

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Modelo ensamblado de aprendizaje automático no supervisado
para la clasificación de inventarios

Andrés Felipe Toscano Serrano

Trabajo de investigación para optar al título de Magister en Ingeniería Industrial

Director

Carlos Eduardo Díaz Bohórquez

M.Sc. en Ingeniería Industrial

Codirector

Henry Lamos Díaz

Ph.D. en Física - Matemáticas

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Estudios Industriales y Empresariales

Maestría en Ingeniería Industrial

Bucaramanga

2026

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

*El presente trabajo investigativo está dedicado a Dios,
por ser el inspirador y artífice de permitirme culminar con éxitos
este proceso de formación profesional.*

*A mis padres Julio César (QEPD) y Lupita,
por su testimonio de amor, trabajo, sacrificio y apoyo incondicional;
gracias a ellos soy el hombre que soy.
Siempre será un orgullo y privilegio ser hijo de personas tan virtuosas.*

*A mi querida hermana Nicole Valentina,
por estar siempre presente, su luz y alegría han sido un motor a lo largo de mi vida.
Por enseñarme a valorar y ver el infinito desde lo finito de cada instante.*

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Agradecimientos

A Dios, por disponer de todo lo necesario en mi camino para cumplir con esta meta profesional.

A la memoria de mi padre y socio de vida, por sembrar en mí el valor del trabajo desde pequeño, motivándome a dar una mejor versión de mí todos los días; sin caer en la rutina que nos distrae y no nos permite saber y crecer en el amor.

A mi madre y colega ingeniera, por ser testimonio que el esfuerzo y el sacrificio tienen su mérito; ser referente y una de las razones que me motivaron a elegir la ingeniería industrial como proyecto de vida, mostrándome que se puede cambiar el mundo generando valor en todo lo que se emprende.

A mi hermana y futura médica, por ser un apoyo invaluable en este proceso, alegrando e iluminando los días más pesados, dándome la tarea y la motivación de ser su modelo a seguir, un hermano apasionado que no renuncia a sus sueños.

A la Universidad Industrial de Santander, donde me he formado como profesional integro, y a la Escuela de Estudios Industriales y Empresariales y todo el equipo docente, especialmente a Carlos Díaz y Henry Lamos, quienes han sido pilares fundamentales en mi aprendizaje, gracias por su apoyo, acompañamiento y orientación a lo largo de mi posgrado.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla de contenido

Introducción	12
1. Planteamiento y justificación del problema	16
1.1 Planteamiento del problema	16
1.2 Justificación	18
2. Objetivos	20
2.1 Objetivo general	20
2.2 Objetivos específicos	20
3. Propositiones	20
4. Metodología	21
4.1 Fase I: Contextualización de la investigación	22
4.2 Fase II: Formulación del modelo ensamblado	23
4.3 Fase III: Evaluación del modelo ensamblado	24
5. Revisión de literatura	26
5.1 Gestión y clasificación de inventarios	27
5.2 Enfoques tradicionales para la clasificación de inventarios	28
5.3 Enfoques de aprendizaje automático para la clasificación de inventarios	30
5.3.1 Aprendizaje supervisado	31
5.3.2 Aprendizaje no supervisado	33
5.3.3 Aprendizaje ensamblado	36
5.4 Discusión de la revisión e identificación de la brecha del conocimiento	39

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

6. Formulación del modelo	40
6.1 Obtención, pre-procesamiento y tratamiento de los datos	40
6.2 Análisis exploratorio de los datos	43
6.3 Selección variables de entrada del modelo	48
6.4 Selección de modelos base	50
6.5 Matriz de Costos	52
6.6 Selección técnica de ensamblaje	57
6.6.1 Selección basada en el desempeño de agrupación	58
6.6.2 Meta-modelo determinístico de Stacking	58
7. Evaluación del modelo	66
8. Conclusiones y Recomendaciones	78
Referencias	84
Apéndices	90

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Lista de tablas

Tabla 1. Algoritmos de aprendizaje no supervisado	35
Tabla 2. Revisión de literatura agrupada por enfoque de clasificación de inventarios	37
Tabla 3. Descripción Retailers	41
Tabla 4. Descripción variables y tipos de datos de los datasets	44
Tabla 5. Estadísticas Descriptivas Retailer 1	45
Tabla 6. Detección de outliers (z-score) Retailer 1	47
Tabla 7. Selección modelos base para la construcción del modelo ensamblado	51
Tabla 8. Tasa de servicio por clase	54
Tabla 9. Desempeño en términos de calidad de agrupamiento	67
Tabla 10. Desempeño en términos de calidad de agrupamiento aplicando PCA	68
Tabla 11. Desempeño en términos de costos	70
Tabla 12. Desempeño en términos de costos aplicando PCA	72
Tabla 13. Delta de costos con respecto al modelo tradicional ABC	74
Tabla 14. Delta de costos con respecto al modelo tradicional ABC aplicando PCA	75

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Lista de figuras

Figura 1. Metodología para el desarrollo del modelo ensamblado no supervisado	21
Figura 2. Fase I: Contextualización del problema	23
Figura 3. Fase II: Formulación del modelo ensamblado	24
Figura 4. Fase III: Evaluación del modelo ensamblado	26
Figura 5. Mapa de calor: Matriz de Correlación Retailer 1	47
Figura 6. Variables seleccionadas para el estudio	50
Figura 7. Flujo modelo ensamblado no supervisado para clasificación de inventarios	65

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Lista de ecuaciones

Ecuación 1. Desviación durante el lead time	53
Ecuación 2. Stock de seguridad	54
Ecuación 3. Costo de mantenimiento	55
Ecuación 4. Costo por faltantes	55
Ecuación 5. Costo total por clase	56
Ecuación 6. Regla de decisión en caso de desacuerdo	59

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Lista de Apéndices

Apéndice A. Recuperación de literatura y análisis bibliométrico	90
Apéndice B. Consultas SQL para la obtención de datos	100
Apéndice C. Análisis Exploratorio de los Datos (EDA)	105
Apéndice D. Evaluación consolidada entre modelos	150

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Resumen

Título: Modelo ensamblado de aprendizaje automático no supervisado para la clasificación de inventarios*

Autor: Andrés Felipe Toscano Serrano

Palabras Clave: Inventario, Clasificación, Aprendizaje Automático, No supervisado, Ensamblado.

Descripción:

La gestión eficiente de inventarios es un pilar estratégico en la cadena de suministro, donde la precisión en la clasificación de productos impacta directamente los costos operativos y el nivel de servicio. Los enfoques tradicionales, como el método ABC, presentan limitaciones ante la creciente complejidad del mercado, la diversificación de la demanda y la expansión del comercio electrónico. En este contexto, las técnicas de aprendizaje automático (ML) emergen como alternativas robustas para abordar estos desafíos.

Esta investigación propone un modelo ensamblado de aprendizaje automático no supervisado para la clasificación de inventarios, desarrollado a partir del análisis de datos de 15 retailers latinoamericanos. El modelo integra tres modelos base (K-means, clustering jerárquico y modelos de mezclas Gaussianas) seleccionados por su capacidad para capturar las características del problema. Las variables de entrada consideradas son: unidades demandadas, desviación estándar de la demanda, nivel promedio de stock y tiempo de reposición. El modelo adopta una estructura modular que permite priorizar criterios de clasificación según los objetivos organizacionales, ya sea en términos de calidad de agrupamiento o minimización de costos, con y sin reducción de dimensionalidad mediante PCA.

Los resultados demuestran que el modelo ensamblado basado en costos supera consistentemente al método ABC tradicional y a los modelos base individuales, generando ahorros totales de \$2,437,821.36 USD. Asimismo, se evidenció que la aplicación de PCA, aunque mejora los indicadores de calidad de clustering (Silhouette Score), puede eliminar información relevante y resultar contraproducente desde una perspectiva económica. El clustering jerárquico mostró mayor estabilidad ante la reducción de dimensionalidad, mientras que K-means y GMM presentaron mayor sensibilidad.

Los hallazgos fueron presentados en la conferencia ICPR2025 y aceptados para publicación en Springer, evidenciando el aporte significativo de esta investigación a la gestión moderna de inventarios.

* Trabajo de Grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Estudios Industriales y Empresariales. Director: Carlos Eduardo Díaz Bohórquez, Magister Ingeniería Industrial. Codirector: Henry Lamos Díaz, Ph.D en Física - Matemática

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Abstract

Title: Unsupervised machine learning ensemble model for inventory classification*

Author: Andrés Felipe Toscano Serrano

Key words: Inventory, Classification, Machine Learning, Unsupervised, Ensemble.

Description:

Efficient inventory management is a strategic pillar in the supply chain, where precision in product classification directly impacts operational costs and service levels. Traditional approaches, such as the ABC method, present limitations in the face of growing market complexity, demand diversification, and the expansion of e-commerce. In this context, machine learning (ML) techniques emerge as robust alternatives for addressing these challenges.

This research proposes an unsupervised machine learning ensemble model for inventory classification, developed from the analysis of data from 15 Latin American retailers. The model integrates three base models (K-means, hierarchical clustering, and Gaussian mixture models) selected for their ability to capture the characteristics of the problem. The input variables considered are: units demanded, standard deviation of demand, average stock level, and replenishment lead time. The model adopts a modular structure that allows classification criteria to be prioritized according to organizational objectives, whether in terms of clustering quality or cost minimization, with and without dimensionality reduction via PCA.

The results demonstrate that the cost-based ensemble model consistently outperforms the traditional ABC method and individual base models, generating total savings of \$2,437,821.36 USD. It was also found that the application of PCA, while improving clustering quality indicators (Silhouette Score), can eliminate relevant information and prove counterproductive from an economic perspective. Hierarchical clustering showed greater stability under dimensionality reduction, whereas K-means and GMM exhibited greater sensitivity.

The findings were presented at the ICPR2025 conference and accepted for publication in Springer, underscoring the significant contribution of this research to modern inventory management.

* Degree Work

** Facultad de Ingenierías Fisicomecánicas. Escuela de Estudios Industriales y Empresariales. Director: Carlos Eduardo Díaz Bohórquez, MSc. in Industrial Engineering. Codirector: Henry Lamos Díaz, Ph.D in Physics - Maths

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Introducción

Las empresas que buscan mantenerse al ritmo del avance tecnológico y la globalización deben ser capaces de gestionar sus existencias de manera eficiente para alcanzar alta calidad, mayor productividad y reducción de costos (Albayrak Ünal et al., 2023) . La gestión del inventario se ocupa principalmente de la planificación y el control del stock en una industria, y es un pilar clave de la gestión de la cadena de suministro, debido a que las estimaciones inexactas desvían los objetivos de nivel de servicio y crean costos adicionales como escasez, pérdida de ingresos o exceso de inventario; lo que conlleva a construir modelos con alta precisión para la planeación de la demanda, pieza importante para garantizar una gestión óptima de inventarios, con base a una correcta clasificación de los mismos (Partovi & Anandarajan, 2004).

Debido a la diversificación de las necesidades de los clientes, la incertidumbre y la competencia del mercado, así como el crecimiento exponencial del comercio electrónico, el desarrollo de tecnologías de inteligencia artificial (IA) está impulsando que las aplicaciones de gestión de inventarios evolucionen hacia procesos más robustos e inteligentes (Ogunfowora & Najjaran, 2023), es así que los métodos de aprendizaje automático (ML) y aprendizaje profundo (DL), que son subramas de la IA, desempeñan un papel importante en este sentido, y en los últimos años se ha evidenciado en numerosos estudios el interés en apostar por técnicas de ML. Estos métodos pueden identificar y analizar patrones rápidamente dentro de conjuntos de datos grandes y variados. Además, la gestión de inventarios aplicando técnicas de ML la convierte en un proceso robusto, eficiente y flexible, maximizando rentabilidad con menores costos, proporciona un mejor nivel de servicio a los clientes y más información contextual (Albayrak Ünal et al., 2023). No obstante, este tipo de algoritmos también tienen sus limitaciones, ya que en cualquier entorno de producción, la precisión y el tiempo de cálculo son igualmente

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

importantes; y así como se puede optar por un sistema con mayor precisión pero inaplicable en tiempo real, también un algoritmo sencillo puede carecer de precisión y no ajustarse bien a los datos.

En los casos mencionados, es necesario lograr un equilibrio entre precisión y tiempo de cálculo, y este puede alcanzarse aplicando técnicas de “Aprendizaje Ensamblado” (EL), las cuales ofrecen un enfoque poderoso para mejorar los modelos de aprendizaje automático, ya que combinan varios modelos, conocidos como “aprendices base”, cuyas predicciones se consolidan para obtener un resultado más preciso y robusto, lo que permite implementar modelos de alta precisión adecuados para las aplicaciones en tiempo real (Dixit, 2017).

En este contexto, los métodos de aprendizaje automático, particularmente los modelos ensamblados no supervisados, emergen como soluciones prometedoras para la clasificación de inventarios en situaciones donde la información etiquetada es escasa o inexistente; es por ello que la presente investigación tiene como objetivo desarrollar un modelo ensamblado de aprendizaje automático no supervisado para la clasificación de inventarios, con el fin de crear un sistema flexible, robusto e inteligente que permita categorizar los productos en función de múltiples características. Al integrar técnicas de aprendizaje ensamblado se busca optimizar los costos asociados a la gestión de inventarios y mejorar la toma de decisiones al proporcionar una clasificación más precisa y adaptable a las necesidades cambiantes de las organizaciones. Por consiguiente, no solo se aborda un vacío identificado en la literatura, sino también se desarrolla un modelo que ofrece una solución práctica y aplicable, que permita a las empresas mejorar sus prácticas de gestión de inventarios y optimizar sus costos en un entorno cada vez más competitivo.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Este documento se encuentra estructurado de la siguiente manera: en la primera sección se presenta el planteamiento del problema, seguido de los objetivos trazados para esta investigación. En la tercer sección se encuentran las proposiciones; posteriormente en las secciones cuatro y cinco, la metodología planteada y la revisión de la literatura respectivamente. En la sexta y séptima sección se expone la formulación y evaluación del modelo propuesto, y finalmente, en la última sección las conclusiones y recomendaciones para futuras investigaciones.

Para desarrollar el modelo ensamblado no supervisado de aprendizaje automático para la clasificación de inventarios se parte de la observación y el análisis objetivo de los fenómenos estudiados en 15 retailers de Latinoamérica y de una revisión exhaustiva de la literatura relacionada con la clasificación de inventarios, encontrando que existen enfoques tradicionales como el método ABC y el análisis basado en la toma de decisiones multicriterio (MCDM); y los enfoques de aprendizaje automático como: el aprendizaje supervisado (árboles de decisión, las máquinas de vectores de soporte y las redes neuronales profundas), aprendizaje no supervisado (K-means, Variational Autoencoders, Clustering jerárquico y modelos de mezclas Gaussianas) y aprendizaje ensamblado. Se formula el modelo desarrollado en lenguaje de programación Python y con base al análisis exploratorio de los datos se establecen las variables de entrada: unidades demandadas, desviación estándar de la demanda, el nivel promedio de stock y el tiempo de reposición en días; se determinan como modelos base: K-means, clustering jerárquico y modelos de mezclas Gaussianas al ser las alternativas que capturaban de mejor forma las características del problema, y buscando alinear el diseño con los objetivos operativos. El modelo sigue una estructura modular para poder priorizar el criterio de clasificación según sea la necesidad, ya sea en términos de calidad de agrupamiento de los clústeres o en términos de costos; como la aplicación, o no, de la reducción de dimensionalidad (PCA) en los datos.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Al evaluar el modelo propuesto para la clasificación de inventarios bajo las políticas de inventario de cada clase relacionadas con el nivel de servicio y su impacto asociado a los costos, se presentan como hallazgos relevantes que al contrastar los resultados del modelo ensamblado basados en costos con el escenario sin PCA, se observa una disminución del ahorro total en alrededor 2x, lo que sugiere que la aplicación de PCA puede eliminar información relevante para la correcta clasificación de los productos. Este patrón se repite en el ensamblado basado en el desempeño del clustering y en los modelos base, donde se identifican con mayor frecuencia deltas negativos, indicando casos en los que la clasificación resultante empeora el desempeño en costos con respecto al modelo tradicional ABC. En particular, K-Means y GMM muestran una mayor sensibilidad a la reducción de dimensionalidad, mientras que el clustering jerárquico mantiene un comportamiento relativamente más estable, aunque con ahorros 3x inferiores a los observados sin PCA. Se concluye que el modelo ensamblado no supervisado basado en costos para los 15 retailers bajo estudio presenta un mejor desempeño económico en comparación con la clasificación ABC tradicional al generar asignaciones que conducen a menores costos totales esperados bajo las políticas de nivel de servicio definidas. Desde la perspectiva de la calidad del agrupamiento, el modelo ensamblado basado en el desempeño del clustering y la aplicación de PCA conducen a incrementos sistemáticos en el Silhouette Score, reflejando una mejor definición estructural de los clústeres. Sin embargo, cuando el foco se traslada al desempeño en términos de costos, el ensamblado basado en costos destaca de forma consistente como la alternativa más eficiente, superando tanto a los modelos base como al enfoque tradicional ABC y generando los mayores ahorros absolutos (\$2,437,821.36 USD) y relativos. Asimismo, la comparación entre escenarios con y sin PCA pone en evidencia que la reducción de dimensionalidad, si bien beneficiosa para la calidad de agrupamiento del clustering, no siempre

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

se traduce en decisiones económicamente superiores e incluso puede resultar contraproducente en ciertos contextos.

Los resultados de esta investigación se han compartido a la comunidad científica mediante la presentación de las ponencias: “Techniques and machine learning models for inventory classification: A systematic Review” y “Unsupervised machine learning ensemble model for inventory classification” en la 28th International Conference on Production Research (ICPR2025). Resultado del desempeño en dicho evento, fueron aceptados los papers para los Proceedings en Springer de ambas ponencias. La revista indexada International Journal of Production Research (Q1) del Reino Unido extendió la invitación y aceptación de publicación del paper ampliado “Unsupervised machine learning ensemble model for inventory classification”. La divulgación de los productos de la presente investigación han contribuido al avance del conocimiento en la gestión de inventarios en el uso de técnicas avanzadas de análisis y aprendizaje automático en entornos reales, y ha evidenciado un aporte significativo a la comunidad científica.

1. Planteamiento y justificación del problema

1.1 Planteamiento del problema

En el entorno actual, caracterizado por una gran diversidad de productos y altas exigencias de los consumidores, mantener una clasificación adecuada del inventario es esencial para evitar costos asociados por una distribución ineficiente de los recursos de las empresas; es por ello, que la gestión eficiente de inventarios se ha convertido en un desafío crítico para las organizaciones que buscan optimizar sus operaciones y minimizar costos para responder de manera ágil, oportuna, eficiente y flexible al dinamismo del mercado.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Las compañías no siempre disponen de inventarios ya clasificados, ni cuentan con las capacidades para etiquetarlos manualmente, lo que restringe el uso de técnicas de aprendizaje automático supervisado, ya que estos modelos requieren datos históricos con inventarios previamente clasificados para poder aprender a categorizar nuevos productos. Además, los modelos supervisados se limitan a clasificar únicamente los productos nuevos, lo que deja fuera gran parte del catálogo de productos; esto es especialmente problemático para empresas que gestionan inventarios grandes y diversos, donde la dinámica de los productos puede variar considerablemente.

En este contexto, un enfoque no supervisado de aprendizaje automático, basado en un modelo ensamblado, ofrece una solución prometedora, debido a que un modelo no supervisado puede clasificar todos los productos del inventario en categorías específicas (por ejemplo, A, B y C) según características inherentes, sin necesidad de datos etiquetados previos. Esto permite una clasificación más completa y adaptable del catálogo, abarcando tanto productos nuevos como antiguos.

Así mismo, considerando que las políticas de inventario se definen a partir de niveles de servicio preestablecidos y que la clase asignada a cada producto determina el nivel de servicio aplicado, el impacto económico de la clasificación resulta necesariamente diferenciado entre productos. La implementación de un modelo ensamblado no supervisado para la clasificación de inventarios permite evaluar el desempeño en términos de costo de la clasificación al asociar niveles de servicio específicos a cada clase. Teniendo en cuenta que los productos clasificados en la clase A representan un mayor impacto económico para la empresa, se asocian a políticas de inventario caracterizadas por niveles de servicio más altos, mientras que los productos

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

clasificados en las clases B y C se asocian a políticas con niveles de servicio más flexibles. La aplicación de estas políticas implícitas de nivel de servicio, alineadas con la clasificación generada por el modelo, puede traducirse en una mejora significativa en la eficiencia operativa y una reducción de los costos generales de inventario.

1.2 Justificación

El desarrollo de este modelo ensamblado de aprendizaje automático no supervisado se justifica desde la adopción de tecnologías avanzadas para la gestión eficiente de los inventarios, dado el uso real de estos modelos de clasificación de inventarios en empresas con recursos limitados para etiquetar sus datos y que pueda abarcar la totalidad del catálogo de productos. Con este modelo ensamblado de aprendizaje automático no supervisado se asignará la categoría de los productos en función de los costos asociados a las políticas de nivel de servicio definidas para cada clase (A, B o C). El meta-modelo integrará las predicciones de los modelos base, evaluando el costo total de cada posible asignación con base a una regla determinista; de esta manera, se espera reducir los costos y por ende, mejorar la rentabilidad y sostenibilidad de las empresas.

Es por esto, que esta investigación puede resultar especialmente útil en el contexto de las pequeñas y medianas empresas, facilitando la transferencia de conocimientos y el avance tecnológico para ellas, logrando una mayor integración entre la teoría y la práctica, contribuyendo al fortalecimiento en la gestión y control de inventario, y a la reducción de costos de este tipo de empresas, que no cuentan con las tecnologías ni blandas ni duras para este tipo de desarrollos.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

En consecuencia, se busca responder la siguiente pregunta de investigación: ¿En qué medida los modelos ensamblados de aprendizaje automático no supervisado mejoran la calidad del agrupamiento y contribuyen a la reducción de costos en los procesos de clasificación de inventarios?

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

2. Objetivos

2.1 Objetivo general

Desarrollar un modelo ensamblado de aprendizaje automático no supervisado para la clasificación de inventarios buscando la reducción en sus costos.

2.2 Objetivos específicos

- Caracterizar el problema de clasificación de inventarios a partir de las técnicas y modelos implementados según lo recopilado en la revisión literaria.
- Formular un modelo ensamblado de aprendizaje automático no supervisado para la clasificación de inventarios que permita la reducción en sus costos.
- Evaluar el modelo propuesto para la clasificación de inventarios bajo las políticas de inventario de cada clase y su impacto asociado a los costos.

3. Propositiones

La presente investigación considera las siguientes proposiciones:

- Los modelos ensamblados no supervisados de aprendizaje automático tienen un mejor desempeño a nivel de calidad de agrupamiento para la clasificación de inventarios que los modelos base de aprendizaje automático de forma individual, en razón a que permite a las empresas optimizar sus recursos y mejorar la toma de decisiones.
- Los modelos ensamblados de aprendizaje automático no supervisados resultan útiles porque permiten analizar grandes volúmenes de datos de manera eficiente, clasificar los datos con base a múltiples variables, identificar patrones complejos y relaciones ocultas que son difíciles de detectar con métodos tradicionales, logrando reducir los costos y

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

mejorar la eficiencia de la cadena de suministro a partir de una clasificación más precisa y robusta.

4. Metodología

La presente investigación adopta un enfoque epistemológico positivista, que se centra en el modelo hipotético-deductivo con la obtención de conocimiento a partir de la observación y el análisis objetivo de los fenómenos estudiados, ya que el conocimiento cuando se desarrolla adecuadamente es congruente con la realidad y exacto (Park et al., 2020). Este enfoque busca garantizar que los resultados sean medibles y verificables, siguiendo principios científicos. La investigación se organiza en tres fases (Ver Figura 1), diseñadas de manera alineada y fundamentadas en los objetivos planteados en la segunda sección, asegurando así la coherencia entre los métodos utilizados y las metas esperadas.

Figura 1.

Metodología para el desarrollo del modelo ensamblado no supervisado



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

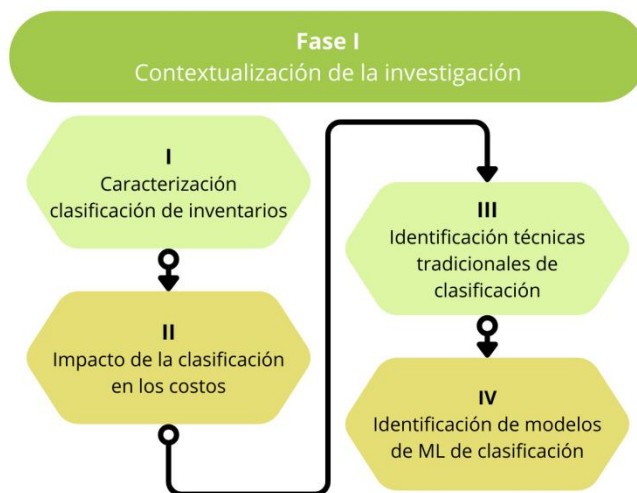
4.1 Fase I: Contextualización de la investigación

La primera fase, tiene como fin contextualizar la investigación mediante una revisión exhaustiva de la literatura existente, la cual se desarrolla bajo la metodología propuesta por Tranfield et al. (2003) . Se iniciará con la caracterización de la clasificación de inventarios, identificando los enfoques y criterios utilizados en distintas industrias para segmentar los productos en categorías específicas. Seguidamente, se analizará el impacto que la clasificación de inventarios tiene sobre los costos, lo cual permitirá justificar la necesidad de optimizar el proceso. En esta etapa también se revisan las técnicas tradicionales de clasificación, como el método ABC, para establecer una línea de referencia en cuanto a la eficiencia y precisión de los modelos. Además, se procede a identificar modelos de aprendizaje automático aplicables a la clasificación de inventarios, abarcando tanto enfoques supervisados como no supervisados, así como el uso de modelos ensamblados que combinan múltiples algoritmos para mejorar la capacidad y robustez de la clasificación. Adicional, en esta fase se realiza el análisis bibliométrico de la literatura recopilada, con el fin de identificar los principales autores, países y editoriales en esta área del conocimiento. A continuación, en la Figura 2 se detalla la secuencia de las actividades desarrolladas en esta primera fase de la investigación.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 2.

Fase I: Contextualización del problema



4.2 Fase II: Formulación del modelo ensamblado

Esta fase comienza con la obtención de datos, la cual consisten en la recopilación de datos reales de inventario por parte de retailers relevantes de Latinoamérica de sus canales digitales, con el fin de asegurar una amplia cobertura del problema de clasificación. Una vez recopilados, se lleva a cabo un preprocesamiento exhaustivo de los mismos para garantizar su calidad y consistencia, lo que incluye la limpieza y transformación de la información disponible.

Luego, se realiza un análisis exploratorio de los datos con el fin de identificar patrones, distribuciones y relaciones que puedan ser útiles para los pasos posteriores.

A continuación, se procede a la selección de las variables de entrada del modelo, utilizando técnicas que permitan identificar las características más relevantes para el proceso de clasificación; adicionalmente, se determinan los modelos base que formarán parte del modelo ensamblado, eligiendo aquellos que mejor capturen las características del problema y por su relevancia y aplicabilidad en la literatura. También se selecciona la técnica de ensamblaje más adecuada, considerando los objetivos del estudio y las características específicas de los datos. El

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

modelo se entrena con los datos preprocesados para optimizar su rendimiento. Cabe resaltar, que el paso a paso de esta fase es una adaptación del autor del marco KDD propuesto por Fayyad et al. (1996), dado su rigor metodológico y su enfoque orientado en la generación de conocimiento a partir de los datos. Así mismo, el modelo ensamblado se desarrolla en el lenguaje de programación Python debido a su amplia cobertura y uso en el ámbito de la inteligencia artificial y del aprendizaje automático. En la siguiente Figura se ilustra el paso a paso de esta fase.

Figura 3.

Fase II: Formulación del modelo ensamblado



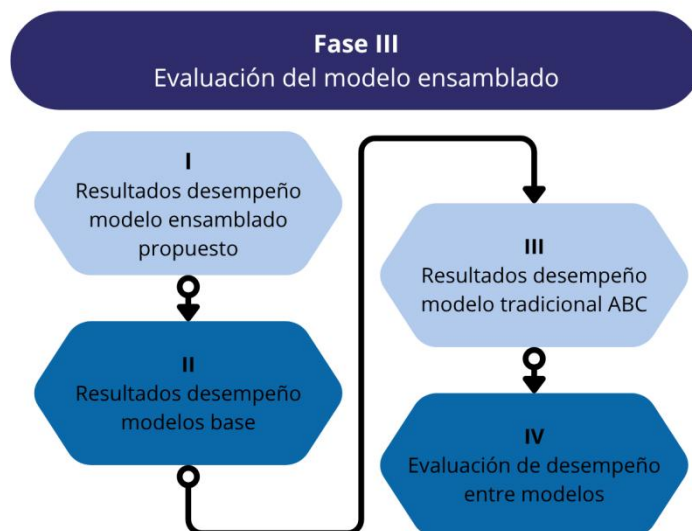
4.3 Fase III: Evaluación del modelo ensamblado

Finalmente, la tercera fase, denominada *Evaluación del modelo* (Ver Figura 4) está orientada a medir el desempeño interno del modelo ensamblado propuesto, así como la evaluación con otras alternativas en términos de los costos asociados a la clasificación de los inventarios.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Para el desempeño interno del modelo, se analiza teniendo en cuenta el índice de silueta, debido a que es la métrica más relevante para modelos no supervisados (Palacio-Niño & Berzal, 2019) ; con el fin de garantizar la precisión, en términos de calidad de agrupamiento, y la capacidad del modelo ensamblado para clasificar correctamente los inventarios. Además, se estudian los resultados de los modelos base de aprendizaje automático, que servirán como referencia para determinar la efectividad del enfoque ensamblado propuesto.

Ahora bien, también se realiza una comparación con métodos tradicionales como el modelo ABC en términos de los costos asociados a la política de inventario de cada categoría; en donde para la presente investigación, se entiende para el autor por política de inventario: “el conjunto de decisiones estratégicas asociadas al nivel de servicio objetivo de cada clase de inventario, las cuales determinan posteriormente el stock de seguridad requerido y los costos esperados de mantenimiento y por faltantes”, lo que permite establecer comparaciones directas y comprender mejor las ventajas y limitaciones de cada técnica. En conjunto, esta fase proporciona una evaluación integral que ayuda a validar la solución propuesta y su aplicabilidad en la clasificación de inventarios.

