dtype: float64
Conclusion:====>
Rechazar la hipótesis nula
Los datos son estacionarios
```

Nota. Fuente: Autor

Un modelo ARIMA(p, d, q) está conformado por 3 parámetros: p representa el orden de la parte autorregresiva(AR), d es el orden de diferenciación (I) y q es el orden de la media móvil (MA). Para encontrar el mejor valor de cada uno de estos parámetros se utilizó la función auto_arima de la librería pmdarima, el valor del parámetro d se configuró de forma manual en 1 para especificar que utilice la primera diferencia. En la ***Ilustración 14*** se puede observar que el mejor modelo seleccionado es ARIMA(1,1,1)(0,0,1)[12], determinado con base en la métrica AIC. Esta métrica es especialmente útil para evaluar modelos ARIMA, ya que equilibra la calidad del ajuste del modelo con su complejidad (Zhang & Meng, 2023).

Ilustración 14

Resultados de la función auto_arima para seleccionar el mejor modelo

```
Performing stepwise search to minimize aic
 ARIMA(1,1,1)(0,0,1)[12] intercept   : AIC=1510.828, Time=1.50 sec
 ARIMA(0,1,0)(0,0,0)[12] intercept   : AIC=1545.848, Time=0.08 sec
 ARIMA(1,1,0)(1,0,0)[12] intercept   : AIC=1532.049, Time=0.58 sec
 ARIMA(0,1,1)(0,0,1)[12] intercept   : AIC=inf, Time=1.49 sec
 ARIMA(0,1,0)(0,0,0)[12]             : AIC=1543.868, Time=0.04 sec
 ARIMA(1,1,1)(0,0,0)[12] intercept   : AIC=inf, Time=0.41 sec
 ARIMA(1,1,1)(1,0,1)[12] intercept   : AIC=1519.379, Time=3.35 sec
 ARIMA(1,1,1)(0,0,2)[12] intercept   : AIC=1513.164, Time=4.28 sec
 ARIMA(1,1,1)(1,0,0)[12] intercept   : AIC=inf, Time=1.32 sec
 ARIMA(1,1,1)(1,0,2)[12] intercept   : AIC=inf, Time=4.61 sec
 ARIMA(1,1,0)(0,0,1)[12] intercept   : AIC=1526.872, Time=0.66 sec
 ARIMA(2,1,1)(0,0,1)[12] intercept   : AIC=1519.453, Time=3.44 sec
 ARIMA(1,1,2)(0,0,1)[12] intercept   : AIC=inf, Time=2.78 sec
 ARIMA(0,1,0)(0,0,1)[12] intercept   : AIC=1538.616, Time=0.13 sec
 ARIMA(0,1,2)(0,0,1)[12] intercept   : AIC=inf, Time=2.53 sec
 ARIMA(2,1,0)(0,0,1)[12] intercept   : AIC=1520.668, Time=0.85 sec
 ARIMA(2,1,2)(0,0,1)[12] intercept   : AIC=inf, Time=2.54 sec
 ARIMA(1,1,1)(0,0,1)[12]             : AIC=1512.569, Time=1.05 sec

Best model:  ARIMA(1,1,1)(0,0,1)[12] intercept
Total fit time: 31.716 seconds
```

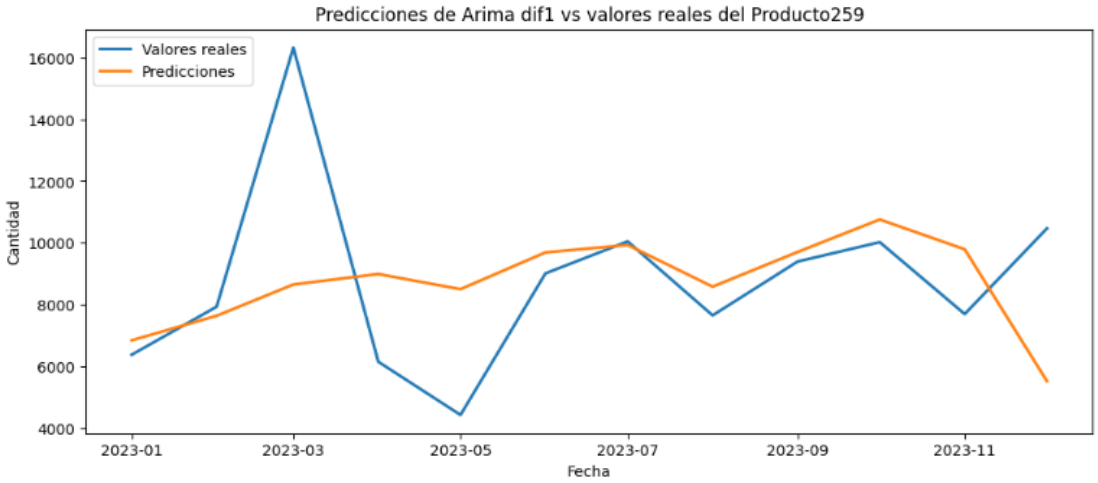
Nota. Fuente: Autor

El modelo ARIMA seleccionado se ajustó y evaluó en el conjunto de prueba correspondiente al año 2023.La **Ilustración 15** muestra la comparación gráfica de los valores reales con las predicciones del modelo ARIMA, estos resultados muestran que,

aunque el modelo capturó bien las tendencias estacionales, tuvo limitaciones en ajustarse completamente a las variaciones mensuales.

Ilustración 15

Comparación de valores de prueba con las predicciones del modelo ARIMA



Nota. Fuente: Autor

4.2.2. Prophet

Prophet es conocido por su capacidad para manejar estacionalidades y eventos especiales. Para este modelo fue necesario renombrar las variables del conjunto de datos, la columna cantidad pasó a llamarse “y”, y la columna Fecha se renombró a “ds”. Con este conjunto de datos inicial el modelo crea sus propias variables para intentar predecir los valores de la serie de tiempo, estas variables se pueden observar en la **Ilustración *16***.

Ilustración 16

Conjunto de datos con variables auto creadas por el modelo Prophet

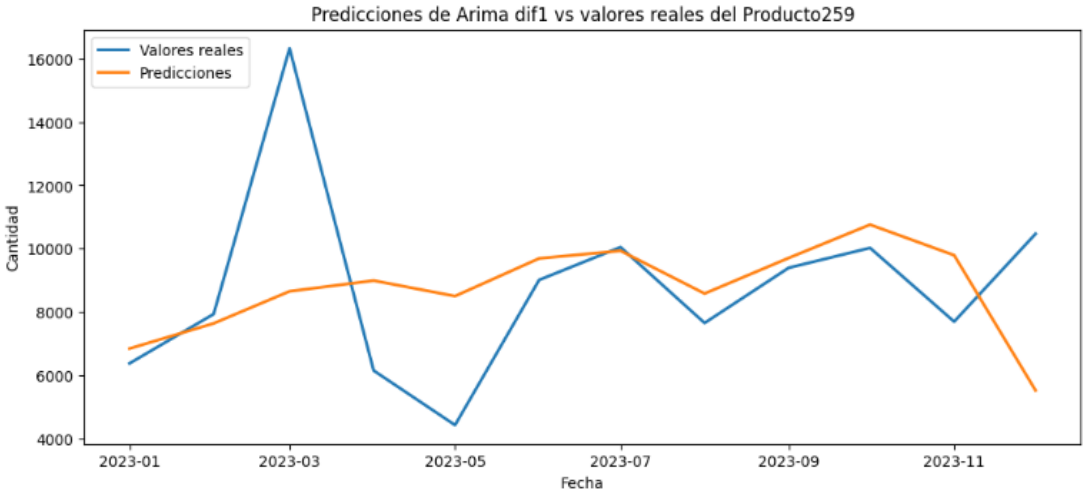
	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive_terms_lower	additive_terms_upper	yearly
0	2016-01-01	5531.504270	2179.916469	6598.804382	5531.504270	5531.504270	-1130.391883	-1130.391883	-1130.391883	-1130.391883
1	2016-02-01	5577.028546	2899.677752	7240.847746	5577.028546	5577.028546	-518.405389	-518.405389	-518.405389	-518.405389
2	2016-03-01	5619.615772	3544.872857	8104.895251	5619.615772	5619.615772	213.584929	213.584929	213.584929	213.584929
3	2016-04-01	5665.140049	1696.652957	6174.973519	5665.140049	5665.140049	-1778.468338	-1778.468338	-1778.468338	-1778.468338
4	2016-05-01	5709.195800	1977.711025	6354.847972	5709.195800	5709.195800	-1729.603543	-1729.603543	-1729.603543	-1729.603543
...	...	...	...	...	...	...	...	...	...	...

Nota. Fuente: Autor

El modelo Prophet produjo predicciones para los 12 meses de 2023, la **Ilustración 17** muestra la comparación gráfica de los valores reales con las predicciones. El modelo logró una mejora en cuanto a las predicciones de ARIMA.

Ilustración 17

Comparación de valores de prueba con las predicciones del modelo Prophet



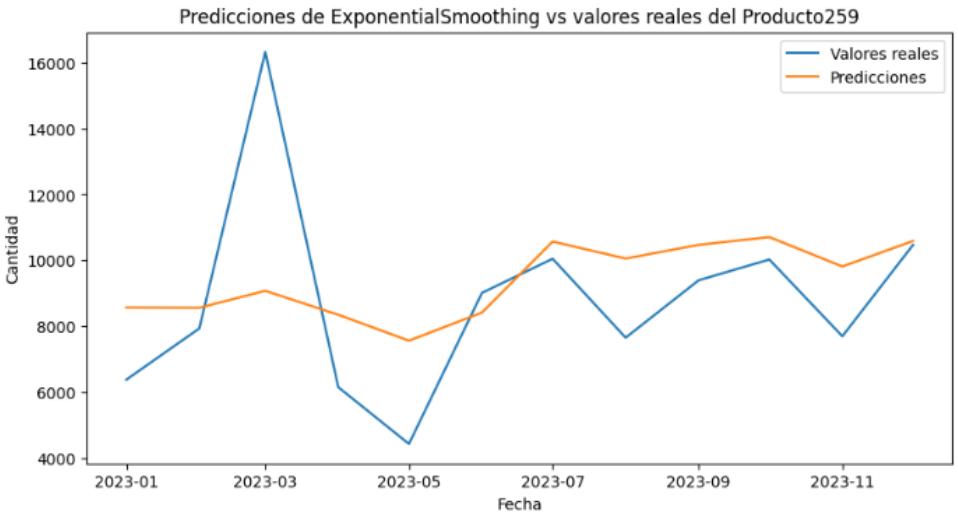
Nota. Fuente: Autor

4.2.3. ExponentialSmoothing

El modelo Exponential Smoothing se configuró con tendencias y estacionalidad aditivas, y un periodo de estacionalidad de 12 (estacionalidad anual). La **Ilustración 18** muestra la comparación gráfica de los valores reales con las predicciones del modelo ExponentialSmoothing.

Ilustración 18

Comparación de valores de prueba con las predicciones del modelo ExponentialSmoothing



Nota. Fuente: Autor

4.2.4. Random forest

Para el modelo RandomForestRegressor, se crearon variables adicionales para capturar la estacionalidad y la tendencia, incluyendo el año, el mes y los retrasos de 1 a 12 meses. Con las colunas de retraso creadas se tuvo que eliminar las columnas donde estas tenían datos faltantes. La **Ilustración 19** muestra el conjunto de datos usado en este modelo.

Ilustración 19

Conjunto de datos usado en el modelo de Random forest

	ds	y	Year	Mes	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12
12	2017-01-01	6104.0	2017	1	6251.0	8554.0	3192.0	8072.0	4953.0	6212.0	5337.0	3776.0	4728.0	6617.0	4046.0	3255.0
13	2017-02-01	5433.0	2017	2	6104.0	6251.0	8554.0	3192.0	8072.0	4953.0	6212.0	5337.0	3776.0	4728.0	6617.0	4046.0
14	2017-03-01	6913.0	2017	3	5433.0	6104.0	6251.0	8554.0	3192.0	8072.0	4953.0	6212.0	5337.0	3776.0	4728.0	6617.0
15	2017-04-01	4752.0	2017	4	6913.0	5433.0	6104.0	6251.0	8554.0	3192.0	8072.0	4953.0	6212.0	5337.0	3776.0	4728.0
16	2017-05-01	6726.0	2017	5	4752.0	6913.0	5433.0	6104.0	6251.0	8554.0	3192.0	8072.0	4953.0	6212.0	5337.0	3776.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
91	2023-08-01	7658.0	2023	8	10051.0	9016.0	4437.0	6157.0	16321.0	7939.0	6384.0	399.0	11490.0	14505.0	10527.0	8967.0
92	2023-09-01	9399.0	2023	9	7658.0	10051.0	9016.0	4437.0	6157.0	16321.0	7939.0	6384.0	399.0	11490.0	14505.0	10527.0
93	2023-10-01	10027.0	2023	10	9399.0	7658.0	10051.0	9016.0	4437.0	6157.0	16321.0	7939.0	6384.0	399.0	11490.0	14505.0
94	2023-11-01	7701.0	2023	11	10027.0	9399.0	7658.0	10051.0	9016.0	4437.0	6157.0	16321.0	7939.0	6384.0	399.0	11490.0
95	2023-12-01	10470.0	2023	12	7701.0	10027.0	9399.0	7658.0	10051.0	9016.0	4437.0	6157.0	16321.0	7939.0	6384.0	399.0

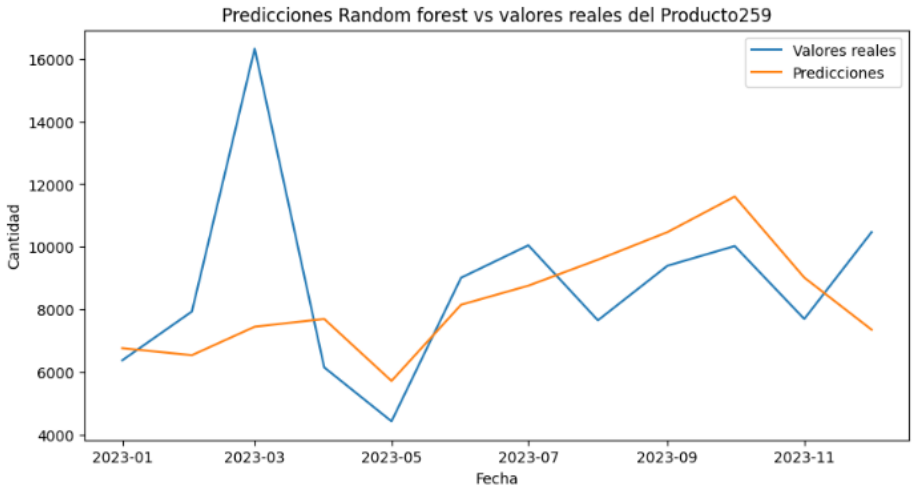
84 rows × 16 columns

Nota. Fuente: Autor

El modelo RandomForestRegressor produjo predicciones que fueron evaluadas en comparación con los valores reales de ventas, la **Ilustración 20** muestra la comparación gráfica de estos valores. Aunque se esperaba que este modelo capturara mejor las complejidades no lineales, su desempeño no fue el mejor en este caso.

Ilustración 20

Comparación de valores de prueba con las predicciones del modelo Random forest