Figura 4.*Fase III: Evaluación del modelo ensamblado*

5. Revisión de literatura

Se presenta de manera organizada los resultados de la revisión exhaustiva de la literatura sobre la clasificación de inventarios, estructurados en cuatro secciones interrelacionadas. Primero, se abordan los fundamentos de la gestión y clasificación de inventarios, destacando su importancia para optimizar costos y mejorar la eficiencia operativa. Luego, se revisan los enfoques tradicionales, como el análisis ABC, que proporcionan una base clásica para la categorización de productos. A partir de ahí, la discusión avanza hacia los enfoques de aprendizaje automático, que ofrecen mejoras en la precisión y adaptabilidad frente a los métodos tradicionales, especialmente en el manejo de grandes volúmenes de datos y patrones complejos. Por último, en la discusión de la revisión, se sintetizan los hallazgos, comparando las ventajas y limitaciones de los enfoques estudiados, se identifica la brecha del conocimiento y se propone un nuevo modelo para la clasificación de inventarios. En el Apéndice A se consolida todo el proceso

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

de cribado junto con el análisis bibliométrico de la literatura recopilada, con el objetivo de poder caracterizar mejor el problema de investigación.

5.1 Gestión y clasificación de inventarios

La gestión de inventarios juega un papel fundamental en las operaciones de la cadena de suministro ya que se centra en supervisar y controlar eficazmente los bienes de una empresa para garantizar una producción fluida y la satisfacción de los clientes (Abidin et al., 2024); es por ello, que resulta clave, la clasificación de los inventarios debido al número de productos en existencias que se pueden llegar a gestionar, convirtiéndose en un aspecto inviable para las empresas a abordar la totalidad del volumen de artículos de forma individual. Por ejemplo, un minorista (retailer) promedio gestiona unos miles de unidades de mantenimiento de existencias (SKU), y en el caso de supermercados de grandes superficies pueden llegar a gestionarse más de 50.000 SKUs (Bala, 2012).

La clasificación de inventarios es el núcleo de la gestión de inventarios, aunque a primera vista, los inventarios corresponden a una gran cantidad de activos almacenados, estos le dan a la empresa el respaldo de la producción para aprovechar economías de escala o tener flexibilidad en cambios en la demanda (Lopez-Soto et al., 2016) ; por tanto, afecta significativamente el rendimiento de los costos, la reposición, las decisiones sobre el almacenamiento y el diseño del proceso de la cadena de suministro (Svoboda & Minner, 2022).

Una clasificación eficaz del inventario permite a las empresas racionalizar sus esfuerzos de control del inventario, priorizar la asignación de recursos; y, en última instancia, mejorar la rentabilidad global (Partovi & Anandarajan, 2004). Del mismo modo, resulta crítico evaluar y abordar las consecuencias que podría acarrear una clasificación errónea de los inventarios,

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

principalmente las implicaciones financieras y costos asociados con la gestión del inventario (Maitra & Kundu, 2023).

5.2 Enfoques tradicionales para la clasificación de inventarios

El análisis ABC, es una técnica tradicional de clasificación de inventarios ampliamente adoptado (Bailkar et al., 2024) en donde clasifica los artículos en función de una única variable, que en la mayoría de los casos corresponde al costo unitario o a las ventas de dicho producto. Este enfoque se basa en el principio de Pareto, según el cual aproximadamente el 80% de los efectos proceden del 20% de las causas. En la gestión de inventarios, esto se traduce en un pequeño porcentaje de artículos (Clase A) que representan una parte significativa del valor total del inventario (Kaabi et al., 2018).

La técnica ABC clasifica los artículos del inventario en tres categorías en función de un solo criterio, así:

- *Clase A*: representan el inventario más crítico, y suelen representar una gran parte del valor total del inventario, pero un pequeño porcentaje del volumen total del inventario.
- *Clase B*: producto que se sitúan en un punto intermedio, con un valor y un volumen moderados.
- *Clase C*: artículos que comprenden una gran porción del volumen total del inventario pero que contribuyen con una porción pequeña del criterio bajo estudio (Qaffas, Hajkacem, et al., 2023).

El análisis ABC tradicional, aunque es fácil de aplicar, tiene sus limitaciones, debido a que se basa exclusivamente en un único criterio y puede no reflejar con exactitud la verdadera importancia de los artículos, especialmente en sistemas de inventario complejos con diversas características de los artículos (Khanorkar & Kane, 2023).

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Ahora bien, bajo el mismo enfoque tradicional, se encuentran las técnicas de Toma de Decisiones Multicriterio (MCDM), las cuales ofrecen una perspectiva más matizada de la clasificación de inventarios al incorporar múltiples criterios al proceso de toma de decisiones, de este modo se abordan las limitaciones del análisis ABC tradicional, que se basa únicamente en un solo criterio, que usualmente es el costo unitario o las ventas de los productos (Khanorkar & Kane, 2023). Los métodos MCDM permiten una clasificación más precisa y exhaustiva al tener en cuenta diversos factores que influyen en las decisiones de gestión de inventarios, como el plazo de entrega, el costo unitario, la criticidad y la dificultad de aprovisionamiento (Mohammaditabar et al., 2012).

Por ejemplo, el modelo de A. Hadi-Vencheh (2010), demuestra una mayor precisión en comparación con el análisis ABC tradicional cuando se compara con la clasificación de expertos. Esta conclusión respalda la adopción de enfoques MCDM para la clasificación de inventarios. Sin embargo, las técnicas de MCDM pueden ser complejas desde el punto de vista informático y requerir la aportación subjetiva de los responsables de la toma de decisiones para la ponderación de los criterios (Khanorkar & Kane, 2023).

Por otra parte, autores como Abidin et al. (2024) presentan un enfoque multicriterio como ABC-HML para refinar las políticas de inventarios en función de múltiples variables como el valor de uso y la urgencia, haciendo énfasis en la importancia de una clasificación precisa en los procesos para la toma de decisiones.

Para abordar el problema de clasificación teniendo en cuenta tanto variables cuantitativas como cualitativas han surgido autores como Chu et al. (2008) y Mahendrawathi E. et al. (2011), los cuales implementaron la técnica ABC-Fuzzy la cual se base en la lógica difusa, incorporando la experiencia y el juicio de pares expertos, demostrando que captura mejor la intuición de la

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

importancia de cada producto comparado con el análisis tradicional ABC y brindando un grado de pertenencia de cada ítem a su respectiva clase.

Así mismo, también se han aplicado métodos como el Simple Aditivo Ponderado -SAW- (Sh. Hematyar, 2011) ; el Proceso de Jerarquía Analítica – AHP- para la clasificación de inventarios, con el fin de contemplar criterios objetivos (precio, demanda y volumen) y subjetivos (criticidad) en la toma de decisiones (Eraslan & Iç, 2020); y el Proceso de Jerarquía Analítica Difusa -FAHP-, en donde Shamsaddini et al. (2015) concluyen que la adición de criterios cualitativos al sistema de clasificación aumenta el número de artículos de clase A y B en comparación con el método tradicional ABC.

5.3 Enfoques de aprendizaje automático para la clasificación de inventarios

Como se ha mencionado previamente, la gestión de inventarios implica el control y manejo de cientos o incluso miles de artículos, lo cual hace inviable centrarse en cada uno por separado. Para enfrentar este desafío, los gerentes están recurriendo cada vez más al aprendizaje automático y a métodos basados en datos para clasificar automáticamente los artículos, materias primas y productos en distintas categorías, esto permite gestionar eficazmente cada grupo del inventario y optimizar los procesos operativos (Qaffas, Ben HajKacem, et al., 2023).

En los últimos años, la investigación ha comenzado a explorar el potencial de combinar la gestión de inventarios con algoritmos de aprendizaje automático para desarrollar métodos de gestión más eficientes y flexibles. En el contexto de la clasificación de inventarios, estas técnicas pueden analizar los datos disponibles y categorizar los productos, proporcionando un enfoque sistemático para la toma de decisiones (García-Barrios et al., 2021).

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Así mismo, los estudios han demostrado que los algoritmos de aprendizaje automático son altamente aplicables a los problemas de clasificación de inventarios. Al evaluarlos bajo diferentes medidas, se ha constatado que son eficaces para identificar patrones y clasificar artículos con precisión (Kartal et al., 2016) ; además, comparaciones realizadas con técnicas tradicionales han evidenciado que los enfoques basados en inteligencia artificial alcanzan niveles de precisión superiores en la clasificación de inventarios (Yu, 2011).

En conjunto, estas evidencias respaldan el uso del aprendizaje automático como una alternativa moderna y efectiva para la gestión de inventarios, facilitando la optimización de recursos y la toma de decisiones informadas. A continuación se detallan los diferentes tipos de aprendizaje supervisado, no supervisado y ensamblado que se recopiló en la literatura.

5.3.1 Aprendizaje supervisado

Los modelos de aprendizaje automático supervisado en la clasificación de inventarios utilizan datos etiquetados, en los que los artículos ya están categorizados en función de criterios predefinidos (Lolli et al., 2019) , y los modelos aprenden de estos datos etiquetados para clasificar artículos de inventario nuevos y desconocidos (Lolli et al., 2017) . Este enfoque aprovecha la potencia de algoritmos como los árboles de decisión, las máquinas de vectores de soporte (SVM) y las redes neuronales profundas (DNN) para descubrir relaciones complejas entre diversas características de los artículos y sus clases correspondientes.

- Los *Árboles de Decisión* son particularmente valiosos en la clasificación de inventarios debido a su interpretabilidad, proporcionando una visión clara de la lógica de clasificación a través de reglas IF-THEN, los autores Svoboda & Minner (2022)

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

proponen esta alternativa de modelo asociado a la minimización de los costos operativos y considerando restricciones generales de la demanda.

- Las *Redes de Retropropagación (BPN)* son la técnica de clasificación más utilizada para entrenar una red neuronal artificial, ya que a través de una arquitectura de retroalimentación permite realizar funciones complejas, como reconocimiento de patrones, predicción y clasificación de inventarios (Yu, 2011).
- *Máquinas de vectores de soporte (SVM)* sobresalen en la clasificación de inventarios basados en varios factores mediante la creación de un límite de decisión que separa las diferentes clases (Bailkar et al., 2024); obtienen una solución global única y son menos propensos a sobreajuste cuando se utilizan parámetros bien definidos (Kartal et al., 2016).
- Las *Redes Neuronales Profundas (DNN)*, conocidas por su eficiencia en el manejo de grandes conjuntos de datos, ya que pueden capturar patrones complejos y relaciones dentro de los datos de inventario para una clasificación precisa (Lolli et al., 2019).
- Las *Redes Neuronales Convolucionales (CNN)*, destacan por capturar y aprender de las dependencias internas y las características estructurales de los datos de inventarios, haciendo que sean muy utilizadas en contextos del mundo real (Weng et al., 2022).

Aprovechando estas técnicas de aprendizaje supervisado, las empresas pueden desarrollar sistemas de clasificación de inventario precisos y eficientes que tengan en cuenta múltiples criterios y se adapten a la dinámica cambiante del inventario. No obstante, se identifican dos limitaciones: (i) los inventarios deben contar con una clasificación para que este tipo de modelos aprenda a predecir esta etiqueta o grupo; (ii) y al ya tener un inventario previamente clasificado,

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

la aplicación de estas técnicas solo se reduce a los productos o ítems nuevos que ingresen al catálogo.

5.3.2 Aprendizaje no supervisado

El aprendizaje no supervisado ofrece un potente enfoque para la clasificación de inventarios cuando no se dispone de categorías o etiquetas predefinidas, ya que descubren patrones ocultos y agrupan los artículos similares en función de sus características inherentes (Jackson et al., 2019).

La agrupación de K-means, es un popular algoritmo de aprendizaje no supervisado que agrupa los artículos de inventario en clústeres en función de su similitud a través de múltiples características. La eficacia de este método es evidente en los estudios en los que demuestra resultados prometedores en la clasificación de inventarios con conjuntos de datos de gran tamaño, logrando una alta calidad de agrupación en comparación con otros métodos de clasificación (Khanorkar & Kane, 2023).

Del mismo modo, Qaffas et al. (2023) propusieron un enfoque de Inteligencia Artificial Explicable (XAI) para la clasificación ABC de múltiples criterios en dos fases, una primera fase destinada a la clasificación por medio del Algoritmo K-means; y una segunda fase, correspondiente a generar explicaciones micro y macro del modelo, proporcionando un esquema sencillo de la contribución de cada criterio en la definición de las clases de inventario.

De manera análoga, Jackson et al. (2019) identificaron que la aplicación de Análisis de Componentes Principales (PCA) aportaba una mejora adicional a los Modelos de K-means en la segmentación de los productos, lo cual se evidenciaba en las pruebas de validez interna. Incluso, Michalski et al. (2021) utilizaron únicamente el método de PCA para la clasificación de ítems, con el fin de evaluar la importancia de los criterios seleccionados para la segmentación.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Y si bien la aplicación de la agrupación en clústeres para la segmentación de SKUs es común, los estudios previos no abordan el problema de la parametrización y el ajuste fino de los parámetros del modelo, Jackson (2022) desarrolló un modelo automatizado para la segmentación de SKU, incorporando el algoritmo de K-means de pequeños lotes con PCA y búsqueda en cuadrícula para el ajuste óptimo de parámetros, con el fin de obtener buenos resultados ahorrando tiempo en el proceso de ajuste del modelo.

Otro ejemplo de este tipo de aprendizaje, es el Variational Autoencoders (VAEs), en el cual se evidencia que presenta un rendimiento superior en comparación con técnicas supervisadas para datos desequilibrados, ofreciendo un gran potencial para mitigar el efecto látigo y mejorar la eficiencia operativa en las cadenas de suministro (Maitra & Kundu, 2023).

Finalmente, los modelos de mezclas gaussianas (GMM) aborda las limitaciones del análisis ABC tradicional mediante la consideración de múltiples criterios sin depender de la ponderación, lo que significa que el algoritmo no implica subjetividad (Zowid et al., 2019).

Los algoritmos de aprendizaje no supervisado resultan valiosos para la clasificación de inventarios, especialmente cuando se trata de grandes conjuntos de datos en los que no resulta práctico definir manualmente las categorías (Jackson et al., 2019). Al aprovechar estas técnicas, las empresas pueden obtener información basada en datos sobre la estructura de su inventario, lo que les permite optimizar las estrategias de gestión de inventarios y mejorar la eficiencia general.

En la Tabla 1 se encuentra un consolidado de los algoritmos no supervisados de aprendizaje automático que se han identificado en la literatura: K-means, K-means + XAI, Análisis de Componentes Principales (PCA), K-means con PCA y Ajuste Automatizado, Variational Autoencoders (VAEs), y Modelos de Mezclas Gaussianas (GMM); y de los cuales se detalla de una manera breve la descripción, los beneficios, limitaciones, y sus respectivos autores.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 1.*Algoritmos de aprendizaje no supervisado*

Algoritmo no supervisado	Descripción	Puntos Clave	Autores destacados
K-means	Agrupación de artículos de inventario en clústeres según similitud en múltiples características.	Beneficios: Alta precisión en grandes conjuntos de datos, eficaz para clasificar inventarios. Limitaciones: Necesita ajustar parámetros, puede ser sensible a la inicialización.	Jackson et al. (2019); Khanorkar & Kane (2023).
K-means + XAI	Enfoque de dos fases: clasificación mediante K-means y explicaciones micro y macro para evaluar criterios de clasificación ABC.	Beneficios: Proporciona transparencia e interpretabilidad en la clasificación de inventarios. Limitaciones: Requiere un enfoque en dos fases, por lo cual puede ser más complejo de implementar.	Qaffas, et al. (2023).
Análisis de Componentes Principales (PCA)	Reducción de dimensionalidad que mejora la segmentación en inventarios al resaltar criterios importantes para la clasificación.	Beneficios: Mejora la precisión y validez interna de la segmentación. Limitaciones: No clasifica por sí solo, necesita combinarse con otros métodos para mejores resultados.	Jackson et al. (2019); Michalski et al. (2021).
K-means con PCA y Ajuste Automatizado	Segmentación de SKUs con K-means de pequeños lotes, PCA y ajuste de parámetros mediante búsqueda en cuadrícula.	Beneficios: Mayor precisión y eficiencia en el ajuste de parámetros, reduce el tiempo de parametrización. Limitaciones: Puede requerir mayor poder computacional para ajustar automáticamente el modelo.	Jackson (2022).
Variational Autoencoders (VAEs)	Modelo de aprendizaje profundo que mejora la clasificación de inventarios para datos desequilibrados y ayuda a mitigar el efecto látigo en la cadena de suministro.	Beneficios: Eficiencia en datos desequilibrados, útil para mejorar la precisión y reducir efectos negativos en la cadena de suministro. Limitaciones: Mayor complejidad computacional comparado con modelos tradicionales.	Maitra & Kundu (2023).
Modelos de Mezclas Gaussianas (GMM)	Agrupación que clasifica inventarios considerando múltiples criterios, sin requerir ponderación y eliminando la subjetividad en la clasificación.	Beneficios: Reduce la subjetividad y facilita la clasificación de inventarios complejos. Limitaciones: Requiere buen ajuste de parámetros y puede ser computacionalmente intensivo en grandes conjuntos de datos.	Zowid et al. (2019).

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

5.3.3 Aprendizaje ensamblado

El aprendizaje ensamblado consiste en combinar múltiples técnicas de aprendizaje automático o integran el aprendizaje automático con otros métodos analíticos para mejorar la precisión y solidez de la clasificación (Aktepe et al., 2018). Estos modelos aprovechan los puntos fuertes de las distintas técnicas para superar las limitaciones de los métodos individuales, lo que conduce a resultados de clasificación más precisos y fiables (Seyedan et al., 2022).

Un ejemplo de aprendizaje ensamblado es la integración de algoritmos de clustering con técnicas de aprendizaje ensamblado. La combinación de los resultados de múltiples algoritmos de agrupación mediante un esquema de votación por mayoría puede mejorar la precisión y la estabilidad de la segmentación de clientes, un paso crucial para la previsión eficaz de la demanda y la gestión de inventarios (Seyedan et al., 2022).

Además, un modelo de aprendizaje ensamblado es el bayesiano promediados (BMA), el cual resulta eficaz para combinar las previsiones generadas por distintos modelos de series temporales, lo que da lugar a predicciones más precisas y sólidas. BMA lo consigue asignando pesos a los modelos individuales en función de su rendimiento, dando mayor peso a los modelos más precisos, lo que da como resultado una predicción agregada final que supera las predicciones de los modelos individuales (Seyedan et al., 2022).

Otro enfoque consiste en integrar el aprendizaje automático con técnicas de optimización, las fuentes destacan el uso del recocido simulado como técnica de optimización metaheurística (Liu et al., 2016), junto con el análisis de agrupación para determinar la clasificación óptima de los artículos en diferentes clases de inventario. Este enfoque integrado pretende minimizar la disimilitud entre los artículos de cada grupo y optimizar al mismo tiempo la política general de

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

control de inventarios de cada clase, lo que conduce a una gestión de inventarios más rentable y eficiente.

Los modelos de aprendizaje ensamblado ofrecen un potente enfoque para la clasificación de inventarios al combinar los puntos fuertes de diferentes técnicas analíticas. Este enfoque conduce a sistemas de clasificación más precisos, robustos y adaptables que pueden manejar con eficacia las complejidades de la gestión moderna de inventarios. Al aprovechar estos modelos avanzados, las empresas pueden obtener una ventaja competitiva optimizando el control de inventarios, reduciendo costos y mejorando la eficiencia operativa general.

A continuación, se presenta un consolidado de la revisión de literatura agrupado por enfoque de clasificación de inventarios, junto con una breve descripción; ventajas y limitaciones y autores destacados (Ver Tabla 2).

Tabla 2.

Revisión de literatura agrupada por enfoque de clasificación de inventarios

Enfoque	Descripción	Puntos Clave	Autores destacados
Tradicional: Análisis ABC	Clasificación de inventarios en tres categorías: A (artículos de mayor valor o impacto), B (intermedios), y C (menor valor). Se basa en el principio de Pareto (80/20).	Beneficios: Fácil de implementar, enfoque en los productos más valiosos. Limitaciones: No considera múltiples criterios como rotación o criticidad del ítem, solo se enfoca en el valor de inventario.	Bailkar et al. (2024); Kaabi et al. (2018).
Tradicional: Toma de decisiones multicriterio (MCDM)	Uso de varios criterios como valor, criticidad, tiempo de reposición, etc., para clasificar los inventarios. Técnicas como AHP o TOPSIS ayudan a priorizar decisiones.	Beneficios: Considera múltiples factores, más flexible y robusto. Limitaciones: Requiere más tiempo y datos para implementar, puede ser complejo de interpretar si se manejan demasiados criterios.	Khanorkar & Kane (2023); Mohammaditabar et al. (2012).
Aprendizaje automático: Supervisado	Técnicas de aprendizaje automático donde un modelo es entrenado con datos etiquetados (inventario previamente clasificado) para predecir categorías futuras.	Beneficios: Alta precisión si hay datos etiquetados de calidad, aprende patrones complejos. Limitaciones: Requiere un conjunto de datos etiquetado y validación, puede ser costoso generar suficientes datos de entrenamiento.	Bailkar et al. (2024); Lolli et al. (2019); Lolli et al. (2017).

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Enfoque	Descripción	Puntos Clave	Autores destacados
Aprendizaje automático: No supervisado	Algoritmos de agrupamiento como K-means o clustering jerárquico que clasifican inventarios sin etiquetas previas, encontrando patrones en los datos no etiquetados.	Beneficios: Útil cuando no se dispone de datos etiquetados, puede descubrir patrones y nuevas categorías de inventario. Limitaciones: La interpretación de los resultados puede ser subjetiva, especialmente si los grupos son difíciles de identificar claramente.	Khanorkar & Kane (2023); Jackson et al. (2019); Zowid et al. (2019).

En resumen, en el enfoque tradicional, se destaca el método ABC, que organiza los inventarios en tres categorías: A, B y C. Esta clasificación se basa en el principio de Pareto, donde aproximadamente el 80% del valor del inventario proviene del 20% de los artículos (categoría A). Este método es fácil de implementar y enfoca la gestión en los productos más valiosos, aunque tiene limitaciones al no considerar otros factores como la rotación o la criticidad del ítem. Otra técnica tradicional es el análisis basado en la toma de decisiones multicriterio (MCDM), que incorpora múltiples factores como el valor, la criticidad y el tiempo de reposición. Este enfoque es más flexible y robusto, utilizando métodos como AHP o TOPSIS para priorizar decisiones, aunque requiere un mayor esfuerzo en la recolección de datos y puede ser complejo de implementar por su interpretabilidad y subjetividad.

Por otro lado, los enfoques de aprendizaje automático ofrecen alternativas más modernas. En el aprendizaje supervisado, se entrenan modelos utilizando datos etiquetados para predecir categorías de inventario futuras, logrando alta precisión si los datos de entrenamiento son de calidad. Sin embargo, este enfoque requiere una cantidad significativa de datos etiquetados y un proceso de validación adecuado. El aprendizaje no supervisado, en cambio, utiliza algoritmos de agrupamiento como K-means para identificar patrones sin etiquetas previas, lo que permite descubrir nuevas categorías y patrones en el inventario. Aunque es útil cuando no se dispone de

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

datos etiquetados, la interpretación de los resultados puede llegar a ser subjetiva especialmente cuando los grupos no se identifican claramente.

5.4 Discusión de la revisión e identificación de la brecha del conocimiento

La clasificación de inventarios es un aspecto fundamental para la gestión eficiente de los mismos, ya que impacta significativamente en la rentabilidad de las empresas al representar una gran parte de sus activos. Sin embargo, la revisión de la literatura revela que son pocos los estudios que han explorado el uso de algoritmos de aprendizaje automático no supervisados para esta tarea. De hecho, solo se han identificado tres investigaciones que aplican este enfoque, cuyos autores son: Jackson et al. (2019), Zowid et al. (2019) y Khanorkar & Kane (2023). Esto resulta especialmente relevante considerando que los modelos no supervisados son una herramienta clave cuando no se dispone de un etiquetado previo de los productos, lo cual es común en el mundo real, particularmente para pequeñas y medianas empresas -pymes- (Zowid et al., 2019).

En el contexto actual, marcado por la globalización y el dinamismo del entorno empresarial, las organizaciones requieren de planes estratégicos y tecnologías avanzadas para la clasificación precisa de sus inventarios (Zowid et al., 2019). Sin embargo, existen barreras financieras que limitan el acceso a estas tecnologías para muchas empresas, especialmente las pymes. Abidin et al. (2024) sugieren que el uso de software de código abierto podría ser una solución viable para implementar nuevas tecnologías sin incurrir en costos elevados.

La incertidumbre en la demanda y la oferta representan un desafío constante para las cadenas de suministro, dificultando el mantenimiento de niveles óptimos de inventario, esta

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

situación puede llevar a la pérdida de ventas o al exceso de inventario, lo que consume hasta un tercio del presupuesto anual de una organización (Rehmani et al., 2021). Además, el análisis independiente de los productos impide aprovechar oportunidades de consolidación y compartir costos, mientras que un enfoque individualizado resulta en una gestión ineficiente del tiempo (García-Barrios et al., 2021).

Finalmente, aunque la inteligencia artificial (IA) ha sido utilizada desde hace décadas, aún no ha alcanzado su máximo potencial en la gestión de la cadena de suministro (Haoud & Bachiri, 2019); así mismo, la literatura sugiere que el uso de modelos ensamblados de aprendizaje automático puede mitigar algunas limitaciones inherentes a los modelos base (Dixit, 2017); no obstante, hasta el momento no se han aplicado modelos ensamblados no supervisados en el área de clasificación de inventarios, evidenciando un espacio que favorece la investigación actual y una oportunidad para futuros estudios.

En conjunto, estos hallazgos indican una brecha en la aplicación de técnicas de aprendizaje automático en la clasificación de inventarios; por tanto, una oportunidad de investigación destacando la necesidad de explorar nuevas metodologías para mejorar la clasificación de inventarios bajo el potencial del enfoque de un modelo ensamblado de aprendizaje automático no supervisado en esta área del conocimiento.

6. Formulación del modelo

6.1 Obtención, pre-procesamiento y tratamiento de los datos

Los datos utilizados para el desarrollo del modelo fueron recolectados por una fuente clasificada, con acceso controlado, y si bien la identificación de dicha fuente no puede hacerse pública por razones de confidencialidad, los registros fueron sometidos a procedimientos de

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

validación interna, garantizando su fiabilidad y rigor metodológico. Los datos corresponden al catálogo de productos de 15 retailers en sus canales de e-commerce, pertenecientes al segmento de Supermercados y Farmacias de cinco países de Latinoamérica. Se seleccionaron tres retailers relevantes por país, distribuidos según el tipo de retailer: pequeño (Small), mediano (Mid) y grande (Large), como se observa en la Tabla 3.

Tabla 3.

Descripción Retailers

ID	País	Subvertical	Tipo de Retailer	# SKUs
1	Colombia	Farmacia	Large	8066
2	Colombia	Supermercado	Mid	6551
3	Colombia	Farmacia	Small	4195
4	Uruguay	Supermercado	Large	2718
5	Uruguay	Supermercado	Mid	1450
6	Uruguay	Supermercado	Small	1201
7	Brasil	Supermercado	Large	6698
8	Brasil	Supermercado	Mid	5870
9	Brasil	Supermercado	Small	3050
10	México	Supermercado	Large	3880
11	México	Supermercado	Mid	3161
12	México	Supermercado	Small	1467
13	Argentina	Supermercado	Large	11693
14	Argentina	Farmacia	Mid	5557
15	Argentina	Supermercado	Small	3512

Junto con la información del catálogo de cada retailer, se obtuvo información adicional relacionada con costos y variables de interés, que sirven como inputs para el modelo propuesto. Para el presente estudio, se seleccionaron productos del catálogo de cada retailer que cumplieran con las siguientes condiciones: (i) ser productos CPGs (Consumer Packaged Goods o bienes de

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

consumo envasados); (ii) no ser productos pesables; y (iii) haber estado publicados y con ventas activas en los últimos 90 días.

La obtención de los datos se realizó mediante consultas SQL (Structured Query Language o Lenguaje de Consulta Estructurado) sobre la base de datos de e-commerce. Tanto el pre-procesamiento como el tratamiento de los datos se efectuaron desde la misma consulta SQL, con el objetivo de garantizar la limpieza, consistencia y trazabilidad de la información desde su origen.

En la primera versión de la consulta, se obtuvieron los datos básicos del catálogo como: órdenes, ventas, unidades vendidas, órdenes canceladas por faltantes, ventas perdidas, unidades demandadas, niveles promedio y medianos de inventario, y días promedio de reposición en una ventana de 90 días.

Posteriormente, para integrar la matriz de costos en el meta-modelo del modelo ensamblado, se desarrolló una segunda versión de la consulta, ampliando la ventana temporal a un año, lo cual permitió estimar variables de largo plazo como: la demanda promedio diaria, la desviación estándar de la demanda, el costo unitario, el costo unitario por faltantes y la desviación estándar del periodo de reposición.

Las consultas SQL empleadas para la obtención de los datos se presentan en el Apéndice B. De este modo, el proceso de obtención, pre-procesamiento y tratamiento de los datos permitió construir una base sólida, limpia y consistente, que refleja con precisión el comportamiento operativo y comercial de los distintos retailers analizados. La aplicación de reglas claras de validación y control de calidad asegura la confiabilidad de los insumos utilizados (datos reales y confiables), reduciendo el riesgo de inconsistencias en las etapas posteriores del análisis y

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

garantizando un punto de partida robusto, lo cual es fundamental para el entrenamiento del modelo propuesto.

6.2 Análisis exploratorio de los datos

El análisis exploratorio de datos (EDA, Exploratory Data Analysis) que se realizó, busca como objetivo comprender la estructura interna de los datos, identificar patrones relevantes, evaluar relaciones entre variables y detectar posibles redundancias o dependencias antes de aplicar el modelo ensamblado no supervisado propuesto. Dado que el estudio abarca múltiples retailers, el EDA se diseñó de forma automatizada y reproducible, permitiendo aplicar el mismo procedimiento analítico a cada conjunto de datos de manera consistente, mediante una plantilla desarrollada en Jupyter Notebook, la cual itera sobre todos los conjuntos de datos (datasets) disponibles en el directorio correspondiente.