Nota. Fuente: Autor

4.2.5. Comparación de los resultados

Las predicciones de cada modelo para los 12 meses del año 2023 se compararon con los valores reales, y se calcularon las métricas de evaluación para determinar el desempeño de cada modelo. La **Tabla 2** contiene las métricas obtenidas en cada modelo aplicado.

Tabla 2

Comparación de las métricas obtenidas de las predicciones (Producto259).

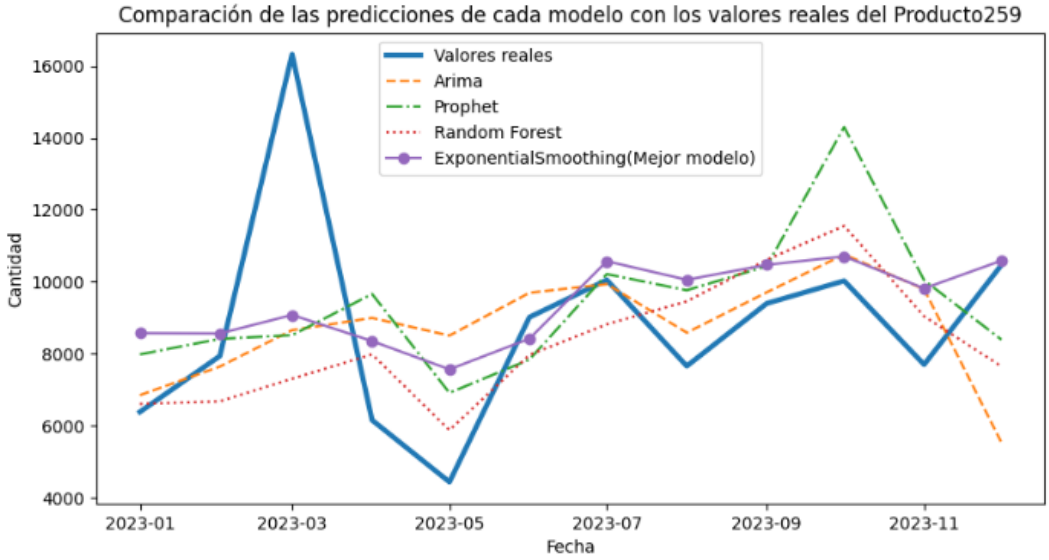
	MAE	MSE	RMSE	MAPE
Arima	2094.32	9533434.68	3087.63	0.2513
Prophet	2419.59	9756695.64	3123.57	0.2813
ExponentialSmoothing	**1907.38**	**7061498.16**	**2657.35**	0.2355
Random Forest	2061.61	8976999.37	2996.16	**0.2137**

Nota. Fuente: Autor

De acuerdo con las métricas de evaluación, ExponentialSmoothing es el mejor modelo en términos de MAE, MSE y RMSE. Sin embargo, en términos de MAPE, Random Forest es el mejor, indicando que tiene el menor porcentaje de error absoluto medio. Dado que la mayoría de las métricas (MAE, MSE y RMSE) favorecen a ExponentialSmoothing, y estas son métricas comúnmente utilizadas para evaluar la precisión de los modelos, podemos concluir que ExponentialSmoothing es el mejor modelo general en este caso. Estos resultados también se lo pude verificar de manera gráfica en **Ilustración 21**.

Ilustración 21

Comparación de las predicciones de cada modelo con los valores reales del Producto259



Nota. Fuente: Autor

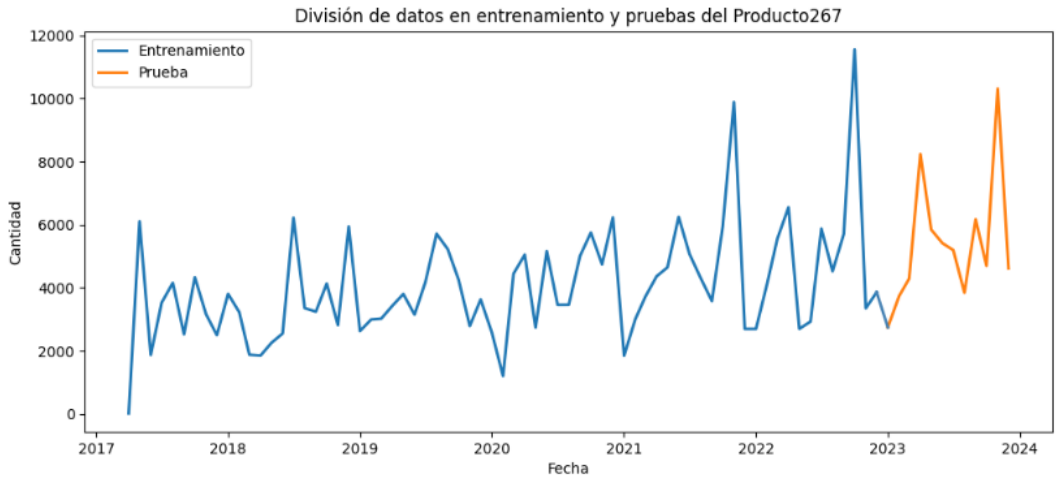
4.2.6. Aplicación de algoritmos en otros productos

Para validar los modelos implementados, se realizó un análisis adicional a dos productos: Producto267 y Producto266, los cuales se encuentran entre los tres principales de la lista según el análisis de Pareto obtenido en el objetivo anterior. En la **Ilustración 22** se presenta la serie temporal correspondiente al Producto267, cuyos datos abarcan desde abril de 2017 hasta diciembre de 2023, acumulando un total de 81 meses en un período

de 7 años. Para la validación de las predicciones de cada modelo, se utilizarán los datos de los 12 meses del año 2023.

Ilustración 22

Serie de tiempo del Producto267 con su respectiva división en datos de entrenamiento y pruebas



Nota. Fuente: Autor

Las predicciones de cada modelo para los 12 meses del año 2023 se compararon con los valores reales, y se calcularon las métricas de evaluación para determinar el desempeño de cada modelo. La **Tabla 3** contiene las métricas obtenidas en cada modelo aplicado.

Tabla 3

Comparación de las métricas obtenidas de las predicciones de cada modelo (Producto267).

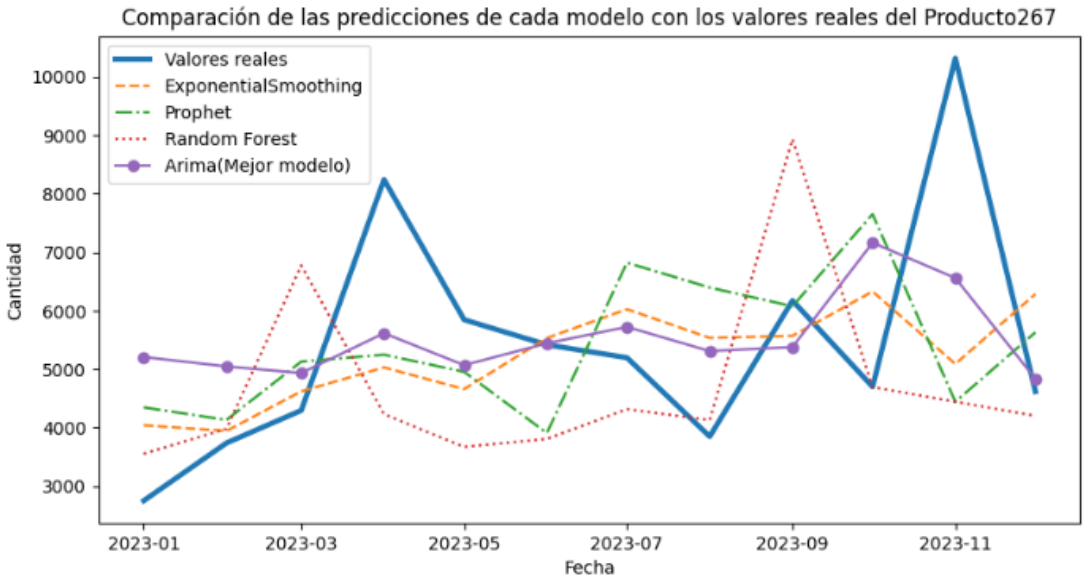
	MAE	**MSE**	**RMSE**	**MAPE**
Arima	**1421.55**	**3246492.04**	**1801.80**	0.2829
Prophet	1861.13	5736509.79	2395.10	0.3402
ExponentialSmoothing	1498.84	4186778.70	2046.16	**0.2604**
Random Forest	1795.08	6119298.65	2473.72	0.2867

Nota. Fuente: Autor

De acuerdo con las métricas de evaluación, ARIMA es el mejor modelo en términos de MAE, MSE y RMSE. Sin embargo, en términos de MAPE, ExponentialSmoothing es el mejor, indicando que tiene el menor porcentaje de error absoluto medio. Dado que la mayoría de las métricas (MAE, MSE y RMSE) favorecen a ARIMA, y estas son métricas comúnmente utilizadas para evaluar la precisión de los modelos, podemos concluir que ARIMA es el mejor modelo para predecir los 12 meses del año 2023 del Producto267. Estos resultados también se lo pude verificar de manera gráfica en la **Ilustración 23**.

Ilustración 23

Comparación de las predicciones de cada modelo con los valores reales del Producto267



Nota. Fuente: Autor

En la **Ilustración 24** se presenta la serie temporal correspondiente al Producto266, cuyos datos abarcan desde abril de 2017 hasta diciembre de 2023, acumulando un total de 81 meses en un período de 7 años. Para la validación de las predicciones de cada modelo, se utilizarán los datos de los 12 meses del año 2023.

Ilustración 24

Serie de tiempo del Producto266 con su respectiva división en datos de entrenamiento y pruebas

Nota. Fuente: Autor

Las predicciones de cada modelo para los 12 meses del año 2023 se compararon con los valores reales, y se calcularon las métricas de evaluación para determinar el desempeño de cada modelo. La **Tabla 4** contiene las métricas obtenidas en cada modelo aplicado.

Tabla 4

Comparación de las métricas obtenidas de las predicciones de cada modelo (Producto266).

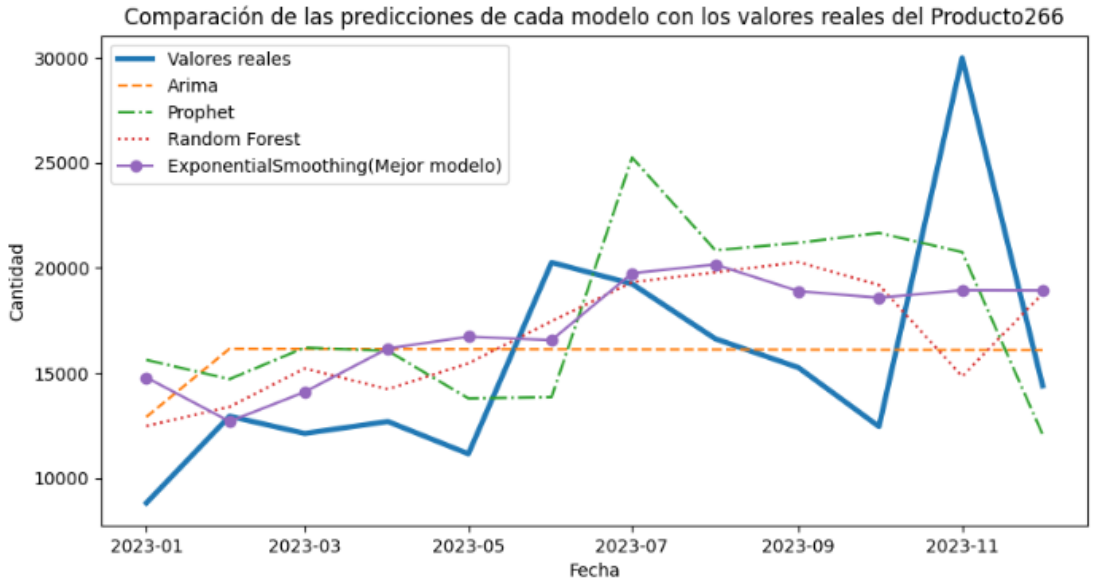
	MAE	MSE	RMSE	MAPE
Arima	3970.28	26437878.81	5141.78	0.26
prophet	5168.40	32544053.87	5704.74	0.35
ExponentialSmoothing	**4196.54**	**25247852.80**	**5024.72**	0.29
Random Forest	4200.07	31804642.05	5639.56	**0.27**

Nota. Fuente: Autor

De acuerdo con las métricas de evaluación, ARIMA es el mejor modelo en términos de MAE y MAPE. Sin embargo, si observamos la **Ilustración 25** , el modelo ARIMA a partir del mes 2 del 2023 a realizado predicciones iguales para los 11 meses restantes, por lo tanto, seleccionaremos al segundo mejor modelo como el más eficiente. ExponentialSmoothing es el segundo mejor modelo debido a que la mayoría de las métricas (MAE, MSE y RMSE) lo favorecen.

Ilustración 25

Comparación de las predicciones de cada modelo con los valores reales del Producto266



Nota. Fuente: Autor

Discusión

Para aplicar el modelo ARIMA se tuvo que transformar la serie temporal de ventas mensuales para garantizar la estacionariedad necesaria para la modelización. En particular, se utilizó la diferenciación para lograr una serie estacionaria (Sri Kalli & Pullagura, 2020). Aunque los resultados del modelo ARIMA mostraron un buen desempeño, con un RMSE de 3087.63, la no linealidad y complejidad de los datos

limitaron su precisión. El modelo ARIMA capturó bien las tendencias estacionales, pero no logró ajustar completamente las variaciones mensuales.

La **Tabla 5** presenta un resumen de los valores correspondientes a cada una de las métricas de evaluación para los productos Producto259, Producto267 y Producto266. En dicha tabla se observan diferencias significativas en el desempeño de los modelos aplicados (ARIMA, Prophet, Exponential Smoothing y Random Forest). Para el Producto259, el modelo que obtuvo el mejor desempeño fue Exponential Smoothing, con los valores más bajos de MAE (1907.38), MSE (7,061,498.16), RMSE (2657.35) y MAPE (0.2355). En el caso del Producto267, el modelo ARIMA se destacó con un desempeño superior, reflejado en una MAE de 1421.55, MSE de 3,246,492.04, RMSE de 1801.80 y MAPE de 0.2829. Para el Producto266, aunque el modelo ARIMA presentó los mejores valores de MAE (3970.28) y MAPE (0.26), sus predicciones fueron constantes a lo largo de los 11 meses de 2023. Debido a esta limitación, se optó por el segundo mejor modelo, Exponential Smoothing, que mostró un rendimiento satisfactorio con una MAE de 4196.54, MSE de 25,247,852.80, RMSE de 5024.72 y MAPE de 0.29. Estos resultados indican que Exponential Smoothing es el modelo más adecuado para los productos Producto259 y Producto266, mientras que ARIMA es el más efectivo para Producto267. Esto es consistente con el estudio de Taylor & Letham (2018) que destacan la efectividad de Exponential Smoothing en la previsión de series temporales con patrones estacionales.

Tabla 5

Resumen de la comparación de métricas obtenidas en las predicciones de cada modelo del Producto259, Producto267 y Producto266

Producto	Modelos	**MAE**	**MSE**	**RMSE**	**MAPE**
Producto 259	**Arima**	2094.32	9533434.68	3087.63	0.2513
	Prophet	2419.59	9756695.64	3123.57	0.2813
	ExponentialSmoothing	**1907.38**	**7061498.16**	**2657.35**	0.2355
	Random Forest	2061.61	8976999.37	2996.16	**0.2137**
Producto 267	**Arima**	**1421.55**	**3246492.04**	**1801.80**	0.2829
	Prophet	1861.13	5736509.79	2395.10	0.3402
	ExponentialSmoothing	1498.84	4186778.70	2046.16	**0.2604**
	Random Forest	1795.08	6119298.65	2473.72	0.2867
Producto 266	**Arima**	~~3970.28~~	~~26437878.81~~	~~5141.78~~	~~0.26~~
	prophet	5168.40	32544053.87	5704.74	0.35
	ExponentialSmoothing	**4196.54**	**25247852.80**	**5024.72**	0.29
	Random Forest	4200.07	31804642.05	5639.56	**0.27**

Nota. Fuente: Autor

4.3.EXISTENCIAS MÍNIMA Y MÁXIMA DE PRODUCTOS MEDIANTE TÉCNICAS DE OPTIMIZACIÓN PARA GARANTIZAR LAS NECESIDADES DE LA EMPRESA.

Para calcular los niveles mínimos y máximos de inventario, primero se recopiló y procesó el detalle de los movimientos de inventario relacionados con las ventas del producto seleccionado. Esta información fue agrupada por año y mes para obtener la suma de las cantidades vendidas. En la **Ilustración *26*** se listan las columnas utilizadas para obtener los valores mínimos y máximos de inventario.

Ilustración 26

Columnas utilizadas en el cálculo de los niveles mínimos y máximos de inventario.

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 94 entries, 2016-03-01 to 2023-12-01
Data columns (total 19 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Producto                94 non-null     object
 1   Anio                    94 non-null     int64
 2   Mes                     94 non-null     int64
 3   Cantidad                94 non-null     float64
 4   LeadTime                94 non-null     int64
 5   L1                      94 non-null     float64
 6   L2                      94 non-null     float64
 7   L3                      94 non-null     float64
 8   ServiceLevel            94 non-null     float64
 9   Total                   94 non-null     float64
 10  Promedio_mensual        94 non-null     float64
 11  Promedio_diario         94 non-null     float64
 12  Desviacion_estandar_ms  94 non-null     float64
 13  LeadTime_mes            94 non-null     float64
 14  ServiceLevel_Z          94 non-null     float64
 15  Inventario_seguridad    94 non-null     float64
 16  PuntoReOrden            94 non-null     float64
 17  Minimo                  94 non-null     float64
 18  Maximo                  94 non-null     float64
dtypes: float64(15), int64(3), object(1)
memory usage: 14.7+ KB
```