Para cada conjunto de datos se generaron de forma sistemática: estadísticas descriptivas básicas, análisis univariado y análisis correlacional. Cabe resaltar que en el presente EDA se excluyó el análisis de valores faltantes en razón que en la sección previa desde la extracción y pre-procesamiento de los datos se hizo el manejo adecuado para no tener datos faltantes.

El análisis exploratorio se realizó sobre el conjunto de variables estructurales, las cuales son: ventas, ordenes, unidades demandadas y vendidas; nivel promedio y mediano de stock; ordenes canceladas y unidades faltantes por stockout; tiempo de reposición (días) y desviación estándar de la demanda. El conjunto de variables numéricas mencionadas son de tipo entero de 64 bits (int64) o de punto flotante de doble precisión (float64), de acuerdo con su naturaleza discreta o continua; como se observa en la siguiente Tabla.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 4.

Descripción variables y tipos de datos de los datasets

Nombre Variable	Descripción Variable	Tipo de dato	Descripción tipo de dato
UNITS	Unidades vendidas	int64	Número entero (64 bits)
ORDERS	Ordenes	int64	Número entero (64 bits)
SALES	Ventas	int64	Número entero (64 bits)
UNITS_STOCKOUT	Unidades faltantes por stockout	int64	Número entero (64 bits)
ORDERS_STOCKOUT	Ordenes canceladas por stockout	int64	Número entero (64 bits)
UNITS_DEMAND	Unidades demandadas	int64	Número entero (64 bits)
STDDEV_DEMAND_FINAL	Desviación estándar de la demanda	float64	Número real con decimales (64 bits)
AVG_STOCK_LEVEL	Promedio del nivel de stock	float64	Número real con decimales (64 bits)
MEDIAN_STOCK_LEVEL	Mediana del nivel de stock	float64	Número real con decimales (64 bits)
LEAD_TIME_DAYS	Tiempo promedio de reposición en días	float64	Número real con decimales (64 bits)

Ahora bien, por razones de claridad y extensión, en el presente capítulo se sintetizan los resultados más relevantes del análisis exploratorio de datos correspondientes a un retailer ($id = 1$), en donde se efectuó un análisis univariado orientado a caracterizar la distribución individual de cada variable, calculando medidas de tendencia central y de dispersión, con el objetivo de evidenciar patrones y posibles sesgos, como se observa en la Tabla 5.

Se identifica una marcada heterogeneidad en el comportamiento de las variables asociadas a la demanda y al inventario. Las variables de volumen, tales como UNITS, ORDERS y UNITS_DEMAND, presentan valores medios elevados acompañados de desviaciones estándar significativamente altas, lo que indica una distribución altamente dispersa y la presencia de observaciones extremas. Este comportamiento es consistente con un entorno de inventarios caracterizado por una demanda irregular, en la cual coexisten productos de baja rotación con otros de consumo considerablemente alto. Asimismo, la diferencia entre los valores promedios y las medianas (percentil 50%) evidencia distribuciones asimétricas sesgadas a la derecha, lo cual

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

indica que la mayoría de los productos tienen una demanda/volumen bajo y hay un número reducido de productos que presentan volúmenes muy altos (productos top sellers), comportamiento común en la dinámica de Retail.

Tabla 5.

Estadísticas Descriptivas Retailer 1

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	8066	470.08	1226.93	1	61	169	423.75	34693
ORDERS	8066	411.69	909.80	2	64	168	403	24940
SALES	8066	2668.29	6876.38	1	369	1083.5	2810	426428
UNITS_STOCKOUT	8066	27.71	82.23	1	2	7	20	2000
ORDERS_STOCKOUT	8066	411.69	909.80	2	64	168	403	24940
UNITS_DEMAND	8066	497.80	1288.50	2	66	179	449	36001
STDDEV_DEMAND_FINAL	8066	0.41	0.38	0.001	0.22	0.30	0.47	9.03
AVG_STOCK_LEVEL	8066	6.18	9.03	1	2.77	3.8	5.92	206.2
MEDIAN_STOCK_LEVEL	8066	5.32	7.20	1	2	3	5	177
LEAD_TIME_DAYS	8066	5.35	14.91	0.02	0.7	1.6	4.33	326

En contraste, las variables asociadas a la variabilidad de la demanda y al tiempo de reposición exhiben valores promedio relativamente bajos, aunque con rangos amplios. En particular, la desviación estándar de la demanda presenta una variabilidad considerable en relación con su media, lo que evidencia distintos niveles de incertidumbre entre los productos analizados. Por su parte, el lead time muestra una dispersión moderada, con la mayoría de los valores concentrados en rangos bajos, pero con la existencia de casos extremos que pueden afectar de forma significativa la gestión del inventario. Finalmente, las métricas de nivel de stock, tanto promedio como mediano, presentan distribuciones concentradas, aunque con valores máximos elevados, lo que sugiere la presencia de inventario de seguridad.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

En cuanto al análisis correlacional, se construyeron matrices de correlación para evaluar la existencia de dependencias lineales significativas entre las variables, en donde se evidencia 5 grupos de variables, así:

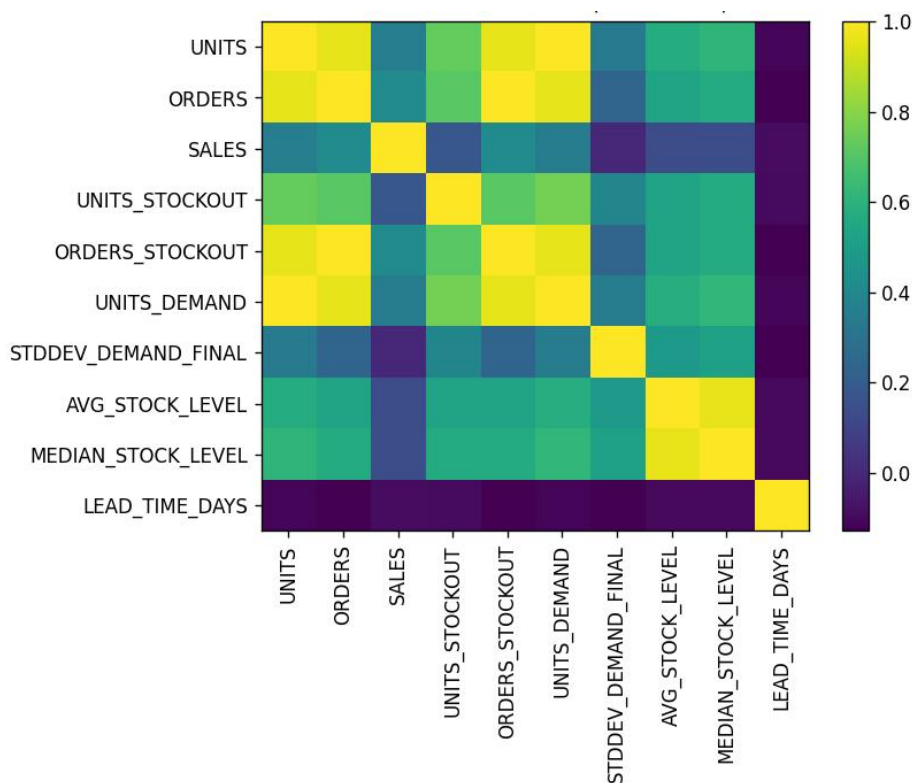
- Demanda: UNITS, ORDERS, UNITS_DEMAND, SALES son variables altamente correlacionadas entre sí, confirmando redundancia informativa (ruido) para el modelo.
- Stockouts: UNITS_STOCKOUT, ORDERS_STOCKOUT tienen una correlación fuerte entre sí y con el grupo de Demanda, evidenciando que este bloque hace referencia a consecuencias y no a variables explicativas para el modelo.
- Inventario: AVG_STOCK_LEVEL y MEDIAN_STOCK_LEVEL poseen una correlación cercana a 1, lo cual es redundancia informativa y debería elegirse solo una para representar este bloque en el modelo.
- Variabilidad de la demanda: STDDEV_DEMAND_FINAL, se evidencia ser una candidata como variable estructural debido a tener una correlación moderada entre la variabilidad de la demanda y los niveles de inventario.
- Tiempo de reposición: LEAD_TIME_DAYS, baja correlación con el resto de variables, lo que indica su carácter exógeno.

En la siguiente Figura se ilustra el mapa de calor de la matriz de correlación.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 5.

Mapa de calor: Matriz de Correlación Retailer 1



De forma complementaria, se implementó un procedimiento de detección de valores atípicos (outliers) basado en el método del z-score, utilizando como umbral $|z| > 3$ para señalar observaciones con comportamiento irregular. Este proceso permitió confirmar que es un proceso estable debido a los pocos outliers y a cuantificar la proporción de outliers por variable, la cual se puede observar en la Tabla 6.

Tabla 6.

Detección de outliers (z-score) Retailer 1

Variable	Conteo Outliers	Porcentaje Outliers
UNITS	114	1.41 %
ORDERS	131	1.62 %
SALES	61	0.76 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Variable	Conteo Outliers	Porcentaje Outliers
UNITS_STOCKOUT	129	1.60 %
ORDERS_STOCKOUT	131	1.62 %
UNITS_DEMAND	117	1.45 %
STDDEV_DEMAND_FINAL	126	1.56 %
AVG_STOCK_LEVEL	127	1.57 %
MEDIAN_STOCK_LEVEL	130	1.61 %
LEAD_TIME_DAYS	108	1.34 %

Para mayor detalle, en el Apéndice C se encuentra de forma extendida los resultados del EDA para todos los retailers. A partir del análisis exploratorio se tomaron las siguientes decisiones clave con el fin de fortalecer la validez interna del estudio: excluir variables del grupo de Stockout, al tratarse de resultados del sistema y no de variables explicativas y evitar el uso simultáneo de variables altamente correlacionadas para reducir multicolinealidad. Así mismo, a pesar de variaciones en la magnitud de los coeficientes, la estructura de correlación se mantuvo estable en todos los retailers analizados. Esto permitió definir un conjunto único de variables estructurales aplicable a todos los casos, garantizando coherencia analítica y comparabilidad, el cual se presentará en la siguiente sección.

6.3 Selección variables de entrada del modelo

Para la clasificación de inventarios, es clave tener presente el volumen de la demanda, la variabilidad de la demanda, los niveles de inventario y el tiempo de reposición. Por tanto, para cubrir este eje operativo y soportado en lo evidenciado en el análisis exploratorio de los datos se establecieron las siguientes variables de entrada para el modelo ensamblado:

La variable “unidades demandadas” (UNITS_DEMAND) representa el volumen total de demanda registrado en el periodo de análisis y constituye una medida directa de la magnitud

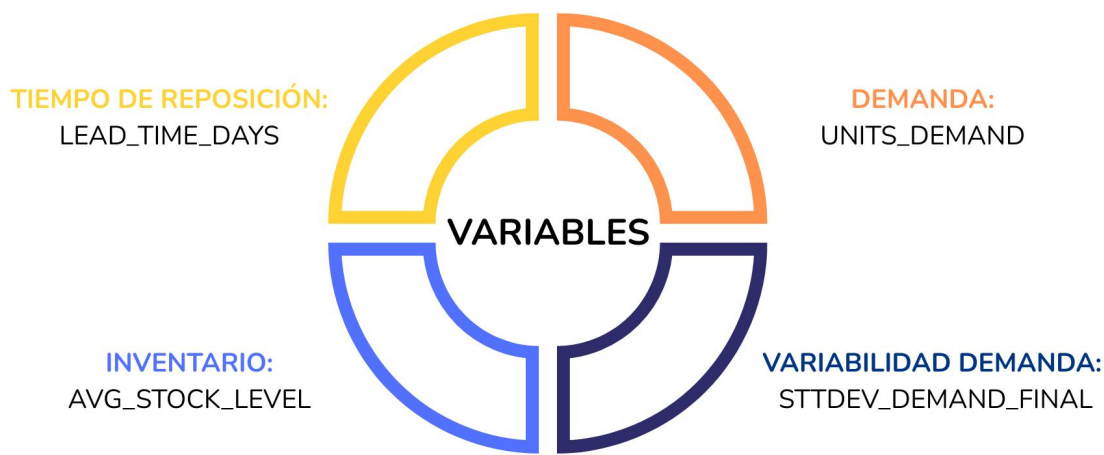
MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

operativa asociada a cada producto. En el presente estudio, esta variable se adopta como criterio principal de priorización, dado que permite capturar el peso relativo de cada ítem dentro del sistema de inventarios, alineándose con enfoques de clasificación ABC por volumen ampliamente utilizados en contextos logísticos. Por otro lado, la “variabilidad de la demanda”, medida a través de la desviación estándar (STDDEV_DEMAND_FINAL), permite caracterizar el grado de incertidumbre asociado al comportamiento de consumo de cada producto.

En complemento, la variable “Tiempo promedio de reposición en días” (LEAD_TIME_DAYS) capta la exposición temporal al riesgo de suministro, variable crítica para determinar los niveles de stock de seguridad y el punto de reorden (Silver, Pyke & Peterson, 1998) . Finalmente, el nivel promedio de inventario (AVG_STOCK_LEVEL) representa el volumen medio de unidades mantenidas en stock a lo largo del periodo analizado. Este indicador permite caracterizar la política de inventarios adoptada y su relación con la dinámica de la demanda y los procesos de reposición, constituyendo una variable clave para el análisis del desempeño del sistema (Axsäter, 2006).

En conjunto, estas cuatro variables describen de manera integral los factores estructurales que condicionan el comportamiento del sistema de inventarios. Específicamente, capturan el nivel y la incertidumbre de la demanda, el riesgo asociado al suministro y la cantidad promedio del stock mantenido por el sistema (Ver Figura 6. Variables seleccionadas para el estudio), proporcionando una base sólida para los modelos base de aprendizaje automático y posterior integración con la matriz de costos en el meta-modelo de consenso.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 6.*Variables seleccionadas para el estudio***6.4 Selección de modelos base**

Dado que la clasificación ABC no constituye un problema supervisado con una variable objetivo claramente definida, cada técnica de clustering tiende a capturar diferentes estructuras latentes en los datos, de modo que el uso de un único algoritmo podría generar una visión parcial o sesgada del comportamiento de los productos del catálogo de cada retailer. Por ello, se adopta un enfoque de complementariedad de modelos, integrando métodos particionales, jerárquicos y probabilísticos que aportan perspectivas analíticas distintas y mutuamente enriquecedoras.

En particular, el algoritmo K-Means, de naturaleza particional, agrupa observaciones en torno a centroides minimizando la varianza intraclase y ofreciendo un modelo rápido y eficiente para la identificación de patrones globales de similitud. Por su parte, el clustering jerárquico permite explorar las relaciones estructurales y niveles de anidamiento entre los productos, facilitando la detección de jerarquías naturales en el inventario. Finalmente, el modelo de mezclas Gaussianas (GMM), de enfoque probabilístico, estima la pertenencia de cada elemento a

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

distintos grupos bajo un marco de verosimilitud estadística, lo que le permite manejar clusters elípticos y solapados y representar la incertidumbre inherente a los datos.

Esta combinación heterogénea de modelos busca capturar tanto la homogeneidad global como las relaciones locales y la variabilidad probabilística del conjunto de datos, proporcionando una clasificación más robusta y resistente al ruido o a valores atípicos. La selección de estos modelos base no solo responde al enfoque de complementariedad, sino que además se fundamenta en lo evidenciado por la revisión de literatura previa, donde estos modelos destacaban por su relevancia en esta área. Como señala Jain (2010), la integración de enfoques particionales, jerárquicos y modelados por probabilidad ofrece una representación más completa y sólida de la estructura subyacente de los datos, lo cual resulta especialmente pertinente en problemas complejos y no supervisados como la clasificación de inventarios. A continuación, se encuentra un resumen de cada modelo base seleccionado para la construcción del modelo ensamblado propuesto junto con la respectiva configuración de hiperparámetros utilizada.

Tabla 7.

Selección modelos base para la construcción del modelo ensamblado

Modelo	Tipo	Descripción	Ventaja	Hiperparámetros
K-means	Particional	Agrupar minimizando la varianza intraclase, por medio de distancias euclidianas.	Rápido, eficiente, base sólida para centroides.	"algorithm": "lloyd", "copy_x": true, "init": "k-means++", "max_iter": 300, "n_clusters": 3, "n_init": "auto", "random_state": 42, "tol": 0.0001, "verbose": 0
Clustering Jerárquico	Jerárquico (Aglomerativo)	Basado en distancias y enlaces, permitiendo explorar la estructura a diferentes niveles.	Detecta jerarquías naturales entre productos.	"compute_distances": false, "compute_full_tree": "auto", "connectivity": null, "distance_threshold": null, "linkage": "ward", "memory": null, "metric": "euclidean", "n_clusters": 3

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Modelo	Tipo	Descripción	Ventaja	Hiperparámetros
Modelo de Mezclas Gaussianas (GMM)	Probabilístico	Maximiza verosimilitud bajo mezcla gaussianas.	Capta incertidumbre y solapamiento entre grupos.	"covariance_type": "full", "init_params": "kmeans", "max_iter": 100, "means_init": null, "n_components": 3, "n_init": 1, "precisions_init": null, "random_state": 42, "reg_covar": 1e-06, "tol": 0.001, "verbose": 0, "verbose_interval": 10, "warm_start": false, "weights_init": null

Cabe resaltar, que con el fin que los resultados fueran interpretables entre modelos base se estableció que se define la clase A/B/C de acuerdo al orden del ranking del valor total de la variable “unidades demandadas” de los grupos o clústeres generados por cada modelo base, es decir, el grupo con mayor valor de unidades demandadas se le asoció la clase A, al siguiente grupo la clase B; y finalmente, la clase C se asoció al grupo con el menor valor de unidades demandadas totales.

6.5 Matriz de Costos

La construcción de la matriz de costos constituyó un componente fundamental dentro del modelo propuesto en razón a que el objetivo general de la presente investigación se centra en desarrollar un modelo ensamblado de aprendizaje automático no supervisado para la clasificación de inventarios buscando la reducción en sus costos. De esta forma se proporciona un marco cuantitativo capaz de evaluar el impacto económico de asignar cada producto a las clases A, B o C dentro de un esquema de clasificación de inventarios. Este procedimiento se desarrolló siguiendo un enfoque metodológico sistemático, orientado a integrar información operativa, como la demanda promedio, la variabilidad ajustada al tiempo de entrega (lead time) y

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

los niveles de servicio (fill rate) junto con los costos asociados relacionado con el mantenimiento y por faltantes.

En primer lugar, se identificaron y prepararon las variables necesarias para la estimación de los costos diarios esperados por clase para cada retailer, incluyendo el costo unitario del producto, la desviación estándar de la demanda durante el tiempo de entrega en días, la demanda promedio diaria y los costos asociados a la no disponibilidad del inventario. Estas variables responden a los fundamentos clásicos de la teoría de inventarios, donde tanto la incertidumbre como el impacto económico de un faltante determinan la necesidad de mantener existencias adicionales.

A partir de la desviación de la demanda y el tiempo de entrega (lead time), se calculó la desviación durante el lead time, de la siguiente manera:

Ecuación 1.

Desviación durante el lead time

$$\text{Desviación durante el lead time } (\sigma_L) = \sigma \cdot \sqrt{\text{lead_time}};$$

Donde σ es la desviación de la demanda.

Cabe resaltar, que se hicieron los siguientes supuestos: en el caso que la desviación de la demanda fuera 0, se fijó un 20% como coeficiente de variación mínimo de la demanda promedio diaria, con el fin de capturar variabilidad residual no observada; y sino se cuenta con información del tiempo de reposición, se estableció un valor de 15 días como dato representativo en un horizonte de reposición típico en contextos operativos similares en retail, adoptando un enfoque conservador ante la incertidumbre.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Posteriormente, se procedió al cálculo del stock de seguridad (SS), el cual depende del nivel de servicio asignado a cada clase. Para cada categoría, se utilizó el valor crítico z de la distribución normal estándar asociado al fill rate correspondiente como se puede observar en la Tabla 8. Tasa de servicio por clase, permitiendo estimar la protección requerida ante la variabilidad de la demanda en el tiempo de entrega mediante la expresión:

Ecuación 2.

Stock de seguridad

$$SS = z \cdot \sigma_L ;$$

Donde z corresponde a la tasa de servicio por clase y σ_L corresponde a la desviación durante el lead time.

Tabla 8.

Tasa de servicio por clase

Clase	Fill Rate	z
A	98%	2.0537
B	95%	1.6449
C	90%	1.2816

Con respecto a la tasa de mantenimiento (carrying rate), se adoptó el 25% anual, debido a que autores como Chopra & Meindl (2019) señalan que la tasa de mantenimiento se sitúa típicamente entre el 20% y el 30% anual, valor ampliamente reconocido en la literatura de gestión de inventarios como representativo del costo total asociado a mantener inventario disponible, considerando componentes como: el costo del capital inmovilizado, seguros, deterioro, obsolescencia, espacio de almacenamiento y gastos administrativos. De este modo, la elección del 25% constituye un valor estándar dentro de las prácticas de la gestión de inventarios

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

aplicable de forma general a múltiples productos y sectores; y adecuado como parámetro base para la estimación del costo de mantenimiento del inventario en el presente modelo.

Así mismo, el costo unitario de los productos se estimó como el 35% del precio de venta, proporción que refleja una relación típica entre margen comercial y estructura de costos en el sector de retail; y en razón que la utilización de un porcentaje fijo facilita la comparación entre productos y asegura consistencia en la estimación del costo de inventario para la construcción de la matriz de costos y el posterior análisis del modelo. A partir de estos valores, se calcularon los costos de mantenimiento (Holding Cost) de inventario a través de la relación:

Ecuación 3.

Costo de mantenimiento

$$\text{Holding Cost } (C_{\text{holding}}) = \frac{\text{carrying_rate}}{365} \cdot \text{unit_cost} \cdot \text{SS};$$

Donde el *carrying_rate* corresponde a la tasa de mantenimiento y SS es el stock de seguridad.

En paralelo, se estimaron los costos por faltantes (Stockout Cost) de inventario, modelados mediante la expresión:

Ecuación 4.

Costo por faltantes

$$\text{Stockout Cost } (C_{\text{stockout}}) = (1 - \alpha) \cdot \mu \cdot \text{unit_cost}_{\text{stockout}};$$

Donde α es la probabilidad de faltantes asociadas al nivel de servicio y μ corresponde a la demanda promedio diaria.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

La cual integra la probabilidad de faltantes dada por el nivel de servicio citados en la Tabla 8, así como la demanda promedio diaria μ y el costo unitario por faltantes, el cual en el caso de no contar con información al respecto se estimó sobre el 70% del precio de venta, dado que este valor aproxima la pérdida económica asociada al margen no capitalizado y a los costos implícitos de los faltantes tales como: pérdida de ventas inmediatas, deterioro del nivel de servicio y efectos indirectos sobre la demanda futura. Asimismo, la adopción de un porcentaje fijo permite estandarizar el tratamiento del faltante entre productos, asegurando consistencia en el modelo y comparabilidad en la evaluación de los resultados.

De este modo, se obtuvo el costo total asociado por cada clase para cada producto, así:

Ecuación 5.

Costo total por clase

$$\text{Costo total por clase (Cost}_x) = C_{\text{holding}_x} + C_{\text{stockout}_x}$$

Con el fin de traducir las diferencias operativas entre los productos en un marco cuantitativo que captura explícitamente las compensaciones entre protección contra la variabilidad y costo económico asociado.

Desde un punto de vista matemático, la matriz de costos se apoya en el principio de riesgo económico en inventarios: niveles de servicio más exigentes (clase A) incrementan el costo de mantener inventario pero reducen significativamente la probabilidad y el impacto de un faltante; mientras que niveles de servicio más bajos (clase C), por el contrario, reducen el costo de mantenimiento pero incrementan la exposición al riesgo de faltantes.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Finalmente, la matriz de costos se integró al meta-modelo del ensamblaje como criterio central para la selección determinística de la clase final en casos de desacuerdo entre los algoritmos de clustering, lo cual se profundizará en la siguiente sección.

6.6 Selección técnica de ensamblaje

El modelo propuesto se fundamenta en una estrategia de ensamblaje heterogéneo, orientada a combinar los resultados de diferentes algoritmos no supervisados aplicados al mismo conjunto de datos. A diferencia de los enfoques supervisados clásicos de “bagging” o “boosting”, que emplean múltiples modelos del mismo tipo con distintas particiones o ponderaciones de los datos, este enfoque integra modelos conceptualmente distintos, capaces de capturar estructuras y patrones complementarios en el problema de clasificación de inventarios. La combinación de métodos particionales, jerárquicos y probabilísticos permite obtener una visión más robusta y generalizable del comportamiento de los productos, reduciendo el sesgo asociado al uso de un único algoritmo.

Con el propósito de dar respuesta a la pregunta de investigación: ¿En qué medida los modelos ensamblados de aprendizaje automático no supervisado mejoran la calidad del agrupamiento y contribuyen a la reducción de costos en los procesos de clasificación de inventarios? En una primera instancia enfocada en la calidad de la agrupación de los clusters, se utilizó una técnica de selección basada en el desempeño de la agrupación determinada por la métrica “coeficiente de silueta”. Y para la relación de la clasificación con los costos se procedió a desarrollar como instancia final un meta-modelo determinístico de stacking, donde la decisión se deriva de una regla determinista que busca clasificar minimizando el costo estimado total, lo cual se describe en detalle a continuación.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

6.6.1 Selección basada en el desempeño de agrupación

Se implementó un mecanismo de ensamblado basado en la selección de acuerdo al desempeño de agrupación por “Silhouette Score”. Dado que no existe una variable objetivo que permita validar externamente las agrupaciones, el Silhouette Score se utilizó como métrica interna de calidad para asignar pesos proporcionales al nivel de cohesión y separación logrado por cada modelo base. De este modo, las decisiones de clasificación final se derivaron de la selección del modelo base con la mayor coherencia estructural según su métrica de silueta, es decir, con la mayor calidad de agrupación.

Este primer enfoque de ensamblado se desarrolló en un entorno de Jupyter Notebook con fines de exploración, funcionando como un “sandbox” previo a la construcción formal del modelo final en un entorno más estructurado. La experimentación permitió identificar el comportamiento comparativo de los algoritmos, sentando las bases para el desarrollo del meta-modelo determinístico de consenso implementado posteriormente.

6.6.2 Meta-modelo determinístico de Stacking

La construcción formal del modelo de ensamblaje se basa en un enfoque de meta-modelado determinístico inspirado en la técnica de stacking (stacked generalization). En su concepto tradicional, el stacking consiste en combinar las predicciones de varios modelos base (base learners) a través de un modelo de nivel superior (meta-learner), que aprende a ponderar las salidas individuales según su desempeño (Wolpert, 1992). En contextos supervisados, este meta-modelo suele adoptar la forma de un algoritmo entrenable, como una regresión logística, una red neuronal o un árbol de decisión, que ajusta sus pesos a partir de ejemplos con etiquetas conocidas. Sin embargo, en la presente investigación, el problema de clasificación de inventarios

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

se desarrolla en un entorno no supervisado, donde no existen etiquetas previas que permitan evaluar la corrección de las agrupaciones generadas.

Ante esta limitación, se propone un meta-modelo determinístico de stacking que no aprende a partir de ejemplos etiquetados, sino que aplica reglas de decisión predefinidas para integrar los resultados de los modelos base. En particular, cuando todos los algoritmos coinciden en la clase asignada (A/B/C), se acepta dicha unanimidad como decisión final; y en los casos de desacuerdo, el meta-modelo selecciona la clase que minimiza el costo total esperado entre las alternativas disponibles, como se relaciona en la siguiente expresión:

Ecuación 6.

Regla de decisión en caso de desacuerdo

$$\text{Tag}_{\text{final}} = \arg \min(\text{Cost}_A, \text{Cost}_B, \text{Cost}_C)$$

Si se presenta un empate entre costos, se aplica un criterio de prioridad jerárquica ($A > B > C$), favoreciendo los productos de mayor impacto operativo. Este mecanismo constituye una forma de stacking no paramétrico, donde la combinación de las etiquetas de los modelos base se realiza mediante una función de consenso determinística que fusiona criterios estadísticos y económicos.

De este modo, el meta-modelo no busca estimar pesos óptimos, sino resolver el desacuerdo entre modelos mediante una lógica de minimización de costos, integrando así la información estructural proveniente del aprendizaje no supervisado con las métricas de impacto operativo derivadas de la matriz de costos, permitiendo obtener una clasificación coherente con los objetivos de las empresas y robusta frente a la variabilidad de los métodos de agrupamiento.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

El código correspondiente al modelo final propuesto para la clasificación de inventarios se encuentra disponible en un [repositorio](#) público de GitHub del autor (2025). Este repositorio contiene la implementación completa del meta-modelo determinístico de stacking, así como los procedimientos de preprocesamiento, modelado y evaluación descritos en el presente trabajo. La estructura del modelo fue diseñada siguiendo principios de modularidad, reproducibilidad y trazabilidad, con el fin de facilitar la ejecución, interpretación y futura extensión del código.

El repositorio sigue la siguiente estructura:

```

ensemble_model_inventory_classification/
├── src/
│   ├── dataset/      # Conjunto de datos en formato .csv
│   ├── results/     # Outputs del modelo
│   ├── utils.py     # Funciones de utilidad
│   └── main.py      # Archivo principal de ejecución
├── requirements.txt  # Archivo con dependencias
├── README.md        # Documentación del proyecto
└── LICENSE          # Licencia MIT

```

El directorio principal, denominado `ensemble_model_inventory_classification/`, está compuesto por una carpeta `src/` que contiene los elementos funcionales del modelo: el subdirectorio `dataset/` almacena los conjuntos de datos en formato `.csv`; la carpeta `results/` concentra las salidas generadas por el modelo (etiquetas, métricas y reportes), siguiendo la siguiente estructura:

```

results/ priority_pca
├── DD_MM_YYYY/      # Fecha
│   └── retailer_x/  # Retailer
│       ├── abc/    # Output modelo ABC Tradicional
│       ├── kmeans/ # Output modelo base Kmeans
│       ├── hc/     # Output modelo base hierarchical clustering
│       ├── gmm/    # Output modelo base GMM
│       ├── merge_models/ # Ensemble modelos base
│       ├── cost_matrix/ # Matriz de costos
│       ├── meta_model/ # Output meta modelo
│       └── performance_evaluation/ # Evaluación de desempeño entre modelos

```

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Además, el repositorio incluye un archivo `requirements.txt` para instalar dependencias; un archivo `README.md` con la documentación técnica y guía de uso, y un archivo `LICENSE` que especifica los términos de uso y distribución del código de acuerdo a la licencia MIT.