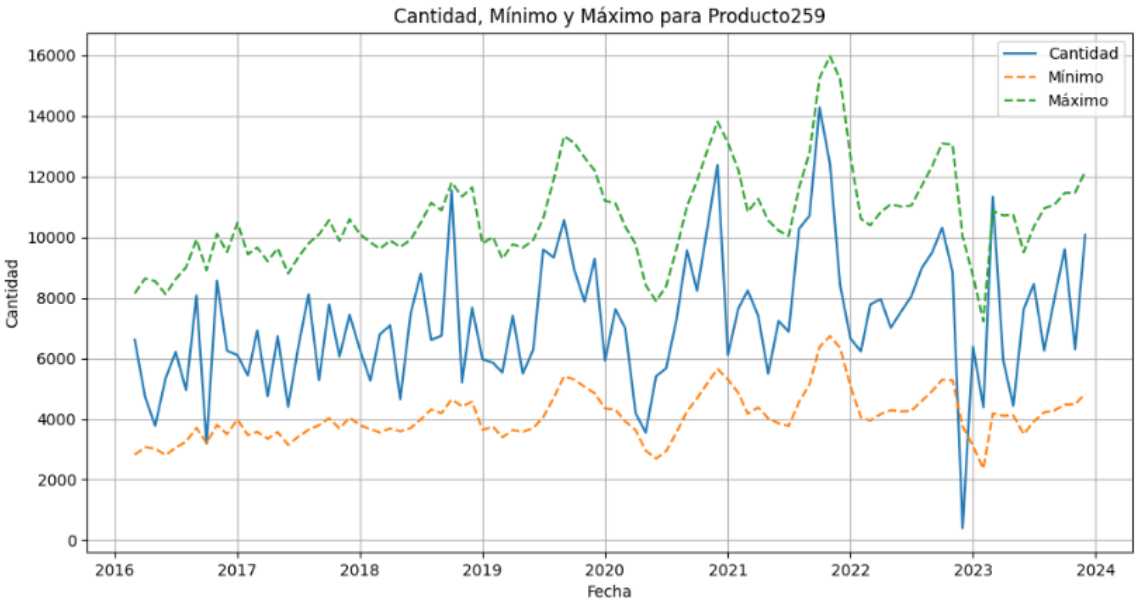
Nota. Fuente: Autor

En la **Ilustración 27** se puede observar que la línea de tiempo de la cantidad vendida del producto 259 se encuentra dentro de los niveles mínimos y máximos calculados. Esta

información se la exportó a un archivo CSV para realizar posteriormente la predicción con los modelos mencionados anteriormente.

Ilustración 27

Serie de tiempo correspondientes a la cantidad de venta, nivel mínimo y máximo de inventario.



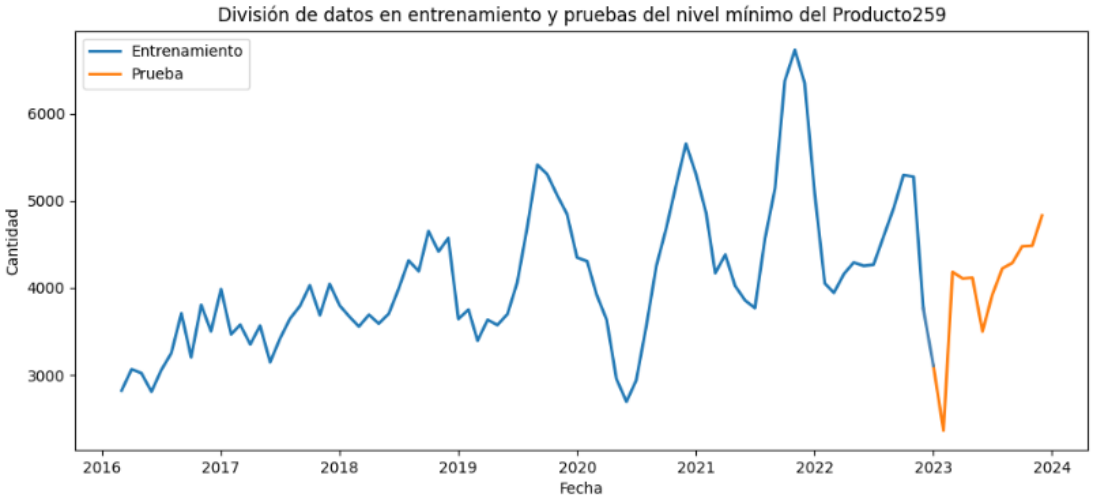
Nota. Fuente: Autor

4.3.1. Análisis de la Serie Temporal de Niveles Mínimos de Inventario

En la **Ilustración 28** se observa la línea de tiempo del nivel mínimo de inventario cuyos datos van desde marzo del 2016 a diciembre del 2023, de estos se tomaron los últimos 12 datos para hacer las validaciones. Se realizó la prueba de Dickey-Fuller aumentada (ADF) para determinar si era estacionaria, se obtuvo un P-value de 0.326826, indicando que la serie original no era estacionaria. Para solucionar esto, se aplicó una primera diferencia, lo cual convirtió la serie en estacionaria con un P-value de 9.854459e-10, menor al umbral de 0.05.

Ilustración 28

Serie de tiempo del nivel mínimo de inventario del producto259 con su respectiva división en datos de entrenamiento y pruebas



Nota. Fuente: Autor

El mejor modelo ARIMA identificado por la función auto_arima fue ARIMA(3,1,0)(0,0,1)[12]. Este modelo se utilizó para predecir los niveles mínimos de inventario. Además del modelo ARIMA, se aplicaron los modelos Prophet, Exponential Smoothing y Random Forest, con las mismas configuraciones utilizadas en el objetivo 2. Se realizaron predicciones con cada modelo para los 12 meses de 2023, estas predicciones se compararon con los valores reales, y se calcularon las métricas de evaluación para determinar el desempeño de cada modelo. En la **Tabla 6** se presentan los resultados obtenidos de las métricas MAE, MSE, RMSE y MAPE.

Tabla 6

Comparación de las Métricas de las predicciones de los niveles mínimos del Producto259

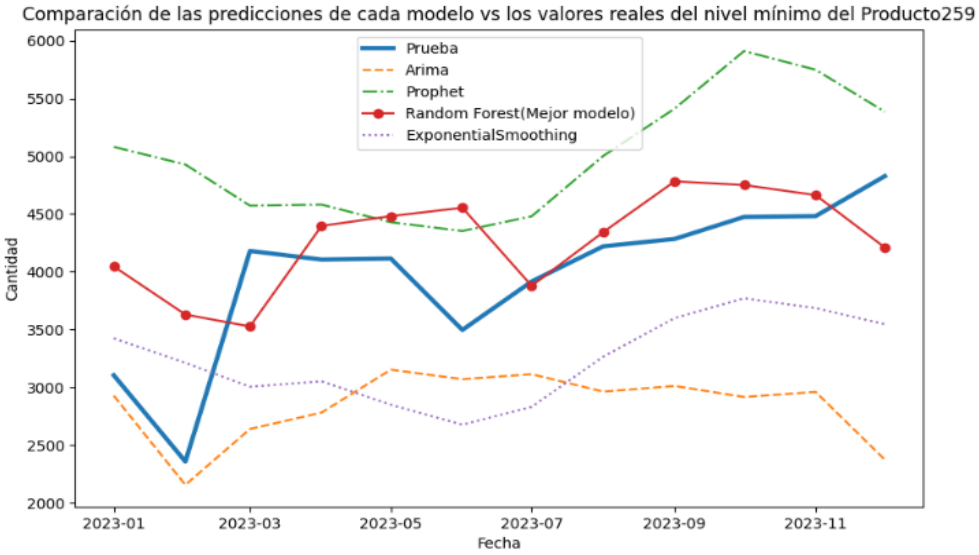
	MAE	MSE	RMSE	MAPE
Arima	1123.66	1656217.05	1286.94	0.2651
Prophet	1025.49	1488992.57	1220.24	0.2970
ExponentialSmoothing	915.26	908654.36	953.23	0.2337
Random Forest	**526.01**	**418074.10**	**646.59**	**0.1537**

Nota. Fuente: Autor

De acuerdo con las métricas de evaluación, el modelo Random Forest presentó los mejores resultados, con los valores más bajos de MAE, MSE y RMSE, y el valor más cercano a 0 para MAPE, indicando que fue el modelo más preciso y confiable para predecir los niveles mínimos de inventario. En la **Ilustración 29** se pueden observar los resultados de las predicciones de cada uno de los modelos.

Ilustración 29

Comparación de las predicciones de cada modelo con el valor real de la serie de nivel mínimo de inventario del Producto259



Nota. Fuente: Autor

4.3.2. Análisis de la Serie Temporal de Niveles Máximos de Inventario

El siguiente paso fue analizar la serie temporal de los niveles máximos de inventario (ver **Ilustración 30**). La prueba ADF inicial dio un P-value de 0.326704, indicando que la serie no era estacionaria. Aplicando una primera diferencia se obtuvo un P-value de 9.718273e-10, confirmando que la serie se convirtió en estacionaria.

Ilustración 30

Serie de tiempo del nivel máximo de inventario del producto259 con su respectiva división en datos de entrenamiento y pruebas

Nota. Fuente: Autor

El mejor modelo ARIMA encontrado fue ARIMA(0,1,3)(0,0,1)[12]. Al igual que para los niveles mínimos de inventario, se aplicaron los modelos Prophet, Exponential Smoothing y Random Forest. Se evaluaron los resultados de las predicciones de cada modelo con los valores originales de los niveles máximos de inventario. En la **Tabla 7** se presentan los resultados obtenidos:

Tabla 7

Métricas de las predicciones de los nivele máximos del Producto259

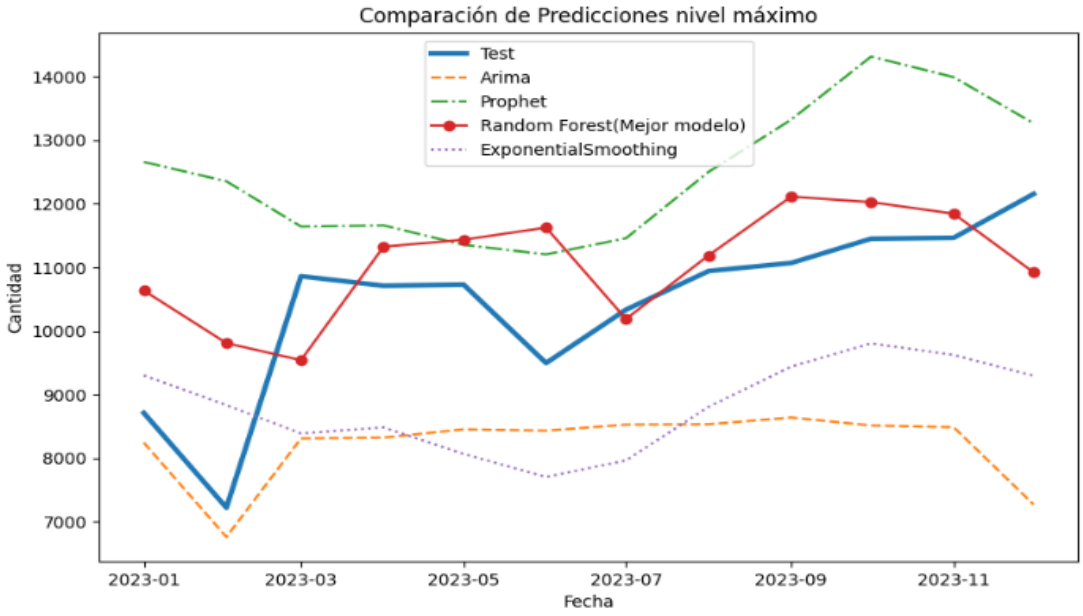
	MAE	MSE	RMSE	MAPE
Arima	2221.40	6274870.53	2504.97	0.2027
Prophet	2048.24	5944149.44	2438.06	0.2149
ExponentialSmoothing	1986.84	4287207.19	2070.56	0.1896
Random Forest	**1064.37**	**1677679.21**	**1295.25**	**0.1122**

Nota. Fuente: Autor

De acuerdo con las métricas de evaluación, El modelo Random Forest nuevamente presentó los mejores resultados, con los valores más bajos de MAE, MSE y RMSE, y el valor más cercano a 0 para MAPE, destacándose como el modelo más adecuado para predecir los niveles máximos de inventario. Estos resultados también se lo pude verificar de forma gráfica en la **Ilustración *31***.

Ilustración 31

Comparación de las predicciones de cada modelo con el valor real de la serie de nivel máximo de inventario



Nota. Fuente: Autor

Discusión

Se construyó un conjunto de datos que incluía las ventas mensuales de productos, tiempos de entrega (LeadTime) y valores de retraso (lags) esto con el objetivo de calcular los niveles mínimo y máximo en cada uno de los meses. Las series de tiempo obtenidas respecto a niveles mínimos y máximos de inventario fue evaluada y transformada para asegurar su estacionariedad.

De acuendo con Kourentzes et al. (2020), en el ámbito de la gestión de inventarios la predicción precisa de los niveles mínimos y máximos es crucial para evitar tanto la falta de stock como el exceso de inventario. La falta de stock puede llevar a pérdidas de ventas y disminución en la satisfacción del cliente, mientras que el exceso de inventario implica costos adicionales de almacenamiento y riesgo de obsolescencia de productos. Por lo tanto, optimizar estos niveles mediante técnicas avanzadas de predicción no solo mejora la eficiencia operativa, sino que también contribuye significativamente a la rentabilidad de la empresa.

Al aplicar los modelos ARIMA, Prophet, Exponential Smoothing y Random Forest para predecir los niveles de inventario mínimo y máximo, se encontró que el Random Forest Regressor ofrecía las predicciones más precisas. Este modelo mostró una capacidad superior para manejar la variabilidad y la complejidad de los datos de inventario, lo que es esencial para mantener un balance óptimo entre la disponibilidad de productos y los costos de almacenamiento.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- La implementación de los algoritmos de agrupación K-means y Agglomerative Clustering resultó efectiva para clasificar los productos según sus características de cantidad, venta y utilidad. Ambos métodos de clustering identificaron patrones similares en los datos logrando identificaron cuatro clusters significativos que ofrecen una perspectiva clara sobre la distribución y el comportamiento de los productos. Además, con el análisis de Pareto reveló que los clústeres 3 y 2 son los principales contribuyentes a la utilidad total de la empresa con una cantidad de 89 de los 269 productos agrupados, esto recalca la importancia de centrarse en estos grupos para maximizar la rentabilidad.
- Se implementaron los modelos ARIMA, Prophet, Exponential Smoothing y RandomForestRegressor lo cual permitió comparar distintos enfoques para predecir el consumo mensual de un producto específico. El modelo Exponential Smoothing mostró el mejor desempeño, con los menores valores de MAE y RMSE, lo que indica que fue más preciso en capturar las tendencias y estacionalidades de los datos. Aunque se esperaba que el modelo RandomForestRegressor capturara mejor las complejidades no lineales, su desempeño fue inferior al de Exponential Smoothing y comparable al de ARIMA y Prophet.
- La construcción y transformación del conjunto de datos fue fundamental para obtener datos precisos y relevantes. El análisis de series temporales aplicado a los niveles de inventario mínimo y máximo, utilizando modelos como ARIMA, Prophet, Exponential Smoothing y Random Forest, demostró que el modelo Random Forest proporciona las mejores predicciones, con métricas de error

significativamente menores. Esto indica que el Random Forest es altamente eficaz para prever los niveles de inventario necesarios, ayudando a mantener un equilibrio óptimo entre la disponibilidad de productos y el control de costos.

5.2. RECOMENDACIONES

- Se recomienda la exploración y aplicación de otros algoritmos de clustering y técnicas de análisis de datos. Algoritmos como DBSCAN (Density-Based Spatial Clustering of Applications with Noise) y SOM (Self-Organizing Maps) podrían proporcionar perspectivas adicionales sobre la agrupación de productos al detectar estructuras no lineales y relaciones más complejas en los datos. La actualización continua del conjunto de datos y la reevaluación periódica de los resultados también son cruciales para mantener la relevancia y precisión de la segmentación en un entorno empresarial dinámico.
- Se recomienda explorar otros algoritmos de aprendizaje automático. Aunque RandomForestRegressor no superó en este caso a los modelos tradicionales, otros algoritmos como XGBoost o LSTM (Long Short-Term Memory) podrían ofrecer mejores resultados al capturar patrones más complejos en los datos. Además, para mejorar la precisión de las predicciones, se sugiere incorporar factores externos que puedan influir en las ventas, como promociones, eventos especiales, cambios en la economía, entre otros. Estos factores pueden ser integrados en los modelos para capturar variaciones no contempladas solo con datos históricos de ventas.
- Se sugiere explorar técnicas adicionales de machine learning y optimización que puedan complementar o mejorar los modelos actuales. Por ejemplo, técnicas de redes neuronales recurrentes (RNN) o modelos de redes neuronales convolucionales (CNN) podrían capturar mejor las relaciones temporales complejas presentes en los datos de inventario. Además, implementar técnicas de ensamble como *model ensembling* podría ayudar a mejorar aún más la precisión de las predicciones al combinar las fortalezas de diferentes modelos.

REFERENCIAS

Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers and Industrial Engineering*, *143*. https://doi.org/10.1016/j.cie.2020.106435

Aditya Satrio, C. B., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, *179*, 524–532. https://doi.org/10.1016/J.PROCS.2021.01.036

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, *07-09-Janu*, 1027–1035.

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

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140–1154. https://doi.org/10.1016/j.ejor.2006.12.004

Chakraborty, A., Dey, K., Ghosh, S., Chakraborty, R., Mitra, I., & Nandy, P. (2023). A Novel Approach for Customer Segmentation and Product Recommendation to Boost Sales using Machine Learning. *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, 97–103. https://ieeexplore.ieee.org/abstract/document/10112415

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Peter Campbell, J. (2020). Introduction to machine learning, neural networks, and deep learning.