El archivo `main.py` constituye el núcleo principal y actúa como el punto de entrada que coordina la ejecución completa del modelo de clasificación de inventarios, permitiendo al usuario interactuar con el proceso mediante la provisión de datasets y la configuración de parámetros clave. El script identifica automáticamente todos los archivos de entrada ubicados en el directorio `dataset/` cuyo nombre siga el patrón `retailer_*.csv`, de modo que el usuario puede incorporar nuevos retailers simplemente añadiendo archivos en formato `.csv` sin necesidad de modificar el código. Asimismo, el usuario puede ajustar diversas variables de entrada, tales como las columnas predictoras empleadas en el modelo (`predictor_columns`), las columnas utilizadas para construir la matriz de costos (`holding_cost_columns` y `stockout_cost_columns`), los niveles de servicio por clase (`fill_rate`), la tasa anual de mantenimiento del inventario (`carrying_rate`), el número de clases (`n_clusters`) y la variable de ordenamiento para el modelo (`ranking_column`).

Una vez definidos estos parámetros, el script ejecuta automáticamente el flujo metodológico: carga y transforma los datos mediante la clase “Preprocess”; calcula los costos esperados de mantenimiento y ruptura a través de “CostMatrix”; y finalmente entrena y combina los modelos base por medio de “EnsembleModel”. Durante la ejecución, el sistema genera mensajes informativos y administra errores sin interrumpir el procesamiento de múltiples retailers, exportando los resultados finales en la estructura de carpetas definidas en el repositorio.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Ahora bien, el archivo `utils.py` concentra la lógica central del modelo y proporciona los componentes reutilizables que son invocados desde `main.py` para implementar el flujo completo de clasificación de inventarios propuesto. En primer lugar, define la función `Banner()`, que muestra en consola la información descriptiva del proyecto y su versión, y sirve como encabezado identificador de la ejecución. A continuación, la función `CostMatrix()` construye la matriz de costos por ítem y por clase a partir de un `DataFrame` de entrada y de un conjunto de parámetros que el usuario configura indirectamente en `main.py`: la columna identificadora del producto (`item_id`), las columnas asociadas al costo de mantenimiento y a la variabilidad de la demanda (`holding_cost_columns`), las columnas vinculadas al costo de faltante y a la demanda esperada (`stockout_cost_columns`), la tasa anual de mantenimiento (`carrying_rate`) y los niveles de servicio por clase (`fill_rate`).

Con estos insumos, la función valida que las columnas existan en el dataset, calcula la tasa diaria de mantenimiento, obtiene el valor crítico z de la distribución normal para cada nivel de servicio, estima el stock de seguridad y deriva tanto el costo de mantenimiento como el costo de faltantes, generando para cada producto los costos totales $Cost_A$, $Cost_B$ y $Cost_C$, además de sus componentes $C_{holding_x}$ y $C_{stockout_x}$. Sobre esta base, las funciones `Cost_ComputeTags()` y `Clustering_ComputeTags()` implementan dos esquemas alternativos de meta-modelo: el primero prioriza el criterio de costos, resolviendo la clasificación final según el mínimo costo esperado cuando los modelos base discrepan; el segundo prioriza el desempeño de clustering, eligiendo la etiqueta del modelo con mayor Silhouette Score. En ambos casos, se construye una estructura que devuelve la etiqueta final (`TAG_FINAL`), el costo asociado (`COST_FINAL`) y los costos por modelo base, lo que permite comparar el modelo ensamblado con el desempeño individual de ABC, K-Means, clustering jerárquico y GMM.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

La clase “Preprocess” se encarga de cargar el dataset desde la carpeta dataset/, seleccionar las columnas predictoras especificadas por el usuario (predictor_columns) y aplicar un pipeline de preprocesamiento que diferencia variables numéricas y categóricas, estandariza las primeras y codifica las segundas mediante one-hot encoding; opcionalmente, puede ejecutar un Análisis de Componentes Principales (PCA) siguiendo el criterio de Kaiser (autovalores > 1), almacenando el número de componentes retenidos y las varianzas explicadas. Por su parte, la clase “EnsembleModel” encapsula todo el flujo de ensamblaje: recibe el DataFrame original, la matriz de características transformada, el número de clases, la columna de identificación (id), la variable de ranking, el nombre del archivo, la matriz de costos, la prioridad del meta-modelo (priority = 'costs' o priority = 'clustering') y la indicación de si se requiere o no PCA. El método `_train_base_learners()` entrena los modelos base (K-Means, Clustering Jerárquico, Gaussian Mixture y el esquema ABC tradicional), calcula etiquetas, Silhouette Score y estadísticas por clase; `_merge_results()` fusiona los resultados de todos los modelos y los integra con la matriz de costos; `_meta_model()` aplica la regla de ensamblaje seleccionada (por costos o por clustering), calcula el Silhouette Score del ensamblaje y genera un DataFrame consolidado con las etiquetas finales y los costos por modelo; `_evaluation()` resume el desempeño de cada modelo en términos de costo total y distribución de SKUs por clase, identificando el modelo económicamente más eficiente; finalmente, `_export_results()` organiza y guarda los resultados en una estructura de carpetas por fecha y dataset, escribiendo tanto archivos CSV con las tablas generadas como archivos JSON con resúmenes de parámetros, etiquetas y métricas.

El método `execute_ensemble()` coordina internamente estas etapas, por lo que el usuario, al ejecutar el modelo desde `main.py`, interactúa con `utils.py` principalmente mediante la configuración de parámetros (columnas de entrada, prioridad del meta-modelo, niveles de

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

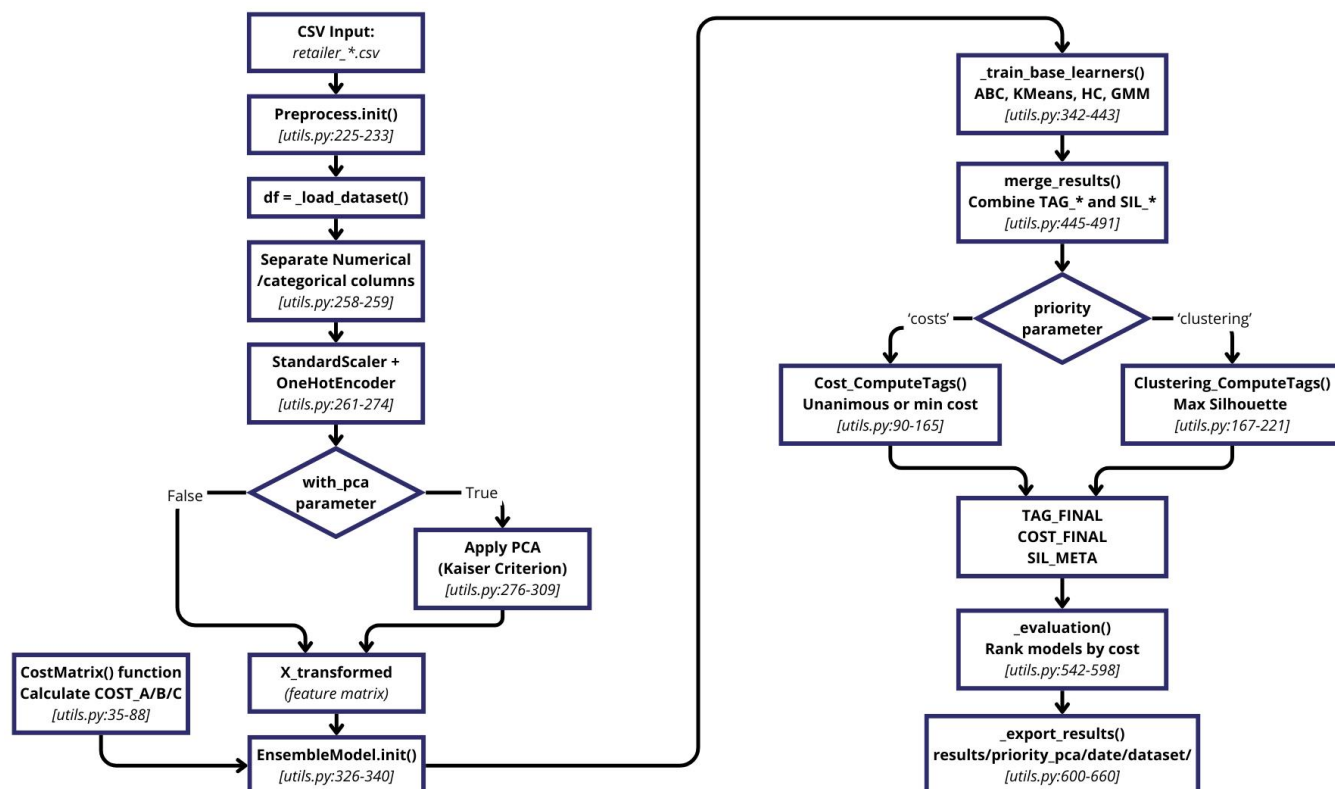
servicio, tasa de mantenimiento, uso o no de PCA), mientras este módulo ejecuta de forma encapsulada el preprocesamiento, el entrenamiento de los modelos base, el ensamblaje, la evaluación y la exportación de resultados de manera consistente y reproducible.

En el siguiente diagrama se ilustra el proceso de principio a fin que sigue el modelo ensamblado propuesto:

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 7.

Flujo modelo ensamblado no supervisado para clasificación de inventarios



Adicional, en esta versión del modelo se llevó a cabo una refactorización del código originalmente desarrollado en el entorno de Jupyter Notebook, donde se había implementado el primer enfoque descrito anteriormente. Esta refactorización, desarrollada en un entorno modular de Python, permitió mejorar la eficiencia computacional y la escalabilidad del modelo, optimizando significativamente los tiempos de ejecución frente a la versión exploratoria inicial.

En la primera versión, el tiempo promedio de procesamiento para un único retailer fue de 11,306 segundos, mientras que la ejecución sobre los quince retailers alcanzó los 82,672 segundos. En contraste, la versión optimizada, implementada en Python con estructuras más eficientes de manejo de datos y paralelización parcial de tareas, redujo dichos tiempos a 9,949 segundos para un retailer y 30,450 segundos para quince retailers. Estas mejoras reflejan una

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

reducción del tiempo total de ejecución superior al 60% en escenarios de mayor carga, demostrando la efectividad de la refactorización en términos de rendimiento, modularidad y capacidad de extensión del modelo.

7. Evaluación del modelo

La evaluación del modelo se estructuró en dos etapas con el fin de dar respuesta a la pregunta de investigación planteada respecto a la capacidad de los modelos ensamblados de aprendizaje automático no supervisado para mejorar la calidad del agrupamiento y contribuir a la reducción de costos en los procesos de clasificación de inventarios. En primer lugar, la evaluación del desempeño de clustering consideró el rendimiento comparativo entre los modelos base y el ensamblado propuesto, utilizando la métrica interna del Silhouette Score para determinar la coherencia estructural de las clases formadas.

Los resultados presentados en la Tabla 9. Desempeño en términos de calidad de agrupamiento, evidencian diferencias claras en el desempeño de los modelos, en donde el ensamblado basado en el desempeño de clustering alcanza de forma general los valores más altos de silueta, con un promedio entre retailers del 68.77% debido a su regla de decisión presentada en la sección previa, coincidiendo en la mayoría de los casos con los obtenidos por el clustering jerárquico (60% de los retailers), lo que indica que este algoritmo fue seleccionado de manera recurrente como el mejor modelo base bajo el criterio de cohesión y separación interna de los clústeres, al lograr un score máximo de 83.78% (retailer 10) y un promedio entre retailers de un 63.63%.

En contraste, K-Means en el 40% de los retailers presenta un desempeño competitivo; sin embargo, en los otros casos muestra un desempeño notoriamente inferior que el modelo base ganador, mostrando una mayor sensibilidad a la forma y dispersión de los datos, mientras que el

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

modelo de mezclas Gaussianas (GMM) exhibe valores de silueta consistentemente bajos, con un valor mínimo de 10.64% (retailer 9) y un promedio entre retailers de 20.75%, lo que sugiere dificultades para capturar adecuadamente la estructura del espacio de características considerado.

Por su parte, el ensamblado de stacking basado en costos registra valores de silueta considerablemente menores, con una disminución de aproximadamente 22 puntos porcentuales (p.p.) con respecto al promedio del coeficiente de silueta entre retailers, lo cual es consistente con su objetivo metodológico, ya que prioriza la minimización del costo total esperado por sobre la calidad de agrupamiento.

Tabla 9.

Desempeño en términos de calidad de agrupamiento

ID	Modelo Ensamblado		K-Means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)
	Stacking basado en costos	Basado en el desempeño del clustering			
1	31.40%	64.83%	58.65%	64.83%	18.26%
2	56.11%	64.23%	64.23%	51.62%	42.03%
3	69.80%	82.26%	66.78%	82.26%	22.51%
4	42.59%	56.27%	56.27%	47.56%	17.10%
5	38.22%	64.09%	61.49%	64.09%	39.64%
6	56.40%	60.72%	60.72%	56.88%	28.86%
7	37.05%	71.10%	71.10%	55.49%	14.99%
8	48.42%	71.65%	71.65%	56.07%	16.39%
9	32.34%	80.69%	80.69%	60.02%	10.64%
10	66.75%	83.78%	54.37%	83.78%	14.00%
11	40.42%	53.79%	47.30%	53.79%	11.90%
12	40.42%	78.04%	49.91%	78.04%	18.48%
13	60.17%	51.09%	50.11%	51.09%	15.04%
14	11.30%	73.02%	65.67%	73.02%	21.83%
15	71.67%	75.97%	52.21%	75.97%	19.55%
MIN	11.30%	51.09%	47.30%	47.56%	10.64%
AVG	46.87%	68.77%	60.74%	63.63%	20.75%
MAX	71.67%	83.78%	80.69%	83.78%	42.03%

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

En conjunto, estos resultados confirman que los ensamblados no supervisados pueden mejorar la calidad del clustering cuando se diseñan explícitamente con ese objetivo, y ya evidencia que existe un compromiso inherente entre maximizar la cohesión estructural de los clústeres y optimizar los costos, aspecto central en la clasificación de inventarios orientada a la toma de decisiones operativas.

Ahora bien, aplicando Análisis de Componentes Principales en el modelo (Ver Tabla 10) se evidencia que el uso de PCA contribuye de manera general a mejorar la calidad del agrupamiento, especialmente en los modelos orientados al desempeño del clustering donde en promedio mejora un 14.20% de manera general. Mientras que en comparación con aquellos sin reducción de dimensionalidad (Tabla 9), el ensamblado basado en el desempeño del clustering replicaba principalmente los resultados del clustering jerárquico, con PCA se observa una mayor estabilidad entre los modelos base, donde el promedio de score de silueta entre los retailers oscila entre el 61.83% y 68.82%. En el caso de K-Means, se observa una mejora más consistente al aplicar PCA, aumentando su silueta en promedio en 7.57 p.p., y reduciendo la variabilidad observada entre retailers y acercando su desempeño al de los métodos jerárquicos, lo cual es coherente con sus supuestos geométricos de clusters compactos y aproximadamente esféricos. Asimismo, el modelo de mezclas Gaussianas (GMM), que sin PCA presentaba los valores de silueta más bajos, experimenta un aumento de 41.08 p.p. en su desempeño relativo promedio, aunque continúa sin superar a los enfoques particionales y jerárquicos.

Tabla 10.

Desempeño en términos de calidad de agrupamiento aplicando PCA

ID	Modelo Ensamblado		K-means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)
	Stacking basado en costos	Basado en el desempeño del clustering			

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

ID	Modelo Ensamblado		K-means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)
	Stacking basado en costos	Basado en el desempeño del clustering			
1	46.54%	74.22%	74.22%	73.89%	65.94%
2	34.98%	72.04%	72.04%	55.68%	61.33%
3	79.97%	80.26%	72.63%	80.26%	64.62%
4	-33.15%	83.06%	69.40%	83.06%	55.81%
5	48.10%	53.31%	50.83%	50.33%	53.31%
6	7.09%	69.69%	62.90%	69.69%	57.95%
7	-16.68%	79.62%	73.09%	79.62%	65.58%
8	8.41%	75.83%	75.83%	71.86%	64.60%
9	-17.25%	77.58%	77.58%	74.07%	58.71%
10	40.89%	67.48%	67.48%	65.27%	66.59%
11	17.13%	62.60%	60.33%	56.04%	62.60%
12	7.74%	68.73%	68.73%	66.68%	62.61%
13	36.01%	74.22%	67.12%	74.22%	62.01%
14	11.69%	66.14%	66.14%	60.14%	61.91%
15	48.88%	71.45%	66.38%	71.45%	63.89%
MIN	-33.15%	53.31%	50.83%	50.33%	53.31%
AVG	21.36%	71.75%	68.31%	68.82%	61.83%
MAX	79.97%	83.06%	77.58%	83.06%	66.59%

Por otro lado, el ensamblado de stacking basado en costos mantiene valores de silueta inferiores tanto con PCA como sin PCA, e incluso aplicando PCA se presentan valores negativos en algunos casos, obteniendo en promedio un score de 21.36% (reducción de ~25 puntos porcentuales), lo que refuerza el hallazgo de que este enfoque no está diseñado para maximizar la cohesión estructural del clustering, sino para mejorar el desempeño en términos de costos.

En conclusión, la comparación indica que la aplicación de PCA potencia la calidad de los agrupamientos y reduce las diferencias entre modelos base, pero no altera el compromiso fundamental identificado previamente entre maximizar la calidad del clustering y priorizar costos en la clasificación de inventarios.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

En segundo lugar, se llevó a cabo una evaluación del desempeño en términos de costos, incorporando no solo los resultados del ensamblaje y de los modelos base, sino también la clasificación tradicional ABC como punto de referencia. Para el modelo tradicional se estableció la variable “unidades demandadas” para realizar la clasificación, asociando a los productos que explican el 80% del valor acumulado a la clase A, a la clase B al 15% adicional del valor acumulado de unidades demandas, y finalmente, a la clase C los items que explican el 5% restante. Para ello, se calcularon los costos esperados por categoría utilizando la matriz de costos derivada de la interacción entre demanda, variabilidad y niveles de servicio, lo que permitió cuantificar el impacto en los costos de cada enfoque como se observa en la Tabla 11.

Tabla 11.

Desempeño en términos de costos

ID	Modelo Ensamblado		ABC	K-means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)
	Stacking based on Costs	Clustering performance based				
1	\$83,861.92	\$112,430.72	\$252,469.98	\$282,414.66	\$112,430.72	\$103,490.97
2	\$11,879.44	\$23,519.64	\$32,178.65	\$23,519.64	\$15,306.55	\$13,738.26
3	\$9,175.98	\$9,732.07	\$24,760.60	\$18,742.40	\$9,732.07	\$19,350.06
4	\$312.52	\$487.95	\$944.04	\$487.95	\$342.05	\$599.10
5	\$77.36	\$137.50	\$208.66	\$141.31	\$137.50	\$92.59
6	\$142.96	\$149.51	\$462.07	\$149.51	\$287.51	\$182.80
7	\$9,284.54	\$12,737.25	\$29,514.52	\$12,737.25	\$11,737.94	\$12,326.32
8	\$11,831.11	\$30,890.22	\$38,393.44	\$30,890.22	\$25,336.97	\$13,298.87
9	\$1,039.98	\$1,932.15	\$3,234.18	\$1,932.15	\$1,924.40	\$1,514.72
10	\$463,224.30	\$694,114.50	\$1,788,022.00	\$1,403,349.00	\$694,114.50	\$542,258.10
11	\$946.68	\$1,657.56	\$2,588.29	\$2,082.31	\$1,657.56	\$1,227.88
12	\$22,808.64	\$32,309.73	\$62,759.90	\$73,446.54	\$32,309.73	\$26,926.69
13	\$252,444.25	\$500,725.28	\$914,048.37	\$825,858.78	\$500,725.28	\$274,795.86
14	\$9,153.39	\$11,028.87	\$26,446.11	\$14,186.28	\$11,028.87	\$12,718.46
15	\$72,653.48	\$114,485.64	\$210,627.09	\$218,780.74	\$114,485.64	\$91,444.35
Total	\$948,836.56	\$1,546,338.59	\$3,386,657.92	\$2,908,718.74	\$1,531,557.28	\$1,113,965.02

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Se observa que el ensamblado basado en costos obtiene de manera consistente los menores costos agregados, alcanzando el valor total de \$948,836.56 USD, siendo el más bajo entre todas las alternativas evaluadas. Este resultado confirma empíricamente que el meta-modelo determinístico orientado a la minimización de costos cumple su objetivo principal, al seleccionar para cada producto la clase que implica el menor impacto económico esperado, integrando explícitamente los costos de mantenimiento y de faltantes.

En contraste, el ensamblado basado en el desempeño del clustering presenta costos totales por un valor de \$1,546,338.59 USD, el cual se traduce en un ~60% más elevado que el modelo ensamblado basado en costos, y es coherente con su función metodológica, ya que prioriza la cohesión estructural de los agrupamientos por sobre los costos. La comparación con el enfoque ABC tradicional evidencia que este último genera costos totales de \$3,386,657.92 USD, siendo +2x el valor del ensamblado basado en el desempeño de clustering y +3x en el caso del modelo basado en costos, lo que sugiere que una clasificación fundamentada únicamente en reglas heurísticas de acumulación porcentual (regla de Pareto) no logra capturar adecuadamente la complejidad ni el costo asociado a los distintos productos.

Entre los modelos base, K-Means y el clustering jerárquico muestran desempeños intermedios, con variabilidad entre retailers, pero sin alcanzar de manera general la eficiencia del ensamblaje basado en costos, mientras que el modelo de mezclas Gaussianas (GMM) mejora respecto a ABC en varios casos, tampoco logra igualar el desempeño económico del modelo propuesto basado en costos, es solo ~17% más costoso, con un valor total de \$1,113,965.02 USD. En conjunto, estos resultados evidencian que la incorporación explícita de un criterio económico dentro del proceso de ensamblado no supervisado permite reducir de forma sustancial el costo

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

total esperado del inventario, reforzando la relevancia de integrar métricas operativas y financieras en la evaluación de modelos de clasificación de inventarios.

Por otro lado, aplicando Análisis de Componentes Principales en el modelo como se observa en la Tabla 12, se evidencia que el uso de PCA contribuye de manera general a aumentar drásticamente los costos para todos los modelos. Se observa que el ensamblado basado en costos continúa siendo el enfoque que minimiza el costo total esperado del inventario, tanto a nivel individual por retailer como en el agregado total, confirmando la robustez del meta-modelo determinístico orientado a la minimización de costos frente a transformaciones del espacio de características/features. No obstante, al comparar con los resultados de no aplicar la reducción de dimensionalidad (Tabla 11), se aprecia que la aplicación de PCA conduce a un incremento de ~136 puntos porcentuales en el costo total agregado para el ensamblado basado en costos respecto al escenario sin PCA, lo que sugiere que, en este caso, la reducción de dimensionalidad puede haber eliminado información relevante para la correcta valoración en términos de costos de los productos.

Tabla 12.

Desempeño en términos de costos aplicando PCA

ID	Modelo Ensamblado		ABC	K-means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)
	Stacking based on Costs	Clustering performance based				
1	\$222,354.94	\$271,477.86	\$252,469.98	\$271,477.86	\$283,032.06	\$275,763.43
2	\$20,875.56	\$31,028.68	\$32,178.65	\$31,028.68	\$22,312.90	\$35,079.42
3	\$9,781.45	\$9,795.32	\$24,760.60	\$10,818.05	\$9,795.32	\$17,825.42
4	\$339.69	\$521.94	\$944.04	\$641.66	\$521.94	\$464.44
5	\$121.94	\$200.36	\$208.66	\$158.67	\$146.73	\$200.36
6	\$246.10	\$324.49	\$462.07	\$297.78	\$324.49	\$388.85
7	\$13,562.60	\$18,120.77	\$29,514.52	\$21,837.60	\$18,120.77	\$22,221.96
8	\$31,452.28	\$38,926.90	\$38,393.44	\$38,926.90	\$34,439.78	\$41,221.95
9	\$1,039.97	\$3,105.98	\$3,234.18	\$3,105.98	\$2,624.61	\$2,384.38

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

10	\$1,122,011.00	\$1,236,212.00	\$1,788,022.00	\$1,236,212.00	\$1,334,791.00	\$1,499,445.00
11	\$946.67	\$2,863.73	\$2,588.29	\$2,534.48	\$2,151.55	\$2,863.73
12	\$60,201.83	\$75,100.13	\$62,759.90	\$75,100.13	\$64,389.34	\$81,326.32
13	\$566,852.94	\$759,642.52	\$914,048.37	\$775,042.89	\$759,642.52	\$843,799.89
14	\$16,775.82	\$19,965.98	\$26,446.11	\$19,965.98	\$18,525.63	\$27,731.04
15	\$170,973.32	\$216,208.17	\$210,627.09	\$196,497.86	\$216,208.17	\$254,729.16
Total	\$2,237,536.11	\$2,683,494.83	\$3,386,657.92	\$2,683,646.52	\$2,767,026.81	\$3,105,445.34

De manera similar, el ensamblado basado en el desempeño del clustering y los modelos base (K-Means, clustering jerárquico y GMM) presentan, en promedio un aumento de 81 puntos porcentuales cuando se aplica PCA, lo que contrasta con las mejoras observadas en términos de Silhouette Score en las tablas de calidad de agrupamiento. Esta divergencia refuerza la idea de que una mayor separabilidad estadística de los clústeres no se traduce necesariamente en decisiones económicamente más eficientes.

En síntesis, la comparación entre las Tablas 11 y 12 indica que, si bien PCA puede mejorar la calidad geométrica del clustering, su impacto sobre el desempeño en costos no mejora de forma directamente proporcional, destacando que la integración explícita de criterios de costo en el meta-modelo tiene un efecto más determinante en la reducción de costos que la transformación del espacio de variables.

Para complementar el estudio de costos, aspecto central de la presente investigación se calcula el delta de costos con respecto al modelo tradicional ABC (sin PCA) como se puede observar en la Tabla 13, donde se permite analizar de manera directa el ahorro económico generado por cada enfoque en comparación con la clasificación heurística de referencia. En términos generales, se observa que el ensamblado basado en costos es el método que logra el mayor ahorro agregado, alcanzando un delta total positivo de \$2,437,821.36 USD, lo que indica

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

una reducción de más de 3x del costo total esperado del inventario frente al modelo ABC. Este resultado confirma que la incorporación explícita de un criterio económico dentro del meta-modelo no solo mejora el desempeño absoluto en costos, sino que también maximiza la ganancia marginal respecto al enfoque tradicional.

Tabla 13.

Delta de costos con respecto al modelo tradicional ABC

ID	Modelo Ensamblado		K-means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)
	Stacking based on Costs	Clustering performance based			
1	\$168,608.06	\$140,039.26	-\$29,944.68	\$140,039.26	\$148,979.01
2	\$20,299.21	\$8,659.01	\$8,659.01	\$16,872.10	\$18,440.39
3	\$15,584.61	\$15,028.53	\$6,018.20	\$15,028.53	\$5,410.53
4	\$631.52	\$456.10	\$456.10	\$601.99	\$344.94
5	\$131.30	\$71.16	\$67.35	\$71.16	\$116.07
6	\$319.11	\$312.55	\$312.55	\$174.56	\$279.27
7	\$20,229.98	\$16,777.27	\$16,777.27	\$17,776.58	\$17,188.21
8	\$26,562.34	\$7,503.22	\$7,503.22	\$13,056.48	\$25,094.57
9	\$2,194.20	\$1,302.03	\$1,302.03	\$1,309.78	\$1,719.46
10	\$1,324,797.70	\$1,093,907.50	\$384,673.00	\$1,093,907.50	\$1,245,763.90
11	\$1,641.61	\$930.73	\$505.98	\$930.73	\$1,360.42
12	\$39,951.26	\$30,450.17	-\$10,686.64	\$30,450.17	\$35,833.21
13	\$661,604.12	\$413,323.10	\$88,189.60	\$413,323.10	\$639,252.52
14	\$17,292.72	\$15,417.24	\$12,259.84	\$15,417.24	\$13,727.66
15	\$137,973.62	\$96,141.45	-\$8,153.65	\$96,141.45	\$119,182.75
Total	\$2,437,821.36	\$1,840,319.33	\$477,939.18	\$1,855,100.64	\$2,272,692.90

Por su parte, el ensamblado basado en el desempeño del clustering también genera ahorros consistentes en todos los retailers, aunque de menor magnitud que el enfoque orientado a costos, lo cual refuerza el hallazgo de que una mejora en la calidad estructural del agrupamiento puede traducirse en beneficios económicos, pero de forma limitada cuando no se optimiza directamente el criterio de costo. Entre los modelos base, el clustering jerárquico presenta un comportamiento similar al del ensamblado basado en el desempeño de agrupamiento debido a su

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

regla de decisión del silhouette score, mientras que K-Means muestra un desempeño más heterogéneo, incluyendo casos con deltas negativos (retailer 1, 12 y 15), traducándose en un ahorro de tan solo \$477,939.18 USD respecto al modelo ABC. El modelo de mezclas Gaussianas (GMM) genera ahorros relevantes en varios casos con un valor agregado de \$2,272,692.90 USD, pero exhibe una mayor variabilidad y no alcanza de forma sistemática los niveles de reducción de costos del ensamblaje basado en costos.

En conclusión, estos resultados muestran que los modelos ensamblados superan consistentemente al enfoque ABC tradicional en términos de costos, siendo el ensamblado basado en costos el que ofrece el mayor beneficio marginal. Asimismo, la presencia de deltas negativos en algunos modelos base subraya el riesgo de aplicar algoritmos de clustering de forma aislada, sin incorporar explícitamente criterios económicos, y refuerza el valor del ensamblaje como mecanismo de mitigación de decisiones subóptimas en la clasificación de inventarios.

Ahora bien, al analizar los resultados correspondientes al delta de costos respecto al modelo tradicional ABC bajo la aplicación de PCA (Ver Tabla 14) se evidencian cómo la reducción de dimensionalidad influye en la capacidad de los distintos enfoques para generar ahorros económicos.

Tabla 14.

Delta de costos con respecto al modelo tradicional ABC aplicando PCA

ID	Modelo Ensamblado		K-means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)
	Stacking based on Costs	Clustering performance based			
1	\$30,115.04	-\$19,007.88	-\$19,007.88	-\$30,562.08	-\$23,293.45
2	\$11,303.10	\$1,149.97	\$1,149.97	\$9,865.75	-\$2,900.76
3	\$14,979.14	\$14,965.27	\$13,942.54	\$14,965.27	\$6,935.17
4	\$604.35	\$422.10	\$302.38	\$422.10	\$479.60
5	\$86.72	\$8.30	\$49.98	\$61.92	\$8.30
6	\$215.97	\$137.58	\$164.29	\$137.58	\$73.22

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

ID	Modelo Ensamblado		K-means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)
	Stacking based on Costs	Clustering performance based			
7	\$15,951.92	\$11,393.75	\$7,676.93	\$11,393.75	\$7,292.57
8	\$6,941.16	-\$533.45	-\$533.45	\$3,953.67	-\$2,828.50
9	\$2,194.21	\$128.20	\$128.20	\$609.57	\$849.80
10	\$666,011.00	\$551,810.00	\$551,810.00	\$453,231.00	\$288,577.00
11	\$1,641.63	-\$275.44	\$53.81	\$436.75	-\$275.44
12	\$2,558.07	-\$12,340.23	-\$12,340.23	-\$1,629.44	-\$18,566.42
13	\$347,195.43	\$154,405.85	\$139,005.48	\$154,405.85	\$70,248.48
14	\$9,670.29	\$6,480.14	\$6,480.14	\$7,920.49	-\$1,284.93
15	\$39,653.78	-\$5,581.07	\$14,129.24	-\$5,581.07	-\$44,102.06
Total	\$1,149,121.81	\$703,163.09	\$703,011.40	\$619,631.11	\$281,212.58

A nivel agregado, el ensamblado basado en costos continúa mostrando un ahorro total positivo y significativo de \$1,149,121.81 USD, lo que confirma su robustez como estrategia orientada a la minimización del costo del inventario incluso cuando se modifica el espacio de características (PCA). No obstante, al contrastar estos resultados con el escenario sin PCA, se observa una disminución del ahorro total en alrededor 2x, lo que sugiere que la aplicación de PCA puede eliminar información relevante para la correcta clasificación de los productos. Este patrón se repite en el ensamblado basado en el desempeño del clustering y en los modelos base, donde se identifican con mayor frecuencia deltas negativos, indicando casos en los que la clasificación resultante empeora el desempeño económico respecto al modelo tradicional ABC.

En particular, K-Means y GMM muestran una mayor sensibilidad a la reducción de dimensionalidad, mientras que el clustering jerárquico mantiene un comportamiento relativamente más estable, aunque con ahorros 3x inferiores a los observados sin PCA. Cabe resaltar que sin aplicar PCA el modelo de mezclas Gaussianas (Top 2) lograba un desempeño agregado en términos de costos similar al modelo ensamblado basado en costos con un ahorro de \$2,272,692.90 USD con respecto al ABC, al aplicar PCA su desempeño en costos lo lleva a

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

ubicarse último en el ranking de ahorros, con un valor de \$281,212.58 USD con respecto al modelo tradicional, lo cual evidencia que este modelo base es un buen punto de partida para obtener una clasificación robusta de inventarios con un buen desempeño en términos de costos sin necesidad siquiera de someterlo a evaluación, aspecto clave como hallazgo dentro del desarrollo de la investigación.

Considerando de manera integrada los resultados obtenidos en las distintas evaluaciones realizadas, se observa que los modelos ensamblados no supervisados aportan mejoras relevantes en la clasificación de inventarios, aunque su impacto depende críticamente del criterio utilizado para su construcción y evaluación. Desde la perspectiva de la calidad del agrupamiento, el modelo ensamblado basado en el desempeño del clustering y la aplicación de PCA conducen a incrementos sistemáticos en el Silhouette Score, reflejando una mejor definición estructural de los clústeres. Sin embargo, cuando el foco se traslada al desempeño en términos de costos, el ensamblado basado en costos destaca de forma consistente como la alternativa más eficiente, superando tanto a los modelos base como al enfoque tradicional ABC, y generando los mayores ahorros absolutos (\$2,437,821.36 USD) y relativos.

Asimismo, la comparación entre escenarios con y sin PCA pone en evidencia que la reducción de dimensionalidad, si bien beneficia la calidad de agrupamiento del clustering, no siempre se traduce en decisiones económicamente superiores e incluso puede resultar contraproducente en ciertos contextos. En consecuencia, los resultados subrayan la importancia de alinear el diseño del modelo con los objetivos operativos, demostrando que la integración explícita de criterios económicos en el ensamblado no supervisado es un factor determinante para mejorar la toma de decisiones en la gestión de inventarios. Para un mayor detalle de los

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

resultados, en el Apéndice D se adjunta la tabla comparativa completa en términos de costos y de calidad de agrupamiento entre modelos.

8. Conclusiones y Recomendaciones

El objetivo trazado de desarrollar un modelo ensamblado de aprendizaje automático no supervisado para la clasificación de inventarios buscando la reducción en sus costos se logró adoptando un enfoque epistemológico positivista centrado en el modelo hipotético-deductivo con la obtención de conocimiento a partir de la observación y el análisis objetivo de los fenómenos estudiados en 15 retailers de Latinoamérica, garantizando resultados medibles, verificables y siguiendo principios científicos a partir de una revisión exhaustiva de la literatura relacionada con la clasificación de inventarios, encontrando que existen enfoques tradicionales como el método ABC y el análisis basado en la toma de decisiones multicriterio (MCDM); y los enfoques de aprendizaje automático como: el aprendizaje supervisado (árboles de decisión, las máquinas de vectores de soporte y las redes neuronales profundas), aprendizaje no supervisado (K-means, Variational Autoencoders, Clustering jerárquico y modelos de mezclas Gaussianas) y aprendizaje ensamblado, hallando que el uso de modelos ensamblados de aprendizaje automático puede mitigar algunas limitaciones inherentes a los modelos base; no obstante, hasta el momento no se han aplicado modelos ensamblados no supervisados en el área de clasificación de inventarios, evidenciando un espacio que favorece la investigación actual, y de esta manera se concluye con la caracterización del problema de clasificación de inventarios.

Seguido con la formulación del modelo desarrollado en lenguaje de programación Python con base al análisis exploratorio de los datos; se concluye que las variables de entrada más relevantes para el modelo son: unidades demandadas, desviación estándar de la demanda, el nivel promedio de stock y el tiempo de reposición en días; así mismo, se determinan como

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

modelos base: K-means, Clustering jerárquico y modelos de mezclas Gaussianas al ser las alternativas que capturaban de mejor forma las características del problema, y buscando alinear el diseño con los objetivos operativos, se formuló el modelo ensamblado de aprendizaje automático no supervisado con una estructura modular para poder priorizar el criterio de clasificación según sea la necesidad, ya sea en términos de calidad de agrupamiento de los clústeres o en términos de costos; como la aplicación, o no, de la reducción de dimensionalidad (PCA) en los datos. Con este enfoque propuesto se buscó analizar la clasificación de los productos, no solo desde una perspectiva estructural, sino también en función de su desempeño en términos de costos, mediante la estimación de los costos esperados de mantenimiento de inventario de seguridad y de faltantes, demostrando que la integración explícita del criterio de costos como regla de decisión en la clasificación de inventarios permite la reducción de los mismos, siendo factor determinante para mejorar la toma de decisiones en la gestión de inventarios.

Al evaluar el modelo propuesto para la clasificación de inventarios bajo las políticas de inventario de cada clase relacionadas con el nivel de servicio y su impacto asociado a los costos, se presentan como hallazgos relevantes que al contrastar los resultados del modelo ensamblado basados en costos con el escenario sin PCA, se observa una disminución del ahorro total en alrededor 2x, lo que sugiere que la aplicación de PCA puede eliminar información relevante para la correcta clasificación de los productos. Este patrón se repite en el ensamblado basado en el desempeño del clustering y en los modelos base, donde se identifican con mayor frecuencia deltas negativos, indicando casos en los que la clasificación resultante empeora el desempeño en costos con respecto al modelo tradicional ABC. En particular, K-Means y GMM muestran una mayor sensibilidad a la reducción de dimensionalidad, mientras que el clustering jerárquico

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

mantiene un comportamiento relativamente más estable, aunque con ahorros 3x inferiores a los observados sin PCA. Se concluye que el modelo ensamblado no supervisado basado en costos para los 15 retailers bajo estudio presenta un mejor desempeño económico en comparación con la clasificación ABC tradicional con un ahorro agregado de \$2,437,821.36 USD, al generar asignaciones que conducen a menores costos totales esperados bajo las políticas de nivel de servicio definidas. Este comportamiento se mantiene de forma consistente en los distintos escenarios evaluados, lo que sugiere que el uso de técnicas no supervisadas y enfoques de ensamblado constituye una alternativa viable, robusta y efectiva para apoyar la toma de decisiones en la gestión de inventarios, particularmente en contextos donde no se dispone de datos etiquetados.

Respondiendo a la pregunta de investigación, un hallazgo relevante es la identificación de un trade-off claro entre las métricas internas de calidad de clustering y el desempeño en términos de costos. Y se observó que la aplicación de técnicas de reducción de dimensionalidad (PCA) mejora en promedio 14.20% entre los retailers bajo estudio la separabilidad estructural de los clústeres en los modelos orientados al desempeño del agrupamiento, reflejada en mayores valores del silhouette score. Sin embargo, dicha mejora estructural no se traduce necesariamente en una reducción de los costos asociados. De manera inversa, los modelos orientados a minimizar el costo esperado presentan en varios casos, valores bajos o incluso negativos del silhouette score, lo que indica que una buena estructura geométrica de los clústeres, no garantiza una asignación de costos eficiente.

En relación con los modelos base del ensamblado se concluye que el modelo de mezclas Guassianas (GMM) presenta un desempeño particularmente favorable en términos de costos cuando no se aplica PCA, alcanzando un ahorro agregado de \$2,272,692.90 USD con respecto al

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

modelo tradicional ABC, a pesar de poseer de manera general el peor desempeño en clustering con un score promedio de silueta de 20.75%. En consecuencia, se puede concluir que el uso de este modelo base sirve como punto de partida para la clasificación de inventarios para la reducción de los costos, sin necesidad de construir matrices de costos complejas como criterio principal de clasificación, lo cual contribuye a disminuir las barreras de adopción del modelo, especialmente en organizaciones con recursos limitados.

El modelo ensamblado que prioriza el criterio de clustering se posiciona como la alternativa que ofrece el mejor balance entre desempeño en términos de calidad de agrupamiento y de costos cuando no se aplica PCA, pues dicho modelo alcanza los mayores valores de silhouette score (68.77%) y, simultáneamente, presenta costos totales esperados de \$1,546,338.59 USD, con un ahorro de \$1,840,319.33 USD con respecto a los de la clasificación ABC tradicional, lo que lo convierte en una solución balanceada capaz de combinar cohesión estructural y eficiencia en costos. En conclusión, los hallazgos de esta investigación ponen de manifiesto que la calidad de un modelo de clasificación de inventarios depende críticamente del criterio de evaluación adoptado y que las métricas tradicionales de clustering no siempre están relacionadas de forma directamente proporcional con los costos. En este sentido, el modelo propuesto constituye una herramienta valiosa para apoyar la toma de decisiones estratégicas en la gestión de inventarios y abre nuevas oportunidades para el desarrollo de enfoques multiobjetivo que integren de manera explícita criterios estructurales, financieros y operativos.

Desde una perspectiva institucional y académica, esta investigación incentiva el desarrollo de nuevas líneas de trabajo dentro del grupo de investigación OPALO de la Escuela de Estudios Industriales y Empresariales, y en otras comunidades científicas, en las áreas de

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

automatización y control y de analítica de datos para facilitar herramientas que permitan cerrar la brecha entre los enfoques tradicionales de gestión de inventarios y las exigencias de la industria moderna caracterizadas por la transformación digital, el uso de inteligencia artificial y el análisis avanzado de datos. A mediano plazo, se espera que los resultados trasciendan del ámbito académico al fortalecimiento de la competitividad industrial mediante la reducción de costos en la gestión de los inventarios en las empresas, promoviendo así la adopción de tecnologías avanzadas en dicha gestión, especialmente en pequeñas y medianas empresas y favoreciendo una mayor integración entre teoría y práctica.

Ahora bien, como líneas de investigación futura, se plantean las posibilidades de ampliar el conjunto de variables analizadas incorporando nuevos predictores de naturaleza operativa, financiera o logística, con el fin de evaluar su influencia en la estructura de agrupamiento y en la clasificación final de inventarios. Asimismo, se propone no restringir el modelo a tres clases (A, B y C) sino explorar configuraciones con un número variable de categorías que otorguen mayor flexibilidad y granularidad en la segmentación, permitiendo su adaptación a distintos entornos y tipos de inventario. Otra extensión relevante consiste en la incorporación de productos pesables o de naturaleza continua, cuya dinámica de consumo y reposición difiere de la de los productos de consumo masivo (CPGs). Otra línea a investigar basada en el feedback recibido en la presentación de las ponencias en la Conferencia ICPR 2025 consistiría en ampliar el conjunto de modelos no supervisados considerados, evaluando el desempeño de diversas técnicas de clustering y seleccionando aquellas con mejor desempeño según el coeficiente de silueta promedio, en función de los datos y variables de entrada.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

El optimizar los hiperparámetros de los modelos base mediante estrategias de búsqueda sistemática como Grid Search, con el objetivo de identificar las combinaciones de parámetros que maximicen el coeficiente de silueta y, en consecuencia, la calidad de los agrupamientos generados, puede ser otra oportunidad de investigación a futuro. Asimismo, se propone evaluar la conveniencia de balancear las clases de acuerdo con las necesidades específicas de los retailers, sin afectar negativamente los costos, estas extensiones contribuirían a fortalecer la robustez, flexibilidad y precisión del enfoque de ensamblado propuesto para la clasificación de inventarios.

Referencias

- Abidin, M. N. A. Z., Bazin, N. E. N., Zebari, D. A., Rahmat, A. N., & Asaad, R. R. (2024). Inventory Categorization Using Multiple Criteria Classification. *Journal of Soft Computing and Data Mining*, 5(1), 142–151. <https://doi.org/10.30880/jscdm.2024.05.01.012>
- Aktepe, A., Ersoz, S., Turker, A. K., Barisci, N., & Dalgic, A. (2018). An inventory classification approach combining expert systems, clustering, and fuzzy logic with the ABC method, and an application. *South African Journal of Industrial Engineering*, 29(1), 49–62. <https://doi.org/10.7166/29-1-1784>
- Albayrak Ünal, Ö., Erkayman, B., & Usanmaz, B. (2023). Applications of Artificial Intelligence in Inventory Management: A Systematic Review of the Literature. *Computational Methods in Engineering*, 2605–2625.
- Axsäter, S. (2006). *Inventory Control*. Springer.
- Bailkar, S., Shenoy, K., Bedekar, A., Bankar, S., & More, P. (2024). Smart Inventory Optimization using Machine Learning Algorithms. *2nd International Conference on Intelligent Data Communication Technologies and Internet of Things, IDCIoT 2024*, 1395–1400. <https://doi.org/10.1109/IDCIoT59759.2024.10467512>
- Bala, P. K. (2012). Improving inventory performance with clustering based demand forecasts. *Journal of Modelling in Management*, 7(1), 23–37. <https://doi.org/10.1108/17465661211208794>
- Chopra, S., & Meindl, P. (2019). *Supply Chain Management: Strategy, Planning, and Operation* (7th ed.). Pearson.
- Chu, C. W., Liang, G. S., & Liao, C. T. (2008). Controlling inventory by combining ABC analysis and fuzzy classification. *Computers and Industrial Engineering*, 55(4), 841–851. <https://doi.org/10.1016/j.cie.2008.03.006>

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

- Dixit, A. (2017). *Ensemble Machine Learning*. Packt Publishing.
- Eraslan, E., & Iç, Y. T. (2020). An improved decision support system for ABC inventory classification. *Evolving Systems, 11*(4), 683–696. <https://doi.org/10.1007/s12530-019-09276-7>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases* (© AAAI) (Vol. 17). www.ffly.com/
- García-Barrios, D., Palomino, K., García-Solano, E., & Cuello-Quiroz, A. (2021). A Machine Learning based Method for Managing Multiple Impulse Purchase Products: An Inventory Management Approach. *Journal of Engineering Science and Technology Review, 14*(1), 25–37. <https://doi.org/10.25103/jestr.141.02>
- Hadi-Vencheh, A. (2010). An improvement to multiple criteria ABC inventory classification. *European Journal of Operational Research, 201*(3), 962–965. <https://doi.org/10.1016/j.ejor.2009.04.013>
- Haoud, N. El, & Bachiri, Z. (2019). *Stochastic Artificial Intelligence benefits and Supply Chain Management: inventory prediction*.
- Jackson, I. (2022). AutoML Approach to Stock Keeping Units Segmentation. *Journal of Theoretical and Applied Electronic Commerce Research, 17*(4), 1512–1528. <https://doi.org/10.3390/jtaer17040076>
- Jackson, I., Avdeikins, A., & Tolujevs, J. (2019). Unsupervised learning-based stock keeping units segmentation. In *Lecture Notes in Networks and Systems* (Vol. 68, pp. 603–612). Springer. https://doi.org/10.1007/978-3-030-12450-2_58
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

- Kaabi, H., Jabeur, K., & Ladhari, T. (2018). A genetic algorithm-based classification approach for multicriteria ABC analysis. *International Journal of Information Technology and Decision Making*, 17(6), 1805–1837. <https://doi.org/10.1142/S0219622018500475>
- Kartal, H., Oztekin, A., Gunasekaran, A., & Cebi, F. (2016). An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification. *Computers and Industrial Engineering*, 101, 599–613. <https://doi.org/10.1016/j.cie.2016.06.004>
- Khanorkar, Y., & Kane, P. V. (2023). Selective inventory classification using ABC classification, multi-criteria decision making techniques, and machine learning techniques. *Materials Today: Proceedings*, 72, 1270–1274. <https://doi.org/10.1016/j.matpr.2022.09.298>
- Liu, J., Liao, X., Zhao, W., & Yang, N. (2016). A classification approach based on the outranking model for multiple criteria ABC analysis. *Omega (United Kingdom)*, 61, 19–34. <https://doi.org/10.1016/j.omega.2015.07.004>
- Lolli, F., Balugani, E., Ishizaka, A., Gamberini, R., Rimini, B., & Regattieri, A. (2019). Machine learning for multi-criteria inventory classification applied to intermittent demand. *Production Planning and Control*, 30(1), 76–89. <https://doi.org/10.1080/09537287.2018.1525506>
- Lolli, F., Ishizaka, A., Gamberini, R., Balugani, E., & Rimini, B. (2017). Decision Trees for Supervised Multi-criteria Inventory Classification. *Procedia Manufacturing*, 11, 1871–1881. <https://doi.org/10.1016/j.promfg.2017.07.326>
- Lopez-Soto, D., Ángel-Bello, F., Soto, D. L., Yacout, S., & Angel-Bello, F. (2016). *Inventory ABC Supervised Classification with Logical Analysis of Data Soumaya Yacout Polytechnique Montréal Inventory ABC Supervised Classification with Logical Analysis of Data*. <https://www.researchgate.net/publication/304158598>

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

- Mahendrawathi E., Nurul Laili E., & Kusumawardani R. (2011). *Classification of hospital pharmaceutical drug inventory items by combining ABC analysis and fuzzy classification*. IEEE.
- Maitra, S., & Kundu, S. (2023). Backorder Prediction in Inventory Management: Classification Techniques and Cost Considerations. *27th International Computer Science and Engineering Conference 2023, ICSEC 2023*, 1–8. <https://doi.org/10.1109/ICSEC59635.2023.10329654>
- Michalski, M., Melani, A., da Silva, R., & de Souza, G. (2021). Applying an Unsupervised Machine Learning Method for Defining Maintenance Significant Items. *Proceedings of the 31st European Safety and Reliability Conference, ESREL 2021, 1 PartF*, 2525–2532. <https://doi.org/10.3850/978-981-18-2016-8>
- Mohammaditabar, D., Hassan Ghodsypour, S., & Obrien, C. (2012). Inventory control system design by integrating inventory classification and policy selection. *International Journal of Production Economics*, *140*(2), 655–659. <https://doi.org/10.1016/j.ijpe.2011.03.012>
- Ogunfowora, O., & Najjaran, H. (2023). Reinforcement and deep reinforcement learning-based solutions for machine. *Journal of Manufacturing Systems*, 244–263.
- Palacio-Niño, J.-O., & Berzal, F. (2019). *Evaluation Metrics for Unsupervised Learning Algorithms*. <http://arxiv.org/abs/1905.05667>
- Park, Y. S., Konge, L., & Artino, A. R. (2020). The Positivism Paradigm of Research. In *Academic Medicine* (Vol. 95, Issue 5, pp. 690–694). Wolters Kluwer Health. <https://doi.org/10.1097/ACM.0000000000003093>
- Partovi, F. Y., & Anandarajan, M. (2004). *Classifying inventory using an artificial neural network approach*. www.elsevier.com/locate/dsw
- Qaffas, A. A., Ben HajKacem, M. A., Ben Ncir, C. E., & Nasraoui, O. (2023). An Explainable Artificial Intelligence Approach for Multi-Criteria ABC Item Classification. *Journal of*

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Theoretical and Applied Electronic Commerce Research, 18(2), 848–866.

<https://doi.org/10.3390/jtaer18020044>

- Qaffas, A. A., Hajkacem, M. A. Ben, Ncir, C. E. Ben, & Nasraoui, O. (2023). Interpretable Multi-Criteria ABC Analysis Based on Semi-Supervised Clustering and Explainable Artificial Intelligence. *IEEE Access*, 11, 43778–43792. <https://doi.org/10.1109/ACCESS.2023.3272403>
- Rehmani, K., Naseem, A., Ahmad, Y., Mirza, M. Z., & Syed, T. H. (2021). Development of a hybrid framework for inventory leanness in Technical Services Organizations. *PLoS ONE*, 16(2 February). <https://doi.org/10.1371/journal.pone.0247144>
- Seyedan, M., Mafakheri, F., & Wang, C. (2022). Cluster-based demand forecasting using Bayesian model averaging: An ensemble learning approach. *Decision Analytics Journal*, 3, 100033. <https://doi.org/10.1016/j.dajour.2022.100033>
- Sh. Hematyar. (2011). *Fuzzy Classification of Gas Power Plant Spare Parts by Combination Statistical Classification Technique, SAW, ABC Analysis*. IEEE.
- Shamsaddini, R., Mahdi Vesal, S., & Management, M. (2015). A new model for inventory items classification through integration of ABC-Fuzzy and fuzzy analytic hierarchy process Khaled Nawaser. In *Int. J. Industrial and Systems Engineering* (Vol. 19, Issue 2).
- Silver, E. A., Pyke, D. F., & Peterson, R. (1998). *Inventory Management and Production Planning and Scheduling* (3rd ed.). Wiley.
- Svoboda, J., & Minner, S. (2022). Tailoring inventory classification to industry applications: the benefits of understandable machine learning. *International Journal of Production Research*, 60(1), 388–401. <https://doi.org/10.1080/00207543.2021.1959078>
- Toscano Serrano, A. F. (2025). *Ensemble Model Inventory Classification [Source code]*. GitHub. https://github.com/Andrest7/ensemble_model_inventory_classification

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review*. In *British Journal of Management* (Vol. 14).
- Weng, F., Guo, B., Suo, X., Wu, X., Hu, Y., & Zheng, Q. (2022). Semi-supervised Graph Convolutional Neural Network Based Classification for Auto Parts Inventory Management. *2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, 1050–1054.
<https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys57074.2022.00166>
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.
[https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- Yu, M. C. (2011). Multi-criteria ABC analysis using artificial-intelligence-based classification techniques. *Expert Systems with Applications*, 38(4), 3416–3421.
<https://doi.org/10.1016/j.eswa.2010.08.127>
- Zowid, F. M., Babai, M. Z., Douissa, M. R., & Ducq, Y. (2019). Multi-criteria inventory ABC classification using Gaussian Mixture Model. *IFAC-PapersOnLine*, 52(13), 1925–1930.
<https://doi.org/10.1016/j.ifacol.2019.11.484>

Apéndices

Apéndice A. Recuperación de literatura y análisis bibliométrico

Actualmente, las empresas que buscan mantenerse al ritmo del avance tecnológico y la globalización deben ser capaces de gestionar sus inventarios de manera eficiente para alcanzar alta calidad, mayor productividad y reducción de costos. Sin embargo, debido a la diversificación de las necesidades de los clientes, la incertidumbre y competencia del mercado, el aumento de los costos de transporte, el crecimiento exponencial de los e-commerces y la nueva tendencia de entrega de última milla; resulta cada vez más complejo la correcta clasificación de inventarios para cumplir con los objetivos estratégicos de las compañías. Con el avance y la aplicación de tecnologías como la inteligencia artificial (IA) y nuevas técnicas de aprendizaje automático (ML) en otras áreas de investigación, se evidencia el interés en apostar por esta clase de algoritmos y modelos, ya que estos permiten identificar y analizar patrones rápidamente dentro de conjuntos de datos; lo que permitiría y transformaría la clasificación de inventarios en un proceso robusto, flexible e inteligente.

Ahora bien, se propone una revisión de literatura con el fin de ofrecer una visión actual del estado del arte en este campo de la investigación, explorando e indagando exhaustivamente los algoritmos de aprendizaje automático utilizados en la clasificación de inventarios, en donde, se pretende consolidar las aplicaciones actuales y señalar posibles direcciones u oportunidades para futuras investigaciones en la clasificación de inventarios.

Para la presente revisión, se parte de la metodología de Tranfield, Denyer, & Smart (2003), y se determina el paso a paso a desarrollar con el objetivo de realizar una búsqueda eficaz y obtener los resultados esperados, tal como se ilustra en la Figura 1.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 1.*Metodología propuesta***Fase I:** Planificación de la revisión

La planificación requerida para la revisión, parte de identificar la necesidad de llevarla a cabo, para la presente revisión esta surge de la importancia de los Inventarios en los activos de las empresas y por tanto, su impacto en la rentabilidad del negocio; así mismo, surge debido a la brecha que existe en la velocidad exponencial en el desarrollo tecnológico y la capacidad de respuesta de las empresas para adaptarse a dichas tecnologías para ser capaces de gestionar de forma eficaz sus inventarios, garantizando calidad, eficiencia y bajos; de igual manera, cabe destacar y enmarcar los algoritmos de aprendizaje automático, contextualizando su importancia en la gestión de inventarios, ya que estos pueden identificar y analizar patrones rápidamente dentro de conjuntos de datos grandes y diversos. Lo que genera que la clasificación de inventarios sea un proceso robusto, objetivo, eficiente y flexible.

Fase II: Selección de las fuentes de información

Parte del éxito de la revisión recae en la calidad de las fuentes de información donde se extrae la información; por tanto, las bases de datos científicas resultan acordes, ya que estas

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

tienen como objetivo reunir toda la producción bibliográfica sobre un área de conocimiento en particular; es decir, recopilar todo el contenido científico y técnico como: artículos de revistas, actas de conferencia, capítulos de libro, entre otros.

Además, teniendo claro que según Sampieri Hernández (2008) , una fuente de información es todo aquello que proporciona datos para reconstruir las bases del conocimiento y satisfacer una demanda de información, y que son un instrumento para el conocimiento, la búsqueda y el acceso a la misma, se definen dentro de las fuentes principales, las bases de datos de “Scopus” y “Web of Science (WOS) dada su relevancia e importancia dentro de la comunidad investigativa. Estas bases de datos seleccionadas para llevar a cabo la recopilación de la información relacionada con los algoritmos de aprendizaje automático aplicados en la gestión de inventarios permiten ofrecer herramientas para realizar un exhaustivo resumen, seguimiento, análisis y visualización de los resultados de la investigación mundial multidisciplinar, con el fin de asegurarse de que se está al corriente de los estudios y trabajos cruciales de todo el mundo. En el caso de las fuentes auxiliares, se acude a artículos de referencia recuperados de bases de datos de acceso libre como “Google Scholar”.

Fase III: Realización de la búsqueda

Se parte de la identificación de palabras clave y términos relevantes del tema; definiendo dos pilares o nubes de términos sobre “*Clasificación de inventarios*” y “*Aprendizaje automático*” para la extracción de los estudios relacionados en bases de datos. De esto modo, se formula la primera ecuación de búsqueda referente a la base de datos “Scopus”, quedando de la siguiente manera:

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Ecuación 1.

Ecuación de búsqueda en Scopus

TITLE-ABS-KEY (((*item OR inventory OR stock OR "Stock keeping unit" OR sku OR abc*)
PRE/1 (*classification OR categori?ation OR segmentation OR clustering OR analysis*))
AND (*"Ensemble learning" OR "Unsupervised learning" OR "unsupervised clustering" OR*
"Deep learning" OR "Reinforcement learning" OR "Machine learning" OR "Artifical
intelligence" OR "IA" OR "ML"))
AND NOT (*"Electronic Trading" OR " Investments" OR " Stock Market Analysis" OR*
"Financial Markets" OR "Commerce" OR "Stock Price" OR "Stock Market" OR "Stock Market
Prediction" OR "Stock Price Prediction" OR "Time Series Analysis" OR "Time? Series" OR
"Stock Predictions" OR "Risk Assessment" OR "Stock Market Forecasting" OR "Psychology" OR
"Financial Forecasting" OR "Stock Exchange" OR "Time-series Data" OR "Price-analysis" OR
"Stock Data" OR "Finance" OR "Building Stocks" OR "Aboveground Biomass" OR "Biomass"))

Para el caso de la base de datos “WOS” se formula una ecuación de búsqueda equivalente a la anterior, así:

Ecuación 2.

Ecuación de búsqueda en Scopus

(**TS** = (*item OR inventory OR stock OR "Stock keeping unit" OR sku OR abc*)
NEAR/1 (*classification OR categorization OR segmentation OR clustering*)
AND (*"Ensemble learning" OR "fuzzy classification" OR "Unsupervised learning" OR*
"unsupervised clustering" OR "Deep learning" OR "Reinforcement learning" OR "Machine
learning" OR "Artificial intelligence" OR IA OR ML))
NOT TS= (*"Electronic Trading" OR " Investments" OR " Stock Market Analysis" OR "Financial*
Markets" OR "Commerce" OR "Stock Price" OR "Stock Market" OR "Stock Market Prediction"
OR "Stock Price Prediction" OR "Time Series Analysis" OR "Time? Series" OR "Stock
Predictions" OR "Risk Assessment" OR "Stock Market Forecasting" OR "Psychology" OR
"Financial Forecasting" OR "Stock Exchange" OR "Time-series Data" OR "Price-analysis" OR
"Stock Data" OR "Finance" OR "Building Stocks" OR "Aboveground Biomass" OR "Biomass"))

Logrando recopilar 179 resultados de “Scopus” y 101 records de “WOS”, en donde se excluyeron 33 artículos duplicados, para un total de 247 artículos únicos de las fuentes principales.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Por otro lado, con respecto a las fuentes auxiliares se tuvieron en cuenta siete artículos de referencia. Teniendo en su totalidad, 254 documentos, para su posterior gestión y selección de estudios totalmente afines con el objeto de estudio en cuestión.