Translational Vision Science and Technology, *9*(2), 1–12. https://doi.org/10.1167/tvst.9.2.14

G V S Abhishek Varma, G V N Akshay Varma, & Adiar Kethaan. (2023). Predictive Insights for Monthly Property Sales Forecasting: An End-to-End Time Series Forecasting. *International Journal of Advanced Research in Science, Communication and Technology*, 598–604. https://doi.org/10.48175/ijarsct-12493

Kiuchi, A., Wang, H., Wang, Q., Ogura, T., Nomoto, T., Gupta, C., Matsui, T., Serita, S., & Zhang, C. (2020). Bayesian Optimization Algorithm with Agent-based Supply Chain Simulator for Multi-echelon Inventory Management. *IEEE International Conference on Automation Science and Engineering*, *2020-Augus*, 418–425. https://doi.org/10.1109/CASE48305.2020.9216792

Kourentzes, N., Trapero, J. R., & Barrow, D. K. (2020). Optimising forecasting models for inventory planning. *International Journal of Production Economics*, *225*, 107597. https://doi.org/10.1016/J.IJPE.2019.107597

López Morales, J. S. (2021). Product Life Cycle. In *Encyclopedia of Sustainable Management* (pp. 1–5). Springer, Cham. https://doi.org/10.1007/978-3-030-02006-4_329-1

Onumanyi, A. J., Molokomme, D. N., Isaac, S. J., & Abu-Mahfouz, A. M. (2022). AutoElbow: An Automatic Elbow Detection Method for Estimating the Number of Clusters in a Dataset. *Applied Sciences (Switzerland)*, *12*(15). https://doi.org/10.3390/app12157515

Patel, P., Sivaiah, B., & Patel, R. (2022). Approaches for finding Optimal Number of Clusters using K-Means and Agglomerative Hierarchical Clustering Techniques.

2022 International Conference on Intelligent Controller and Computing for Smart Power, ICICCSP 2022. https://doi.org/10.1109/ICICCSP53532.2022.9862439

Paul, A., Ghosh, S., Das, A. K., Goswami, S., Das Choudhury, S., & Sen, S. (2020). A Review on Agricultural Advancement Based on Computer Vision and Machine Learning. *Advances in Intelligent Systems and Computing*, *937*, 567–581. https://doi.org/10.1007/978-981-13-7403-6_50

Plotnikova, V., Dumas, M., & Milani, F. (2020). Adaptations of data mining methodologies: A systematic literature review. *PeerJ Computer Science*, *6*, 1–43. https://doi.org/10.7717/PEERJ-CS.267

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, *134*, 93–101. https://doi.org/10.1016/J.ESWA.2019.05.028

Sri Kalli, N., & Pullagura, H. T. (2020). Predicting Total Business Sales using Time Series Analysis. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 475–482. https://doi.org/10.32628/cseit206485

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, *72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep Learning for Time Series Forecasting: A Survey. *Https://Home.Liebertpub.Com/Big*, *9*(1), 3–21. https://doi.org/10.1089/BIG.2020.0159

Vandeput, N. (2020). Inventory optimization: Models and simulations. In *Inventory Optimization: Models and Simulations*. DE GRUYTER.

Varkalys, M. (2021). *Machine learning based inventory optimization respecting supplier order line fees MINDAUGAS VARKALYS KTH ROYAL INSTITUTE OF TECHNOLOGY SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE*.

Vijesh Chaudhary, Kundurthi Bharadwaja, Radhey Shyam Meena, Dr. Purnendu Bikash Acharjee, Dr. Niraj C. Chaudhari, & Dr. R. Gopinathan. (2023). Exploring the Use of Machine Learning in Inventory Management for Increased Profitability. *A Journal for New Zealand Herpetology*, *12*(01), 658–659.

Villegas, M. A., Pedregal, D. J., & Trapero, J. R. (2018). A support vector machine for model selection in demand forecasting applications. *Computers and Industrial Engineering*, *121*, 1–7. https://doi.org/10.1016/j.cie.2018.04.042

Wan, T., & Hong, L. J. (2022). *Large-Scale Inventory Optimization: A Recurrent-Neural-Networks-Inspired Simulation Approach*. http://arxiv.org/abs/2201.05868

Xie, Y., Jin, M., Zou, Z., Xu, G., Feng, D., Liu, W., & Long, D. (2022). Real-Time Prediction of Docker Container Resource Load Based on a Hybrid Model of ARIMA and Triple Exponential Smoothing. *IEEE Transactions on Cloud Computing*, *10*(2), 1386–1401. https://doi.org/10.1109/TCC.2020.2989631

Yonar, H., Yonar, A., Tekindal, M. A., & Tekindal, M. (2020). Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods. *Eurasian Journal of Medicine and Oncology*, *4*(2), 160–165. https://doi.org/10.14744/ejmo.2020.28273

Zhang, Y., & Meng, G. (2023). Simulation of an Adaptive Model Based on AIC and BIC ARIMA Predictions. *Journal of Physics: Conference Series*, *2449*(1), 012027. https://doi.org/10.1088/1742-6596/2449/1/012027

ANEXOS

Anexo 1: Informe de COMPILATIO

	UNIVERSIDAD TÉCNICA ESTATAL QUEVEDO **FACULTAD DE POSGRADO**	

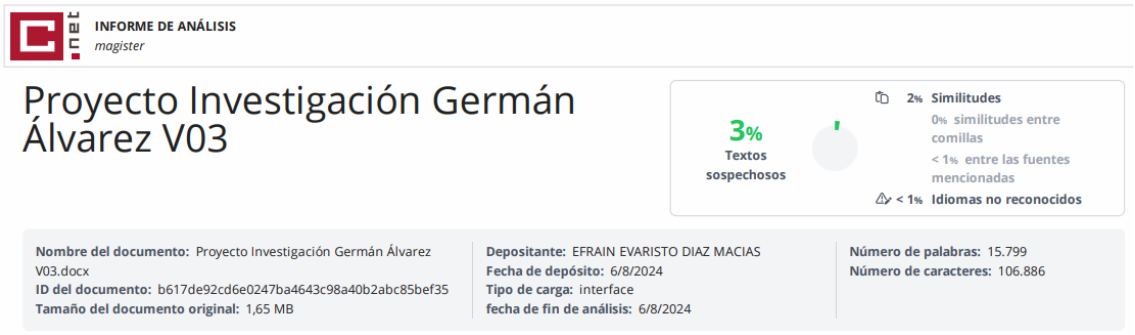
PARA: Ing. Byron Oviedo Bayas, PhD
Decano de Posgrado UTEQ

DE: Ing. Efraín Evaristo Díaz Macías, M. Sc.

ASUNTO: Informe Proyecto de Investigación

FECHA: Quevedo, 06 de agosto de 2024

Adjunto al presente sírvase encontrar el documento final del proyecto de investigación titulado: **Optimización de inventario en la gestión de la cadena de suministro en empresas de consumo masivo**, elaborado por el **Ing. Gilberto Germán Álvarez Carpio** posgradista de la **Maestría en Ciencia de Datos**. El Proyecto de investigación fue elaborado bajo mi dirección según lo asignado en el contrato PG-1239-2023-C-CIV-SERV-PROF de fecha 27 de noviembre de 2023, el mismo que cumple el informe de la herramienta COMPILATIO, el cual avala los niveles de originalidad, en un 97% del trabajo investigativo.



Atentamente,

Firmado electrónicamente por:
EFRAIN EVARISTO DIAZ MACIAS

Ing. Efraín Evaristo Díaz Macías, M. Sc.
DIRECTOR DE PROYECTO DE INVESTIGACIÓN