Fase IV: Análisis y selección de información

Así pues, para la selección de estudios, se parte de un primer filtrado en donde se tienen el título y el resumen como criterios para tener en cuenta, en donde la inclusión del registro se debe a si existe una verdadera relación de estos con el tema bajo estudio, es decir, si aporta a la clasificación de inventarios aplicando algoritmos de aprendizaje automático. Seguido de esto, se procede a la etapa de recuperación de los documentos para su posterior evaluación para su elegibilidad, la cual consiste en una revisión exhaustiva con la lectura completa del artículo o documento, con el fin de corroborar la coherencia de los estudios con la temática del presente proyecto.

Dentro del análisis y selección de la evidencia relevante para la revisión de literatura se hace uso del diagrama de la metodología PRISMA para organizar la información adecuadamente, permitiendo el filtrado de información y resumir de manera gráfica la evidencia, brindando al lector comprender rápidamente los procedimientos básicos utilizados en la búsqueda exhaustiva y examinar la criba para la selección de estudios a lo largo del proceso de revisión.

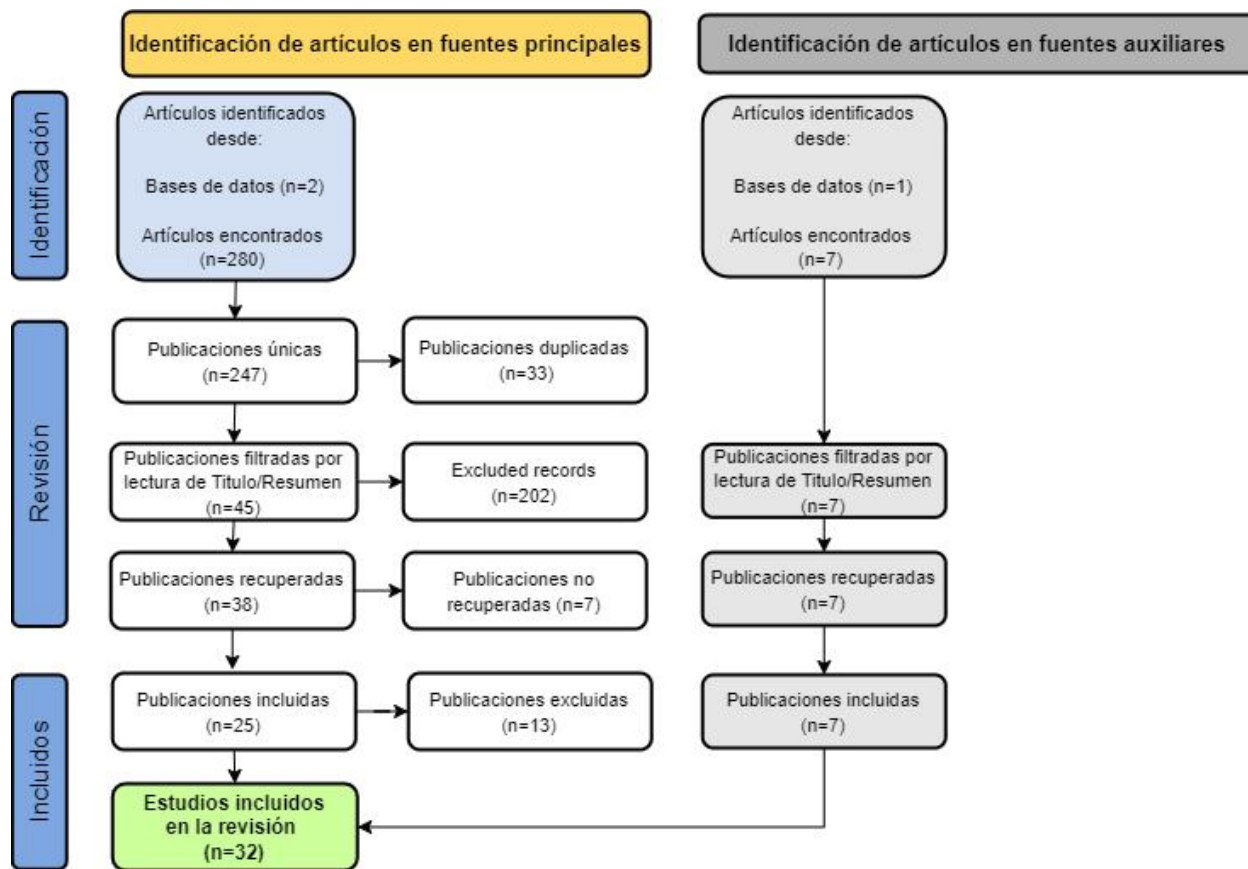
A partir del diagrama de flujo PRISMA se filtran los 247 registros hallados, así: 202 registros excluidos por filtrado de Title/Abstract; siete publicaciones de tipo “Conference Paper” no fueron recuperadas para su lectura completa, y 13 registros excluidos en la evaluación de elegibilidad; obteniendo 25 artículos incluidos por parte de las fuentes de datos principales, y que

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

en adición a las siete publicaciones de fuentes auxiliares, se tiene un total compilado de 32 estudios para la revisión como se observa en la Figura 2.

Figura 2.

Diagrama PRISMA



Nota. Diseño adaptado por el Autor, de los diagramas de flujo propuestos por Boers, Mayo-Wilson y Stovold et al.

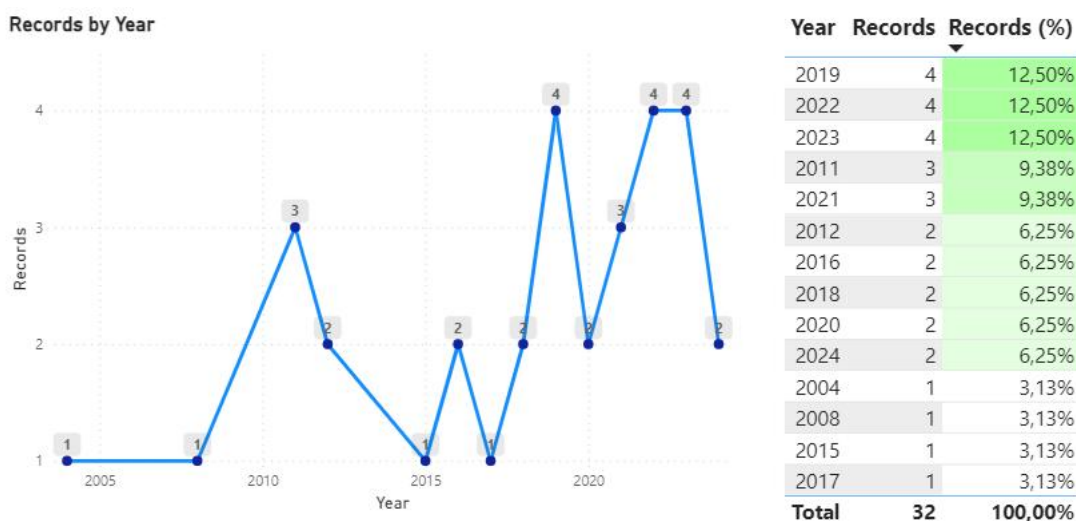
Con los estudios incluidos se procede a realizar la extracción de información relevante y el análisis bibliométrico para cerrar esta fase de análisis.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

En la Figura 3, se observa la frecuencia de publicación por año sobre aplicaciones de ML en clasificación de inventarios, donde se observa que los dos años previos (2022-2023) y en el año 2019 es donde más investigación se ha realizado en este campo con cuatro publicaciones.

Figura 3.

Frecuencia de Publicación por año



Del mismo modo, se destacan autores como Rimini B.; Qaffas, AA; Lolli F.; Jackson, I; Ishizaka A.; Gamberini R.; Ben Ncir, CE; Ben HajKacem, M.-A. y Balugani E. quienes han contribuido a esta área de investigación con dos publicaciones en “Journals”, seguido de los demás autores con un artículo (Ver Figura 4).

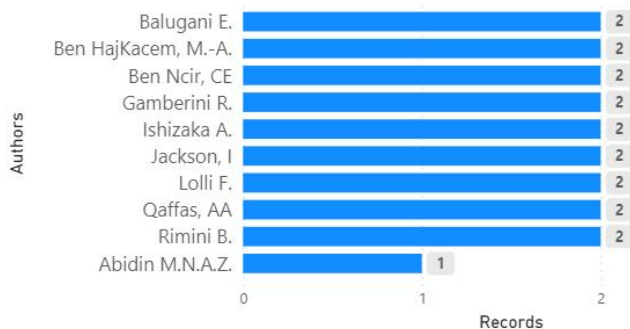
MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 4.

Publicaciones por

Autor

Top 10 Authors



Authors	Records	Records (%)
Balugani E.	2	2,06%
Ben HajKacem, M.-A.	2	2,06%
Ben Ncir, CE	2	2,06%
Gamberini R.	2	2,06%
Ishizaka A.	2	2,06%
Jackson, I	2	2,06%
Lolli F.	2	2,06%
Qaffas, AA	2	2,06%
Rimini B.	2	2,06%
Abidin M.N.A.Z.	1	1,03%

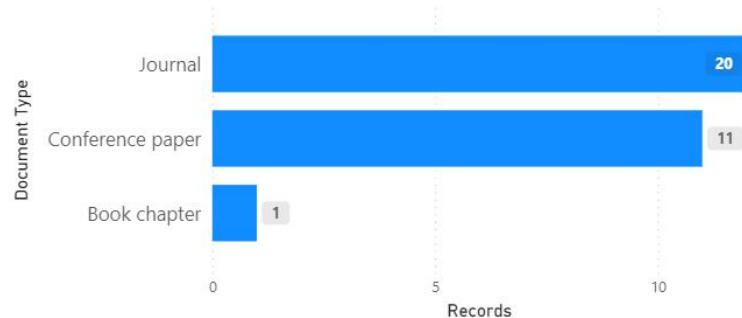
Para ver la distribución de artículos a nivel de tipo de documento, podemos observar a continuación, que el tipo “Journal” es el tipo de documento más utilizado con 20 publicaciones, lo cual representa el 62.5% del total. Seguido de “Conference Papers” con 11 records y finalmente solo una publicación de tipo “Book Chapter”.

Figura 5.

Publicaciones por tipo de

documento

Records by Document type



Document Type	Records	Records (%)
Journal	20	62,50%
Conference paper	11	34,38%
Book chapter	1	3,13%
Total	32	100,00%

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Dicho esto, se identificaron los top cinco publishers, donde se destaca el “Institute of Electrical and Electronics Engineers Inc.” con cuatro publicaciones (12.5%), seguido de “Elsevier Ltd”, “MDPI” y “Springer”, con dos publicaciones cada uno, como se ilustra en la Figura 6.

Figura 6.

Publicaciones por publisher

Top 5 Publishers



Publisher	Records	Records (%)
Institute of Electrical and Electronics Engineers Inc.	4	12,50%
Elsevier Ltd	2	6,25%
MDPI	2	6,25%
Springer Science and Business Media Deutschland GmbH	2	6,25%
Elsevier B.V.	1	3,13%

Fase V: Presentación de información

Comprende la sección del análisis de la literatura presentada en la segunda sección del documento de propuesta de investigación

Referencias

Sampieri Hernández, R. (2008). Metodología de la Investigación. En R. Sampieri Hernández, *La*

Idea: Nace un proyecto de investigación (págs. 30-39). México: Mc. Graw-Hill.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing

Evidence-Informed Management Knowledge by Means of Systematic Review. *British*

Journal of Management, 207-222.

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Apéndice B. Consultas SQL para la obtención de datos

Versión 1.

```

-----
WITH BASE AS (
-----
    SELECT DISTINCT
        $COUNTRY AS COUNTRY
        , $RETAILER_ID AS RETAILER_ID
-----
), SALES AS (
-----
    SELECT DISTINCT
        CURRENT_DATE AS SNAPSHOT
        , RP.COUNTRY
        , RP.RETAILER
        , RP.RETAILER_ID
        , RP.PRODUCT_ID
        , RP.SKU
        , RP.NAME
        , RP.UNITS_L90D AS UNITS
        , RP.ORDERS_L90D AS ORDERS
        , RP.TOTAL_PRICE_L90D AS SALES
        , RP.IS_TOP_SELLER
    FROM retailer_products RP
    INNER JOIN BASE B
        ON RP.COUNTRY = B.COUNTRY
        AND RP.RETAILER_ID = B.RETAILER_ID
    WHERE 1=1
        AND RP.STATUS_PRODUCT_ID = 'published'
        AND RP.ORDERS_L90D > 0
        AND RP.CATEGORY NOT IN ('pesables')
-----
), STOCKOUT AS (
-----
    SELECT DISTINCT
        T.COUNTRY
        , T.CP_BRAND_ID AS RETAILER_ID
        , T.PRODUCT_ID
        , ROUND(SUM(CASE WHEN (T.FINAL_STATE IN ('stockout','substituted'))
            OR (T.SKU_COUNT = 1 AND T.FINAL_STATE = 'cso')
            OR (FINAL_STATE = 'cso' AND STOCKOUT = 1)
            THEN T.TOTAL_PRICE ELSE 0 END)) AS LOST_SALES
        , ROUND(SUM(CASE WHEN (T.FINAL_STATE IN ('stockout','substituted'))
            OR (T.SKU_COUNT = 1 AND T.FINAL_STATE = 'cso')
            OR (FINAL_STATE = 'cso' AND STOCKOUT = 1)
            THEN T.UNITS ELSE 0 END)) AS UNITS_STOCKOUT
        , ROUND(COUNT(CASE WHEN (T.FINAL_STATE IN ('stockout','substituted'))
            OR (T.SKU_COUNT = 1 AND T.FINAL_STATE = 'cso')
            OR (FINAL_STATE = 'cso' AND STOCKOUT = 1)
            THEN T.ORDER_ID ELSE 0 END)) AS ORDERS_STOCKOUT
        , AVG(T.CREATE_STOCK) AS AVG_STOCK_LEVEL
        , MEDIAN(T.CREATE_STOCK) AS MEDIAN_STOCK_LEVEL
    FROM tipificador T
    INNER JOIN BASE B
        ON T.COUNTRY = B.COUNTRY
        AND T.CP_BRAND_ID = B.RETAILER_ID
    WHERE 1=1
        AND T.DAY::DATE >= (CURRENT_DATE - 90)
    GROUP BY ALL
    HAVING UNITS_STOCKOUT > 0
-----
), TIME_REPOSITION AS (
-----
    //-----
    WITH ORDER_EVENTS AS (
    //-----
        SELECT DISTINCT

```

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

```

        T.COUNTRY
        , T.CP_BRAND_ID AS RETAILER_ID
        , T.PRODUCT_ID
        , T.CREATED_AT_ORDER
        , CASE WHEN (T.FINAL_STATE IN ('stockout','substituted'))
                OR (T.SKU_COUNT = 1 AND T.FINAL_STATE = 'cso')
                OR (T.FINAL_STATE = 'cso' AND T.STOCKOUT = 1)
                THEN TRUE ELSE FALSE END AS IS_STOCKOUT
        , ROW_NUMBER() OVER (PARTITION BY T.COUNTRY, T.PRODUCT_ID ORDER BY T.CREATED_AT_ORDER)
AS ROW_NUM
FROM tipificador T
  INNER JOIN BASE B
    ON T.COUNTRY = B.COUNTRY
    AND T.CP_BRAND_ID = B.RETAILER_ID
WHERE 1=1
  AND T.DAY::DATE >= (CURRENT_DATE - 90)
  ORDER BY CREATED_AT_ORDER DESC
//-----
), JOIN_EVENTS AS (
//-----
  SELECT
    O1.COUNTRY
    , O1.RETAILER_ID
    , O1.PRODUCT_ID
    , O1.CREATED_AT_ORDER AS STOCKOUT_START
    , O2.CREATED_AT_ORDER AS RESTOCK_TIME
    , DATEDIFF('DAY', O1.CREATED_AT_ORDER, O2.CREATED_AT_ORDER) AS DAYS_TO_RESTOCK
  FROM ORDER_EVENTS O1
    INNER JOIN ORDER_EVENTS O2
      ON O1.PRODUCT_ID = O2.PRODUCT_ID
      AND O1.ROW_NUM + 1 = O2.ROW_NUM
      AND O1.IS_STOCKOUT = TRUE
      AND O2.IS_STOCKOUT = FALSE
//-----
)
//-----
SELECT DISTINCT
  COUNTRY
  , RETAILER_ID
  , PRODUCT_ID
  , AVG(DAYS_TO_RESTOCK) AS AVG_DAYS_TO_RESTOCK
FROM JOIN_EVENTS
GROUP BY ALL
-----
)
-----
SELECT DISTINCT
  S.*
  , COALESCE(SO.UNITS_STOCKOUT, 0) AS UNITS_STOCKOUT
  , COALESCE(SO.ORDERS_STOCKOUT, 0) AS ORDERS_STOCKOUT
  , COALESCE(SO.LOST_SALES, 0) AS LOST_SALES
  , (S.UNITS + COALESCE(SO.UNITS_STOCKOUT, 0)) AS UNITS_DEMAND
  , ROUND(COALESCE(SO.AVG_STOCK_LEVEL, 0), 2) AS AVG_STOCK_LEVEL
  , ROUND(COALESCE(SO.MEDIAN_STOCK_LEVEL, 0), 2) AS MEDIAN_STOCK_LEVEL
  , ROUND(COALESCE(TR.AVG_DAYS_TO_RESTOCK, 0), 2) AS AVG_DAYS_TO_RESTOCK
FROM SALES S
  LEFT JOIN STOCKOUT SO
    ON S.COUNTRY = SO.COUNTRY
    AND S.RETAILER_ID = SO.RETAILER_ID
    AND S.PRODUCT_ID = SO.PRODUCT_ID
  LEFT JOIN TIME_REPOSITION TR
    ON S.COUNTRY = TR.COUNTRY
    AND S.RETAILER_ID = TR.RETAILER_ID
    AND S.PRODUCT_ID = TR.PRODUCT_ID
WHERE 1=1
;

```

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Versión 2.

 WITH BASE AS (

```
SELECT DISTINCT
  $COUNTRY AS COUNTRY
  , $RETAILER_ID AS RETAILER_ID
```

), SALES AS (

```
SELECT DISTINCT
  CURRENT_DATE AS SNAPSHOT
  , RP.COUNTRY
  , $OWNERSHIP AS OWNERSHIP
  , $ID AS ID
  , RP.RETAILER
  , RP.RETAILER_ID
  , RP.PRODUCT_ID
  , RP.SKU
  , RP.NAME
FROM retailer_product RP
  INNER JOIN BASE B
    ON RP.COUNTRY = B.COUNTRY
    AND RP.RETAILER_ID = B.RETAILER_ID
WHERE 1=1
  AND RP.STATUS_PRODUCT_ID = 'published'
  AND RP.ORDERS_L90D > 0
  AND RP.CATEGORY_ID NOT IN ('pesables')
```

), STOCKOUT AS (

```
SELECT DISTINCT
  T.COUNTRY
  , T.CP_BRAND_ID AS RETAILER_ID
  , T.PRODUCT_ID
  , ROUND(SUM(CASE WHEN (T.FINAL_STATE IN ('stock'))
    THEN T.TOTAL_PRICE ELSE 0 END)) AS SALES
  , ROUND(COUNT(CASE WHEN (T.FINAL_STATE IN ('stock'))
    THEN T.ORDER_ID ELSE 0 END)) AS ORDERS
  , ROUND(SUM(CASE WHEN (T.FINAL_STATE IN ('stock'))
    THEN T.UNITS ELSE 0 END)) AS UNITS
  , STDDEV(CASE WHEN (
    (T.FINAL_STATE IN ('stock','stockout','substituted'))
    OR (T.SKU_COUNT = 1 AND T.FINAL_STATE = 'cso')
    OR (FINAL_STATE = 'cso' AND STOCKOUT = 1)
  )
    THEN T.UNITS ELSE 0 END) AS STDDEV_DEMAND
  , ROUND(SUM(CASE WHEN (T.FINAL_STATE IN ('stockout','substituted'))
    OR (T.SKU_COUNT = 1 AND T.FINAL_STATE = 'cso')
    OR (FINAL_STATE = 'cso' AND STOCKOUT = 1)
    THEN T.TOTAL_PRICE ELSE 0 END)) AS LOST_SALES
  , ROUND(SUM(CASE WHEN (T.FINAL_STATE IN ('stockout','substituted'))
    OR (T.SKU_COUNT = 1 AND T.FINAL_STATE = 'cso')
    OR (FINAL_STATE = 'cso' AND STOCKOUT = 1)
    THEN T.UNITS ELSE 0 END)) AS UNITS_STOCKOUT
  , ROUND(COUNT(CASE WHEN (T.FINAL_STATE IN ('stockout','substituted'))
    OR (T.SKU_COUNT = 1 AND T.FINAL_STATE = 'cso')
    OR (FINAL_STATE = 'cso' AND STOCKOUT = 1)
    THEN T.ORDER_ID ELSE 0 END)) AS ORDERS_STOCKOUT
  , AVG(T.CREATE_STOCK) AS AVG_STOCK_LEVEL
  , MEDIAN(T.CREATE_STOCK) AS MEDIAN_STOCK_LEVEL
FROM tipificador T
  INNER JOIN BASE B
    ON T.COUNTRY = B.COUNTRY
    AND T.CP_BRAND_ID = B.RETAILER_ID
WHERE 1=1
  AND T.DAY::DATE >= (CURRENT_DATE - 365)
```

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

```

GROUP BY ALL
HAVING UNITS_STOCKOUT > 0
-----
), TIME_REPOSITION AS (
-----
//-----
WITH ORDER_EVENTS AS (
//-----
    SELECT DISTINCT
        T.COUNTRY
        , T.CP_BRAND_ID AS RETAILER_ID
        , T.PRODUCT_ID
        , T.CREATED_AT_ORDER
        , CASE WHEN (T.FINAL_STATE IN ('stockout','substituted'))
            OR (T.SKU_COUNT = 1 AND T.FINAL_STATE = 'cso')
            OR (T.FINAL_STATE = 'cso' AND T.STOCKOUT = 1)
            THEN TRUE ELSE FALSE END AS IS_STOCKOUT
        , ROW_NUMBER() OVER (PARTITION BY T.COUNTRY, T.PRODUCT_ID ORDER BY T.CREATED_AT_ORDER)
AS ROW_NUM
FROM tipificador T
INNER JOIN BASE B
    ON T.COUNTRY = B.COUNTRY
    AND T.CP_BRAND_ID = B.RETAILER_ID
WHERE 1=1
    AND T.DAY::DATE >= (CURRENT_DATE - 365)
ORDER BY CREATED_AT_ORDER DESC
//-----
), JOIN_EVENTS AS (
//-----
    SELECT
        O1.COUNTRY
        , O1.RETAILER_ID
        , O1.PRODUCT_ID
        , O1.CREATED_AT_ORDER AS STOCKOUT_START
        , O2.CREATED_AT_ORDER AS RESTOCK_TIME
        , DATEDIFF('DAY', O1.CREATED_AT_ORDER, O2.CREATED_AT_ORDER) AS DAYS_TO_RESTOCK
FROM ORDER_EVENTS O1
INNER JOIN ORDER_EVENTS O2
    ON O1.PRODUCT_ID = O2.PRODUCT_ID
    AND O1.ROW_NUM + 1 = O2.ROW_NUM
    AND O1.IS_STOCKOUT = TRUE
    AND O2.IS_STOCKOUT = FALSE
//-----
)
//-----
SELECT DISTINCT
    COUNTRY
    , RETAILER_ID
    , PRODUCT_ID
    , AVG(DAYS_TO_RESTOCK) AS AVG_DAYS_TO_RESTOCK
FROM JOIN_EVENTS
GROUP BY ALL
-----
)
-----
SELECT DISTINCT
    S.* EXCLUDE(RETAILER, RETAILER_ID, SKU, NAME)
    , COALESCE(SO.UNITS, 0) AS UNITS
    , COALESCE(SO.ORDER, 0) AS ORDERS
    , COALESCE(SO.SALES, 0) AS SALES

    , COALESCE(SO.UNITS_STOCKOUT, 0) AS UNITS_STOCKOUT
    , COALESCE(SO.ORDER_STOCKOUT, 0) AS ORDER_STOCKOUT
    , IFF(COALESCE(SO.LOST_SALES, 0) = 0, (DIV0(COALESCE(SO.SALES, 0), COALESCE(SO.UNITS, 0)) * 0.70),
SO.LOST_SALES) AS UNITS_STOCKOUT_COST

    , (SO.UNITS + COALESCE(SO.UNITS_STOCKOUT, 0)) AS UNITS_DEMAND
    , DIV0(UNITS_DEMAND, 365) AS AVG_DAILY_DEMAND

    , IFF(SO.STDDEV_DEMAND = 0, AVG_DAILY_DEMAND * 0.2, SO.STDDEV_DEMAND) AS STDDEV_DEMAND_FINAL

```

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

```

, (DIV0(COALESCE(SO.SALES, 0), COALESCE(SO.UNITS, 0))* 0.35) AS UNIT_COST
, ROUND(COALESCE(SO.AVG_STOCK_LEVEL, 0),2) AS AVG_STOCK_LEVEL
, ROUND(COALESCE(SO.MEDIAN_STOCK_LEVEL, 0),2) AS MEDIAN_STOCK_LEVEL

, IFF(ROUND(COALESCE(TR.AVG_DAYS_TO_RESTOCK, 0), 2)) = 0, 15, ROUND(TR.AVG_DAYS_TO_RESTOCK,2)) AS
LEAD_TIME_DAYS
, (STDDEV_DEMAND_FINAL * SQRT(LEAD_TIME_DAYS)) AS STDDEV_LEAD_TIME
FROM SALES S
LEFT JOIN STOCKOUT SO
  ON S.COUNTRY = SO.COUNTRY
  AND S.RETAILER_ID = SO.RETAILER_ID
  AND S.PRODUCT_ID = SO.PRODUCT_ID
LEFT JOIN TIME_REPOSITION TR
  ON S.COUNTRY = TR.COUNTRY
  AND S.RETAILER_ID = TR.RETAILER_ID
  AND S.PRODUCT_ID = TR.PRODUCT_ID
WHERE 1=1
AND SO.SALES > 0
;

```

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Apéndice C. Análisis Exploratorio de los Datos (EDA)

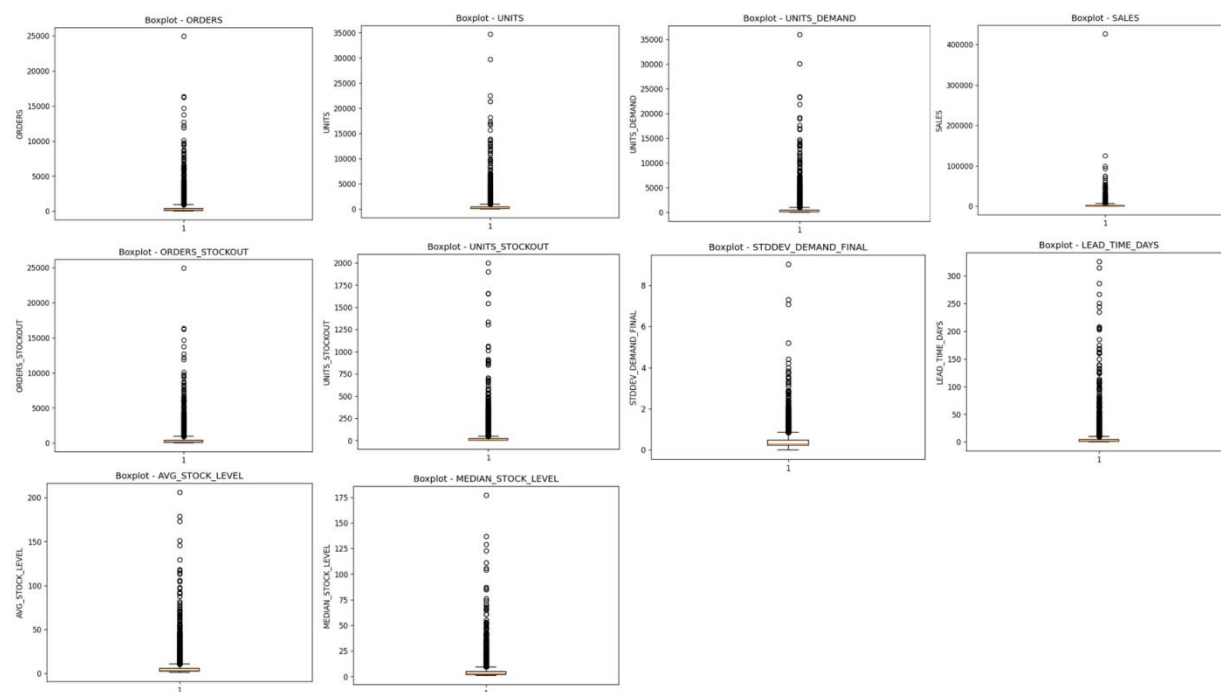
Retailer 1

Tabla 1.

Estadísticas Descriptivas Retailer 1

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	8066	470.09	1226.93	1	61	169	424	34693
ORDERS	8066	411.70	909.80	2	64	168	403	24940
SALES	8066	2668.30	6876.38	1	369	1084	2810	426428
UNITS_STOCKOUT	8066	27.71	82.24	1	2	7	20	2000
ORDERS_STOCKOUT	8066	411.70	909.80	2	64	168	403	24940
UNITS_DEMAND	8066	497.80	1288.50	2	66	179	449	36001
STDDEV_DEMAND_FINAL	8066	0.41	0.39	0	0	0	0	9
AVG_STOCK_LEVEL	8066	6.19	9.04	1	3	4	6	206
MEDIAN_STOCK_LEVEL	8066	5.33	7.20	1	2	3	5	177
LEAD_TIME_DAYS	8066	5.35	14.92	0	1	2	4	326

Figura 1.

Boxplots Retailer 1

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 2.

Histogramas Retailer 1

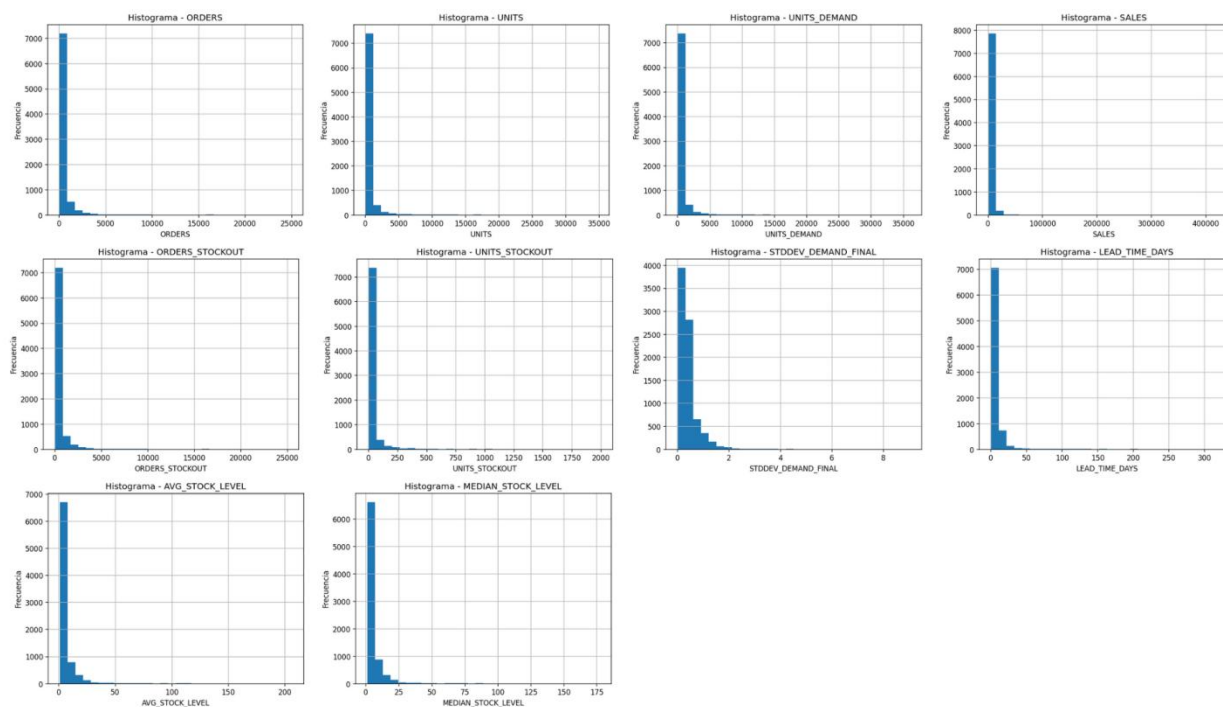
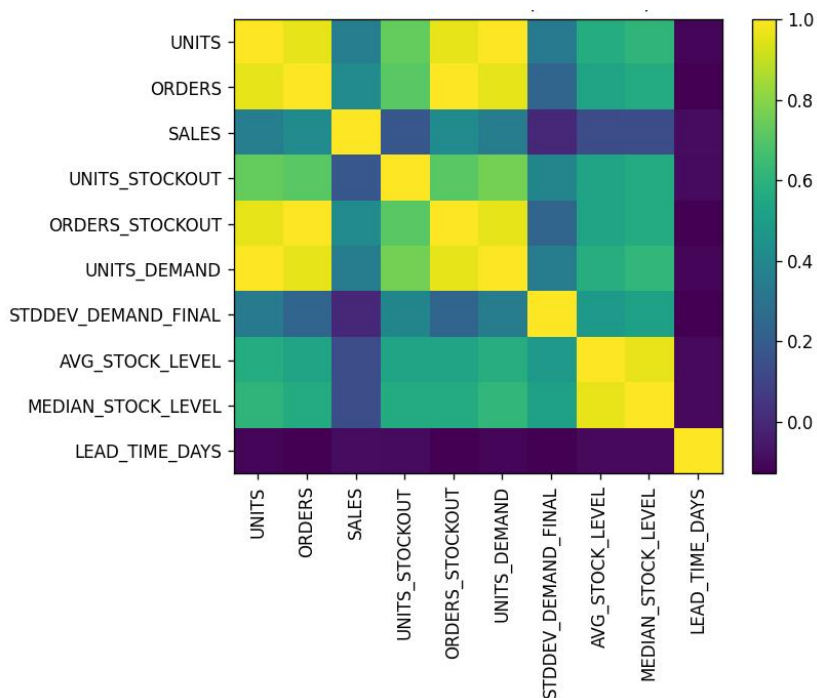


Figura 3.

Mapa de calor: Matriz de Correlación Retailer 1



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 2.*Detección de outliers (z-score) Retailer 1*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	127	1.57 %
LEAD_TIME_DAYS	108	1.34 %
MEDIAN_STOCK_LEVEL	130	1.61 %
ORDERS	131	1.62 %
ORDERS_STOCKOUT	131	1.62 %
SALES	61	0.76 %
STDDEV_DEMAND_FINAL	126	1.56 %
UNITS	114	1.41 %
UNITS_DEMAND	117	1.45 %
UNITS_STOCKOUT	129	1.60 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

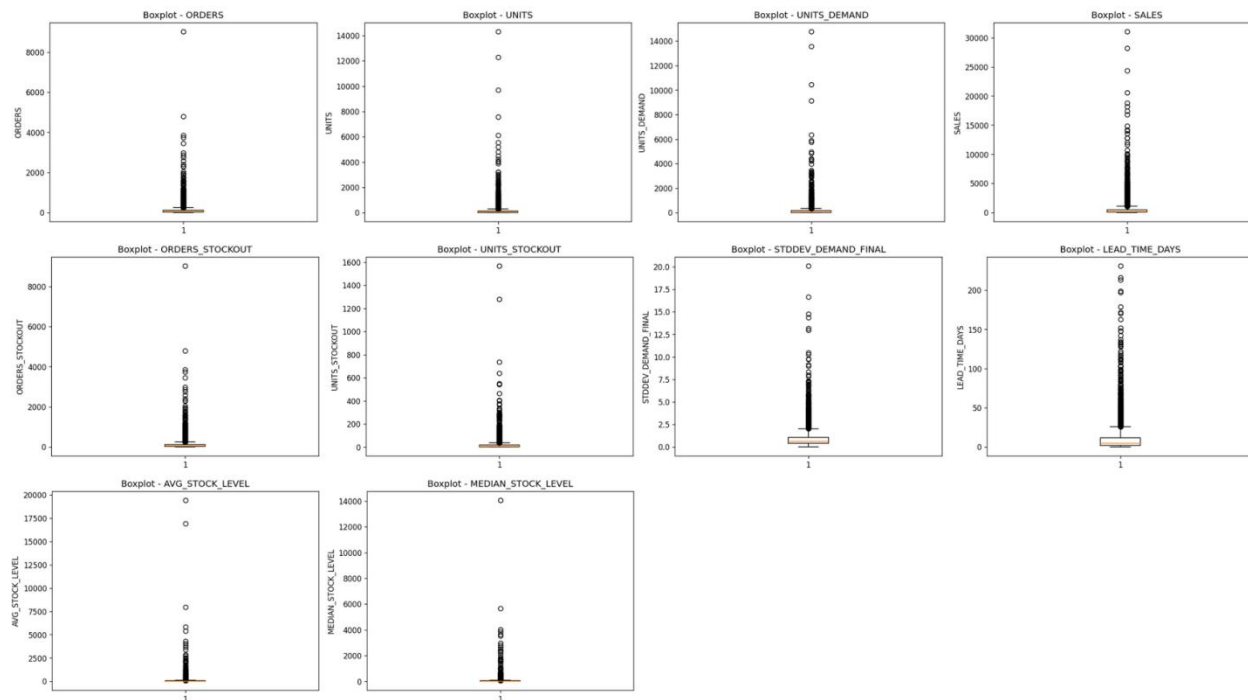
Retailer 2

Tabla 3.

Estadísticas Descriptivas Retailer 2

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	6551	138.21	404.91	1	21	52	125	14316
ORDERS	6551	108.74	239.61	2	21	49	109	9023
SALES	6551	530.87	1300.13	1	67	179	477	31048
UNITS_STOCKOUT	6551	16.59	41.97	1	3	6	16	1568
ORDERS_STOCKOUT	6551	108.74	239.61	2	21	49	109	9023
UNITS_DEMAND	6551	154.79	436.96	2	26	61	144	14782
STDDEV_DEMAND_FINAL	6551	0.89	1.01	0	0	1	1	20
AVG_STOCK_LEVEL	6551	70.14	392.05	5	15	25	49	19418
MEDIAN_STOCK_LEVEL	6551	51.95	244.66	5	13	22	41	14065
LEAD_TIME_DAYS	6551	9.70	15.75	0	2	5	12	231

Figura 5.

Boxplots Retailer 2

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 6.

Histogramas Retailer 2

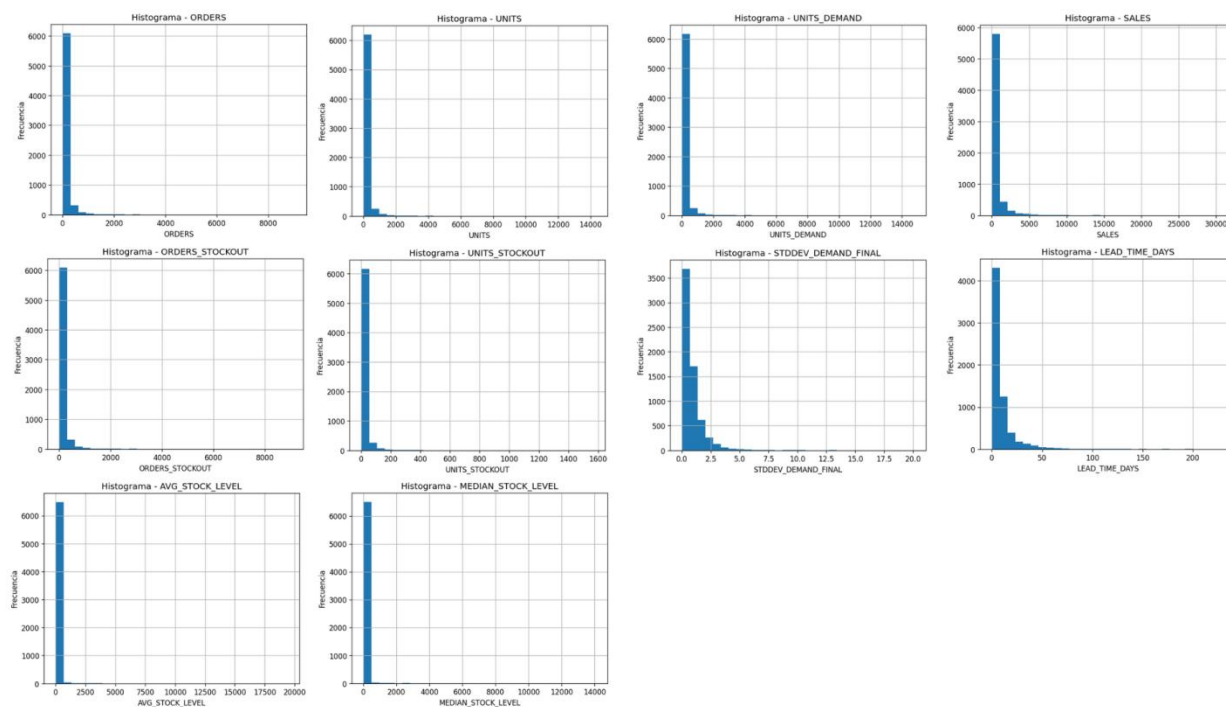
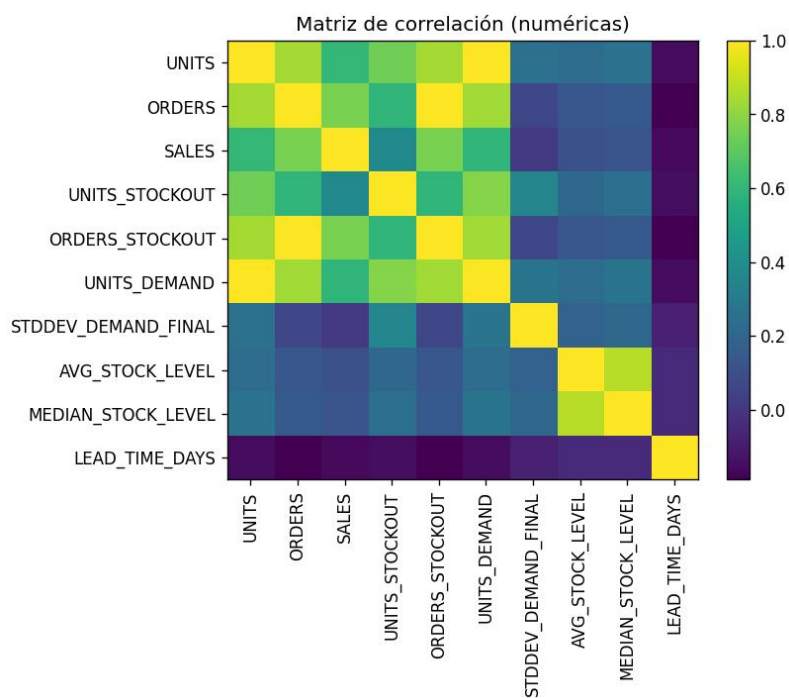


Figura 7.

Mapa de calor: Matriz de Correlación Retailer 2



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 4.*Detección de outliers (z-score) Retailer 2*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	33	0.50 %
LEAD_TIME_DAYS	131	2.00 %
MEDIAN_STOCK_LEVEL	33	0.50 %
ORDERS	89	1.36 %
ORDERS_STOCKOUT	89	1.36 %
SALES	111	1.69 %
STDDEV_DEMAND_FINAL	107	1.63 %
UNITS	61	0.93 %
UNITS_DEMAND	62	0.95 %
UNITS_STOCKOUT	80	1.22 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

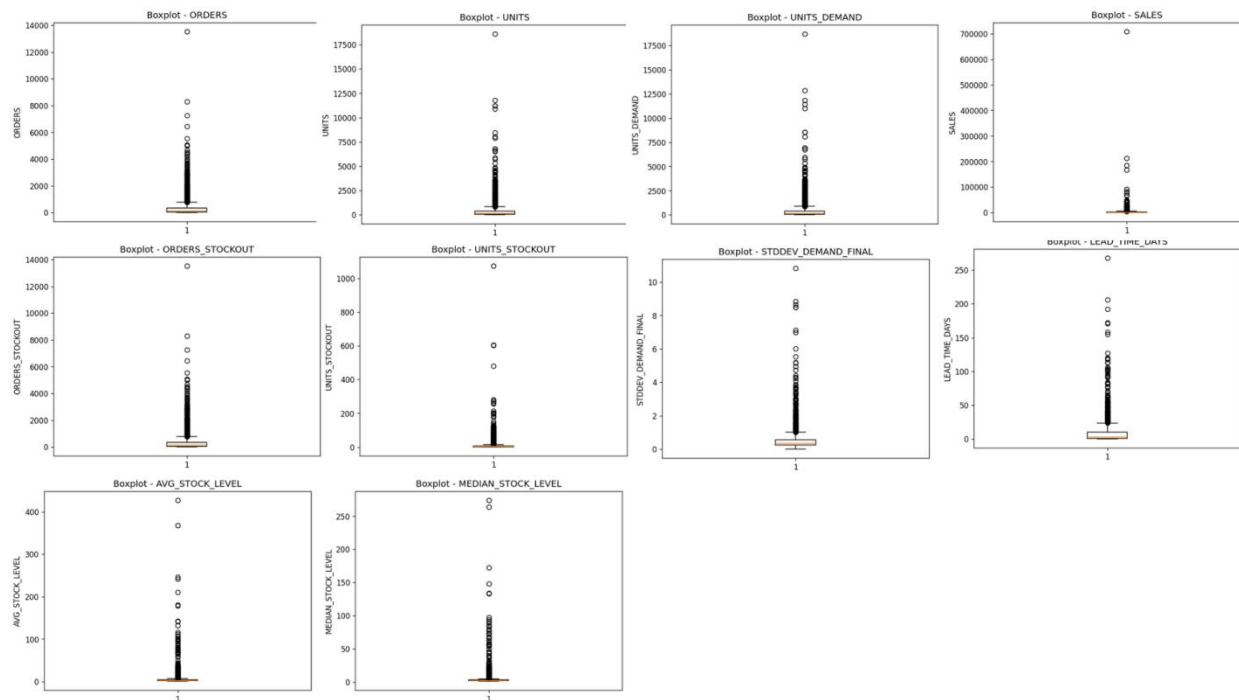
Retailer 3

Tabla 5.

Estadísticas Descriptivas Retailer 3

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	4157	366.82	772.10	1	57	149	374	18615
ORDERS	4157	318.27	577.43	2	56	140	348	13532
SALES	4157	2679.61	13015.24	1	244	829	2510	708336
UNITS_STOCKOUT	4157	8.86	29.01	1	1	3	7	1074
ORDERS_STOCKOUT	4157	318.27	577.43	2	56	140	348	13532
UNITS_DEMAND	4157	375.69	789.48	2	60	153	384	18704
STDDEV_DEMAND_FINAL	4157	0.49	0.58	0	0	0	1	11
AVG_STOCK_LEVEL	4157	5.12	14.26	1	2	2	4	427
MEDIAN_STOCK_LEVEL	4157	3.80	9.71	1	2	2	3	274
LEAD_TIME_DAYS	4157	7.49	14.40	0	1	3	10	268

Figura 8.

Boxplots Retailer 3

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 6.*Detección de outliers (z-score) Retailer 3*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	38	0.91 %
LEAD_TIME_DAYS	71	1.71 %
MEDIAN_STOCK_LEVEL	37	0.89 %
ORDERS	81	1.95 %
ORDERS_STOCKOUT	81	1.95 %
SALES	15	0.36 %
STDDEV_DEMAND_FINAL	68	1.64 %
UNITS	73	1.76 %
UNITS_DEMAND	72	1.73 %
UNITS_STOCKOUT	40	0.96 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Retailer 4

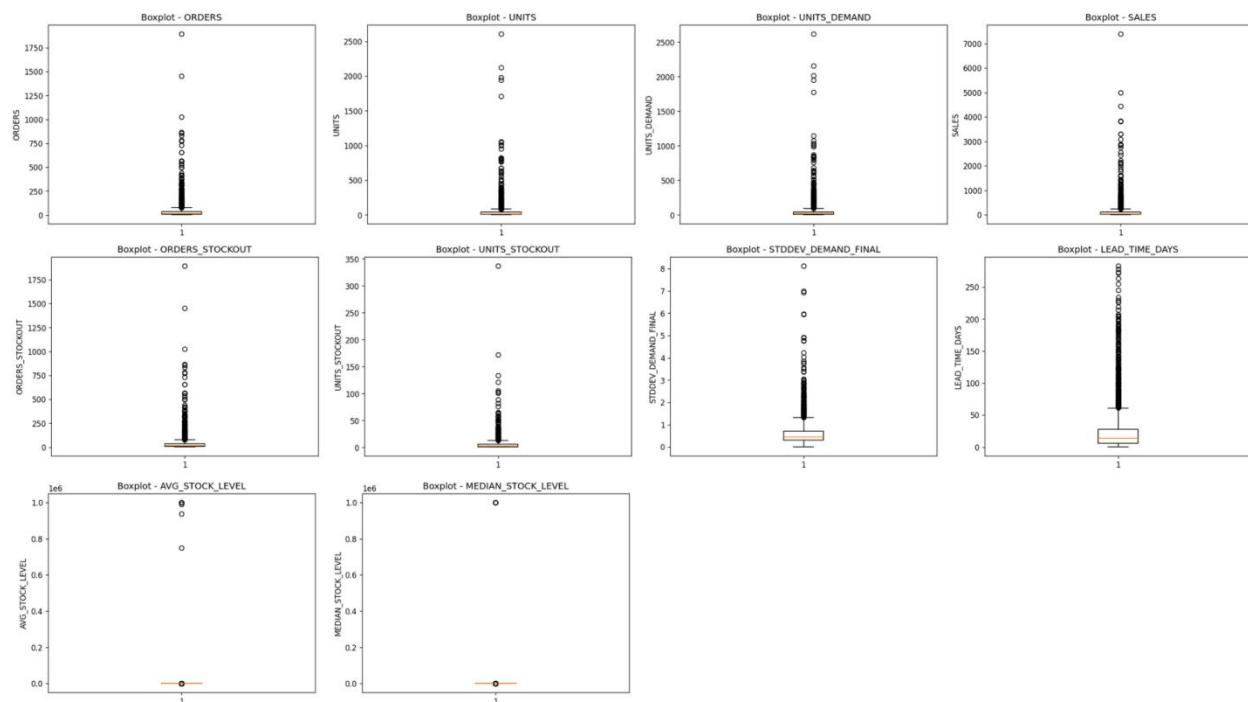
Tabla 7.

Estadísticas Descriptivas Retailer 4

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	2718	45.45	126.43	1	7	16	38	2609
ORDERS	2718	39.39	85.16	2	9	18	37	1895
SALES	2718	130.59	337.28	1	21	49	112	7398
UNITS_STOCKOUT	2718	5.90	11.59	1	1	3	6	337
ORDERS_STOCKOUT	2718	39.39	85.16	2	9	18	37	1895
UNITS_DEMAND	2718	51.35	130.96	2	10	21	45	2620
STDDEV_DEMAND_FINAL	2718	0.59	0.61	0	0	0	1	8
AVG_STOCK_LEVEL	2718	5097.00	70098.83	3	18	30	61	999999
MEDIAN_STOCK_LEVEL	2718	5201.59	71593.81	3	16	27	50	999999
LEAD_TIME_DAYS	2718	25.63	36.17	0	6	14	28	283

Figura 11.

Boxplots Retailer 4



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 12.

Histogramas Retailer 4

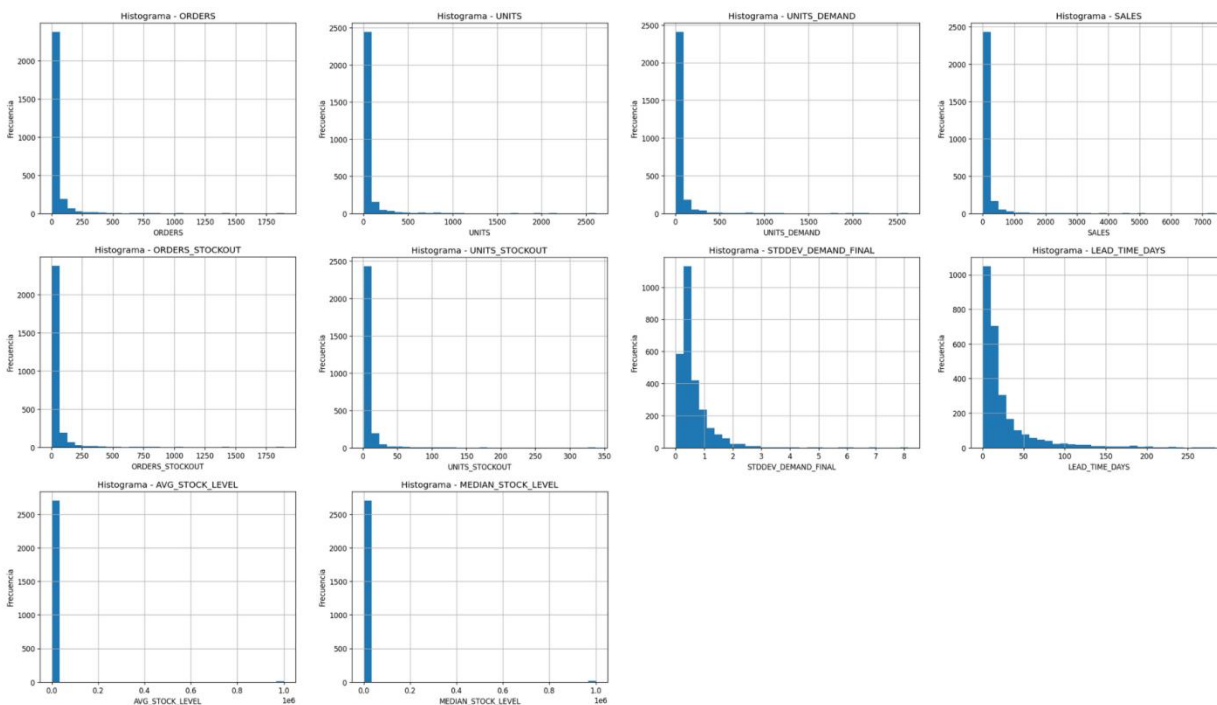
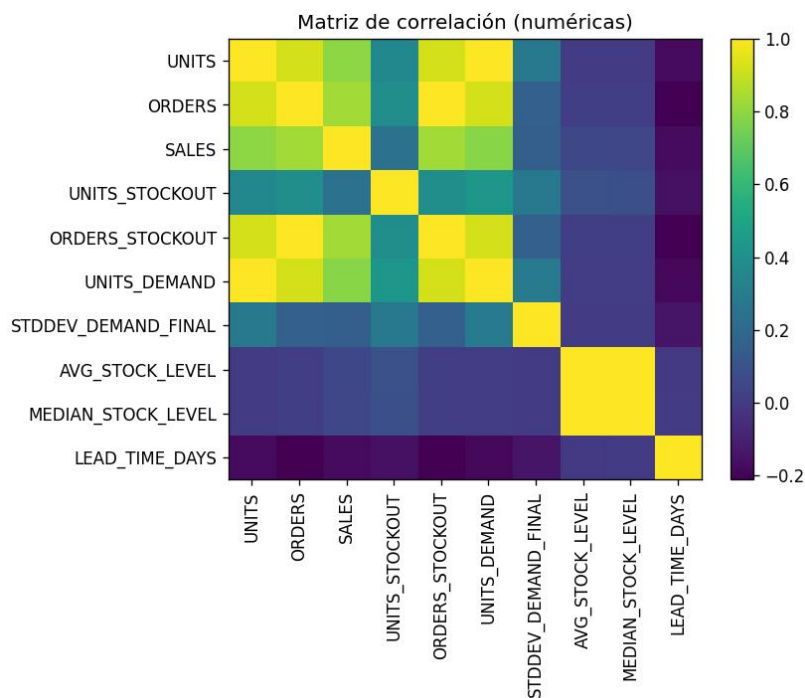


Figura 13.

Mapa de calor: Matriz de Correlación Retailer 4



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 8.*Detección de outliers (z-score) Retailer 4*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	14	0.52 %
LEAD_TIME_DAYS	76	2.80 %
MEDIAN_STOCK_LEVEL	14	0.52 %
ORDERS	45	1.66 %
ORDERS_STOCKOUT	45	1.66 %
SALES	41	1.51 %
STDDEV_DEMAND_FINAL	42	1.55 %
UNITS	38	1.40 %
UNITS_DEMAND	37	1.36 %
UNITS_STOCKOUT	36	1.32 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Retailer 5

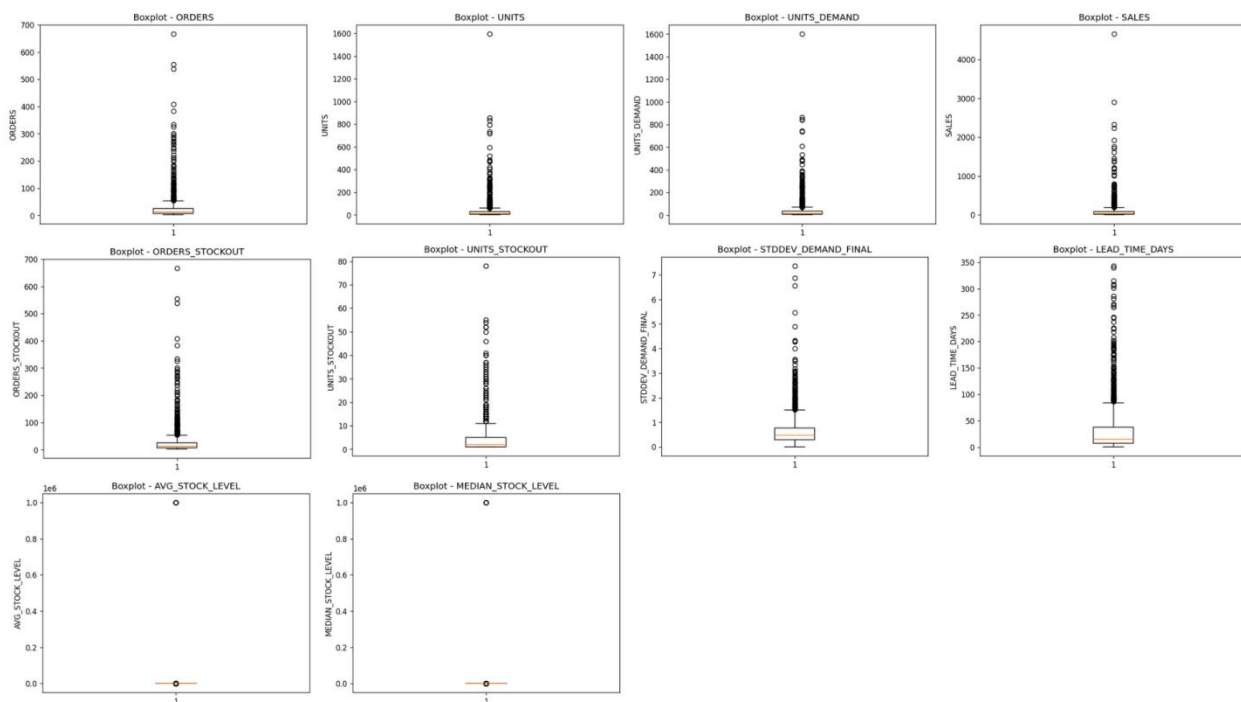
Tabla 9.

Estadísticas Descriptivas Retailer 5

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	1450	33.18	81.92	1	6	13	28	1596
ORDERS	1450	26.39	46.97	2	7	13	26	667
SALES	1450	92.43	232.26	1	16	34	87	4658
UNITS_STOCKOUT	1450	4.16	6.21	1	1	2	5	78
ORDERS_STOCKOUT	1450	26.39	46.97	2	7	13	26	667
UNITS_DEMAND	1450	37.34	84.60	2	8	16	33	1602
STDDEV_DEMAND_FINAL	1450	0.65	0.70	0	0	0	1	7
AVG_STOCK_LEVEL	1450	7649.75	86792.19	2	17	29	63	999999
MEDIAN_STOCK_LEVEL	1450	7635.04	86793.44	1	15	25	50	999999
LEAD_TIME_DAYS	1450	32.43	44.94	0	7	15	38	343

Figura 14.

Boxplots Retailer 5



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 10.*Detección de outliers (z-score) Retailer 5*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	11	0.76 %
LEAD_TIME_DAYS	39	2.69 %
MEDIAN_STOCK_LEVEL	11	0.76 %
ORDERS	28	1.93 %
ORDERS_STOCKOUT	28	1.93 %
SALES	18	1.24 %
STDDEV_DEMAND_FINAL	25	1.72 %
UNITS	24	1.66 %
UNITS_DEMAND	25	1.72 %
UNITS_STOCKOUT	31	2.14 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Retailer 6

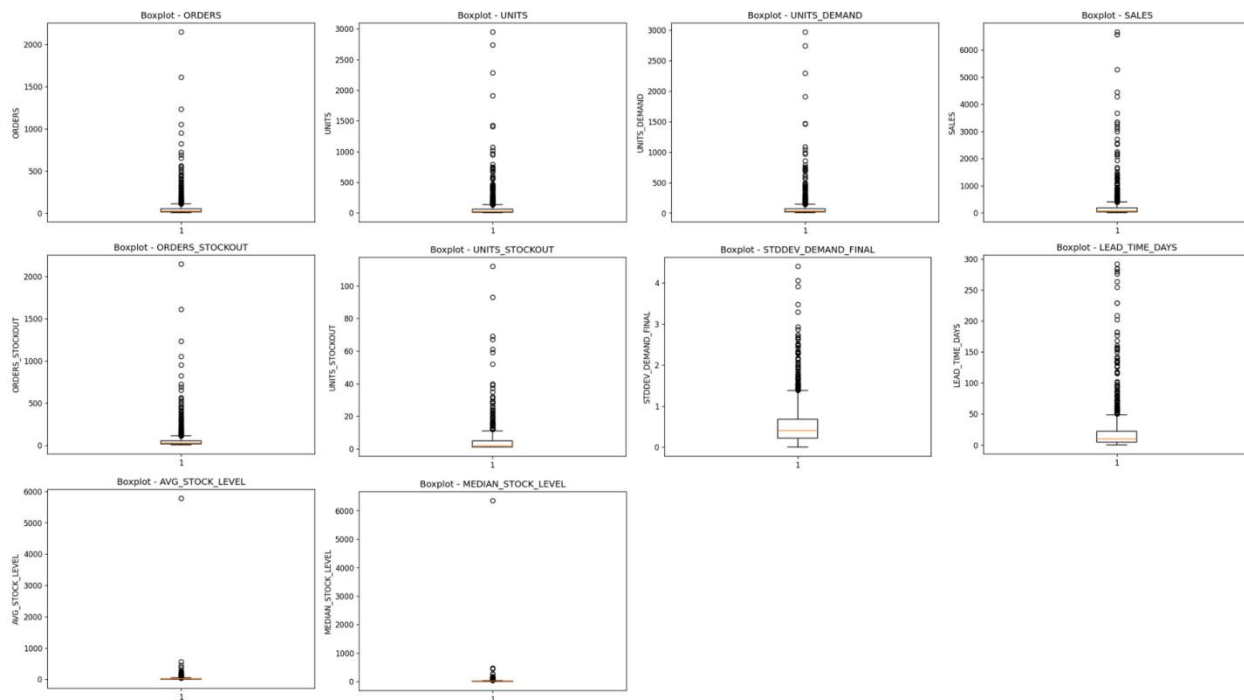
Tabla 11.

Estadísticas Descriptivas Retailer 6

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	1201	71.00	187.17	1	12	27	60	2951
ORDERS	1201	55.24	120.75	2	13	25	52	2150
SALES	1201	204.35	503.96	1	32	75	179	6663
UNITS_STOCKOUT	1201	4.59	7.37	1	1	2	5	112
ORDERS_STOCKOUT	1201	55.24	120.75	2	13	25	52	2150
UNITS_DEMAND	1201	75.59	190.12	2	15	30	66	2973
STDDEV_DEMAND_FINAL	1201	0.53	0.54	0	0	0	1	4
AVG_STOCK_LEVEL	1201	29.45	170.33	2	9	15	25	5780
MEDIAN_STOCK_LEVEL	1201	25.47	185.06	1	8	13	22	6353
LEAD_TIME_DAYS	1201	21.50	34.93	0	4	10	22	292

Figura 17.

Boxplots Retailer 6



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 12.*Detección de outliers (z-score) Retailer 6*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	2	0.17 %
LEAD_TIME_DAYS	33	2.75 %
MEDIAN_STOCK_LEVEL	1	0.08 %
ORDERS	19	1.58 %
ORDERS_STOCKOUT	19	1.58 %
SALES	20	1.67 %
STDDEV_DEMAND_FINAL	29	2.41 %
UNITS	18	1.50 %
UNITS_DEMAND	19	1.58 %
UNITS_STOCKOUT	19	1.58 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Retailer 7

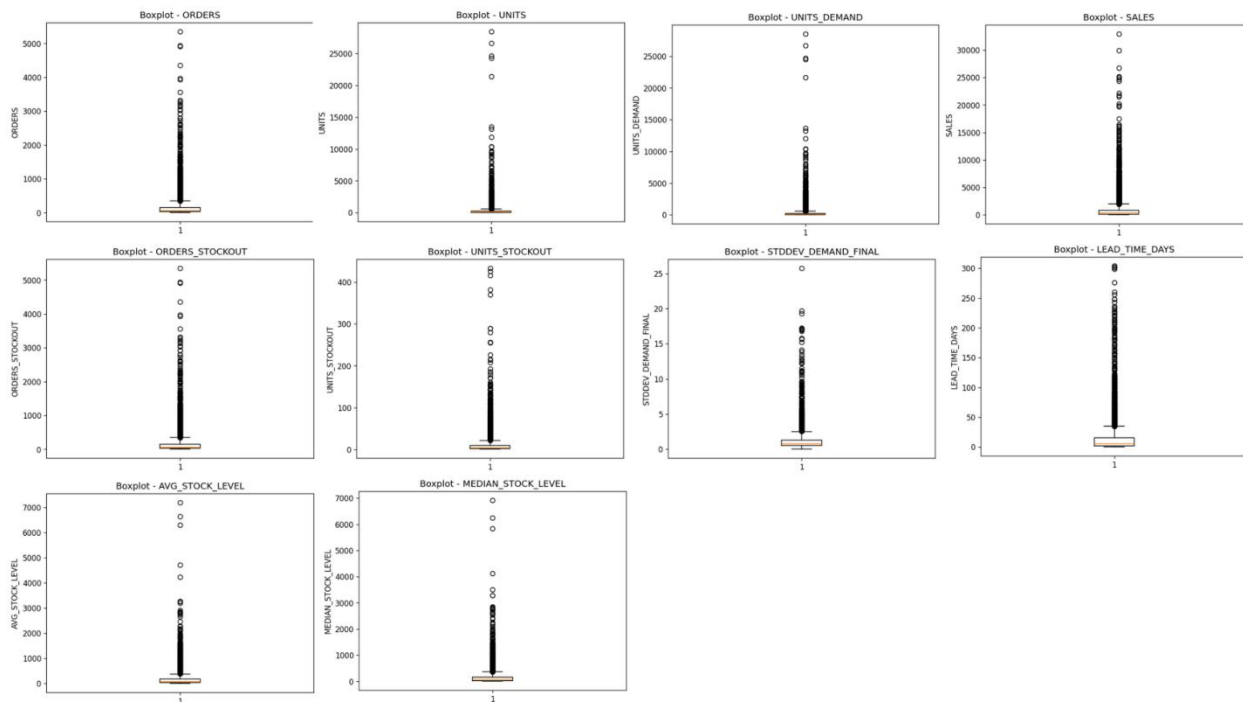
Tabla 13.

Estadísticas Descriptivas Retailer 7

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	6698	316.97	1015.32	1	33	87	248	28473
ORDERS	6698	159.99	320.27	2	26	60	155	5350
SALES	6698	909.56	1934.17	1	112	305	856	32936
UNITS_STOCKOUT	6698	10.62	23.30	1	2	4	10	433
ORDERS_STOCKOUT	6698	159.99	320.27	2	26	60	155	5350
UNITS_DEMAND	6698	327.59	1025.19	2	38	93	261	28526
STDDEV_DEMAND_FINAL	6698	1.15	1.41	0	0	1	1	26
AVG_STOCK_LEVEL	6698	166.07	311.99	2	33	74	176	7196
MEDIAN_STOCK_LEVEL	6698	157.47	297.64	3	32	69	166	6918
LEAD_TIME_DAYS	6698	13.66	26.04	0	2	6	15	304

Figura 20.

Boxplots Retailer 7



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 14.*Detección de outliers (z-score) Retailer 7*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	114	1.70 %
LEAD_TIME_DAYS	138	2.06 %
MEDIAN_STOCK_LEVEL	121	1.81 %
ORDERS	132	1.97 %
ORDERS_STOCKOUT	132	1.97 %
SALES	135	2.02 %
STDDEV_DEMAND_FINAL	108	1.61 %
UNITS	79	1.18 %
UNITS_DEMAND	79	1.18 %
UNITS_STOCKOUT	125	1.87 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

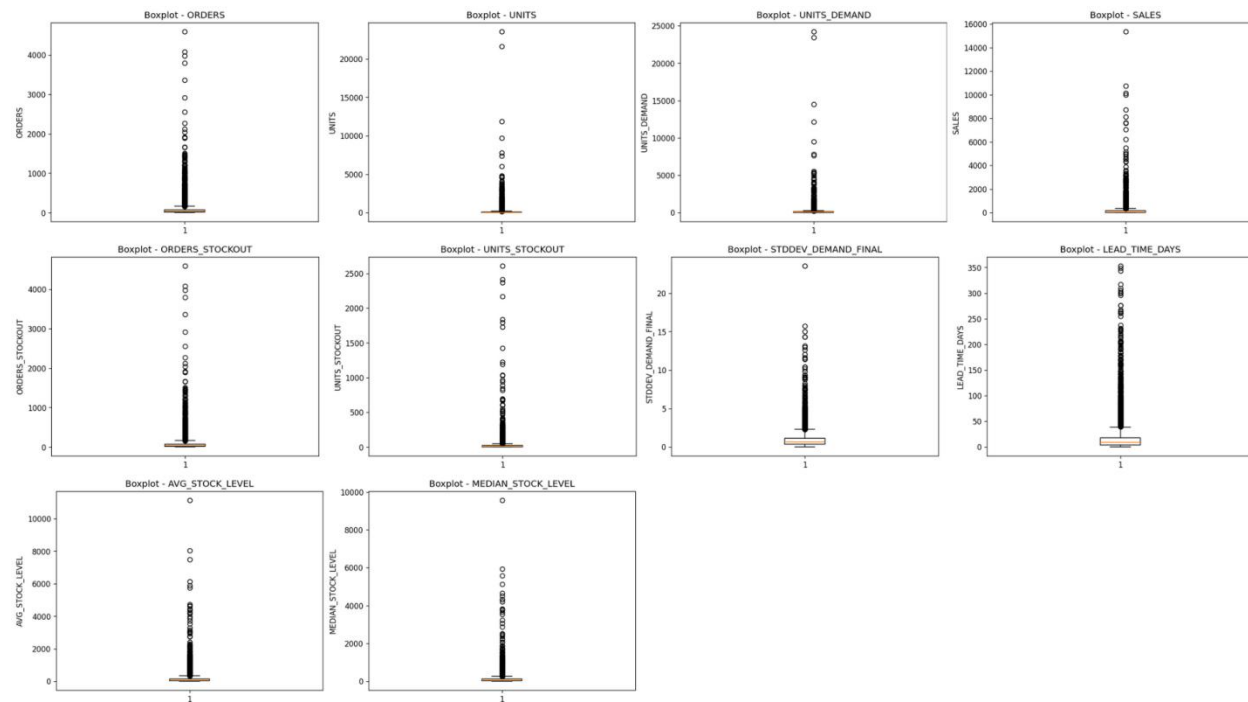
Retailer 8

Tabla 15.

Estadísticas Descriptivas Retailer 8

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	5870	130.85	582.89	1	13	33	86	23551
ORDERS	5870	84.42	203.53	2	15	33	73	4590
SALES	5870	191.28	545.03	1	23	60	154	15363
UNITS_STOCKOUT	5870	26.76	98.23	1	3	9	21	2609
ORDERS_STOCKOUT	5870	84.42	203.53	2	15	33	73	4590
UNITS_DEMAND	5870	157.61	652.87	2	19	44	109	24236
STDDEV_DEMAND_FINAL	5870	0.97	1.20	0	0	1	1	24
AVG_STOCK_LEVEL	5870	173.27	405.57	3	38	71	159	11132
MEDIAN_STOCK_LEVEL	5870	134.72	321.79	1	31	57	124	9563
LEAD_TIME_DAYS	5870	18.06	31.01	0	4	9	18	353

Figura 23.

Boxplots Retailer 8

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 16.*Detección de outliers (z-score) Retailer 8*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	79	1.35 %
LEAD_TIME_DAYS	139	2.37 %
MEDIAN_STOCK_LEVEL	78	1.33 %
ORDERS	98	1.67 %
ORDERS_STOCKOUT	98	1.67 %
SALES	71	1.21 %
STDDEV_DEMAND_FINAL	115	1.96 %
UNITS	51	0.87 %
UNITS_DEMAND	54	0.92 %
UNITS_STOCKOUT	49	0.83 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

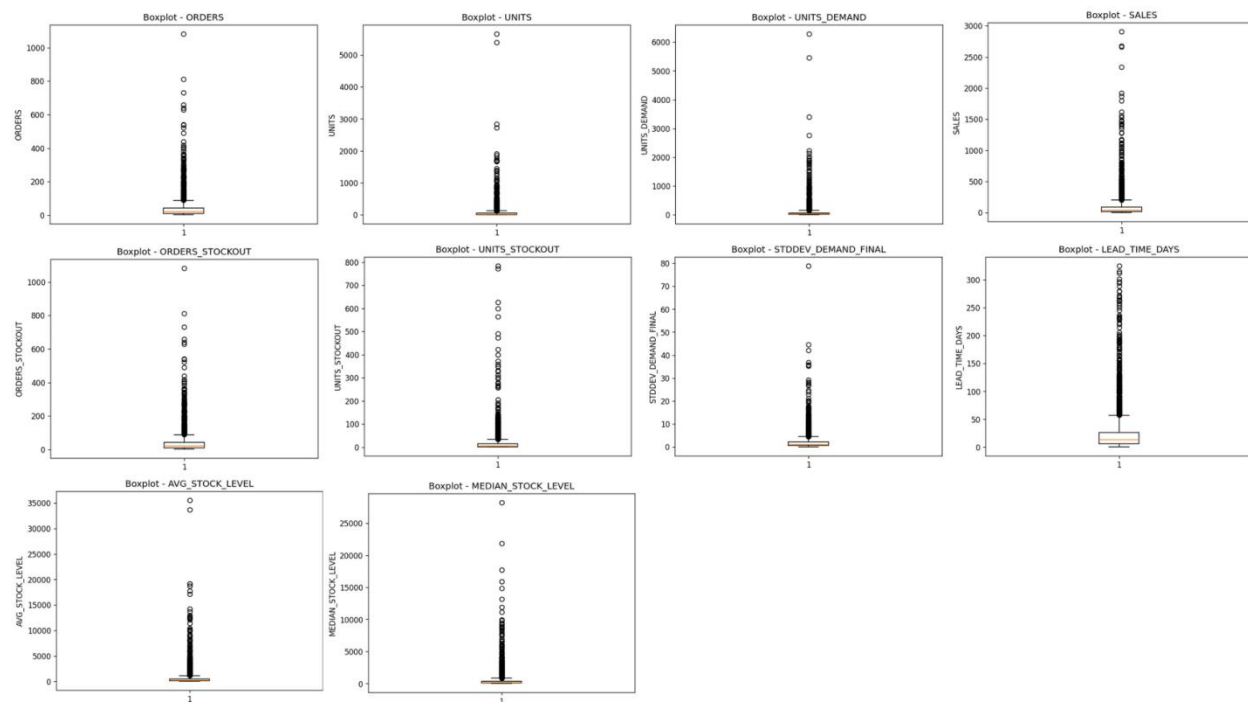
Retailer 9

Tabla 17.

Estadísticas Descriptivas Retailer 9

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	3050	71.71	224.21	1	9	22	55	5659
ORDERS	3050	38.94	62.44	2	10	21	42	1082
SALES	3050	93.22	194.78	1	14	35	91	2908
UNITS_STOCKOUT	3050	17.01	43.49	1	2	6	15	784
ORDERS_STOCKOUT	3050	38.94	62.44	2	10	21	42	1082
UNITS_DEMAND	3050	88.72	251.81	2	14	30	72	6285
STDDEV_DEMAND_FINAL	3050	2.11	3.54	0	1	1	2	79
AVG_STOCK_LEVEL	3050	620.59	1686.10	13	85	195	504	35554
MEDIAN_STOCK_LEVEL	3050	473.06	1286.69	11	66	146	375	28225
LEAD_TIME_DAYS	3050	25.32	38.12	0	6	13	27	325

Figura 26.

Boxplots Retailer 9

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 27.

Histogramas Retailer 9

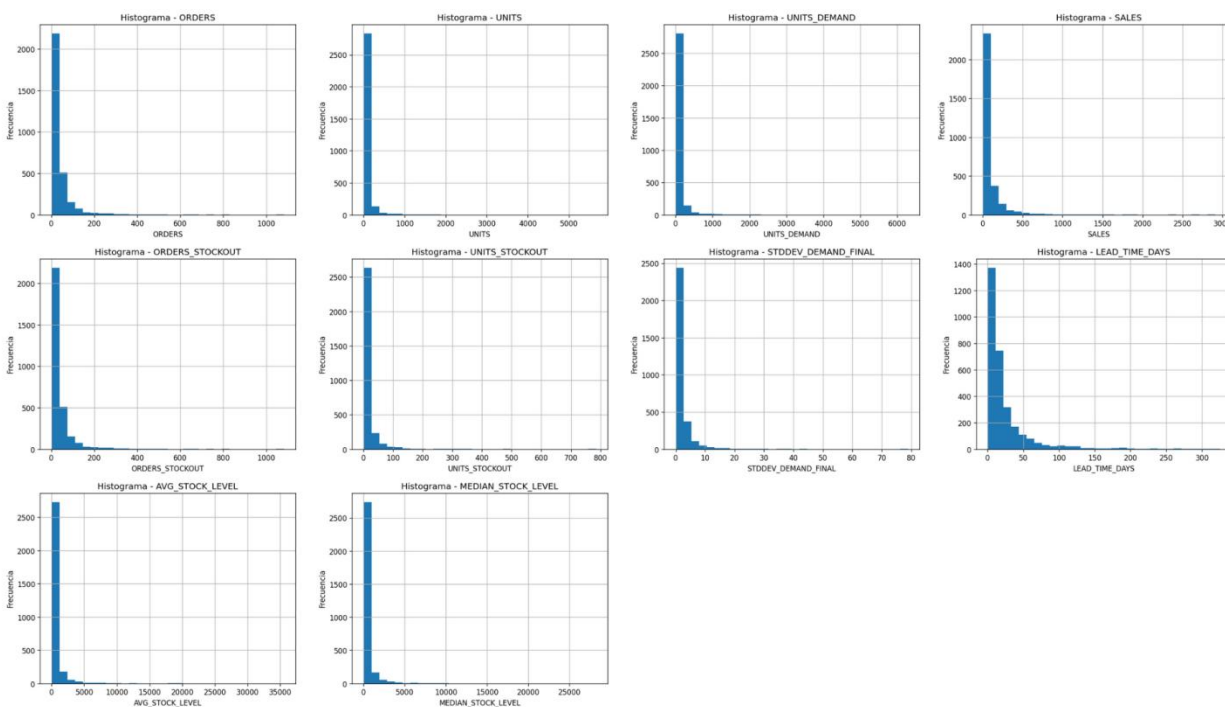
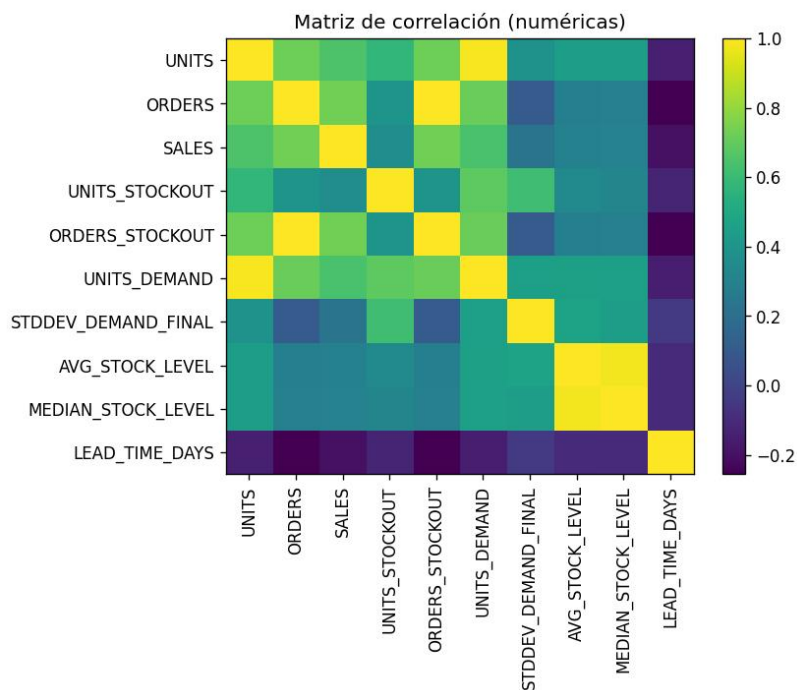


Figura 28.

Mapa de calor: Matriz de Correlación Retailer 9



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 18.*Detección de outliers (z-score) Retailer 9*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	50	1.64 %
LEAD_TIME_DAYS	77	2.52 %
MEDIAN_STOCK_LEVEL	46	1.51 %
ORDERS	65	2.13 %
ORDERS_STOCKOUT	65	2.13 %
SALES	59	1.93 %
STDDEV_DEMAND_FINAL	55	1.80 %
UNITS	43	1.41 %
UNITS_DEMAND	48	1.57 %
UNITS_STOCKOUT	33	1.08 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Retailer 10

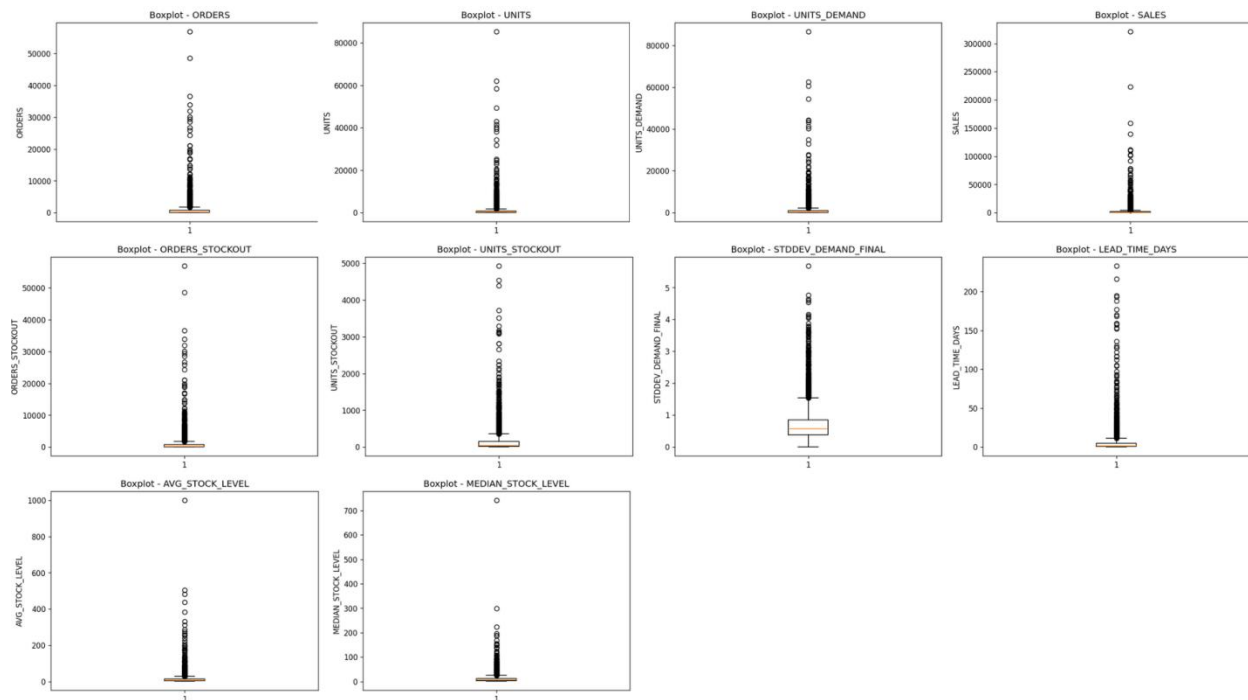
Tabla 19.

Estadísticas Descriptivas Retailer 10

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	3880	1004.66	3313.27	1	45	197	752	85316
ORDERS	3880	841.87	2483.12	2	52	211	717	56959
SALES	3880	2531.70	9815.64	1	92	412	1646	321160
UNITS_STOCKOUT	3880	149.66	319.54	1	12	44	148	4926
ORDERS_STOCKOUT	3880	841.87	2483.12	2	52	211	717	56959
UNITS_DEMAND	3880	1154.31	3535.09	2	61	251	918	86673
STDDEV_DEMAND_FINAL	3880	0.72	0.56	0	0	1	1	6
AVG_STOCK_LEVEL	3880	13.51	30.18	1	4	8	14	1000
MEDIAN_STOCK_LEVEL	3880	10.37	19.34	1	3	6	12	742
LEAD_TIME_DAYS	3880	5.86	15.21	0	0	2	5	233

Figura 29.

Boxplots Retailer 10



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 20.*Detección de outliers (z-score) Retailer 10*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	47	1.21 %
LEAD_TIME_DAYS	59	1.52 %
MEDIAN_STOCK_LEVEL	52	1.34 %
ORDERS	53	1.37 %
ORDERS_STOCKOUT	53	1.37 %
SALES	39	1.01 %
STDDEV_DEMAND_FINAL	92	2.37 %
UNITS	45	1.16 %
UNITS_DEMAND	47	1.21 %
UNITS_STOCKOUT	71	1.83 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Retailer 11

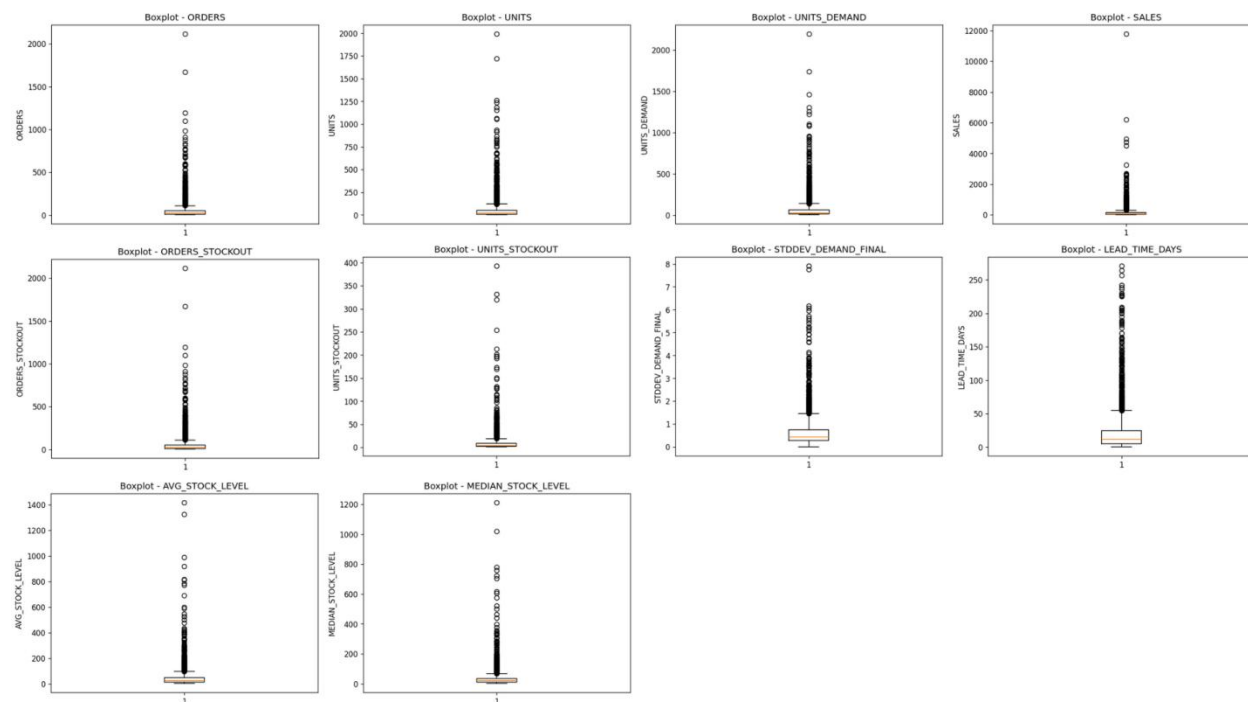
Tabla 21.

Estadísticas Descriptivas Retailer 11

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	3161	54.23	114.38	1	8	20	52	1993
ORDERS	3161	51.12	101.66	2	10	22	50	2117
SALES	3161	151.15	382.70	1	21	54	138	11794
UNITS_STOCKOUT	3161	9.46	19.51	1	2	4	9	393
ORDERS_STOCKOUT	3161	51.12	101.66	2	10	22	50	2117
UNITS_DEMAND	3161	63.69	125.69	2	12	26	62	2194
STDDEV_DEMAND_FINAL	3161	0.62	0.71	0	0	0	1	8
AVG_STOCK_LEVEL	3161	43.61	75.35	1	13	24	46	1416
MEDIAN_STOCK_LEVEL	3161	33.14	58.66	1	11	19	34	1212
LEAD_TIME_DAYS	3161	22.06	32.02	0	5	12	25	271

Figura 32.

Boxplots Retailer 11



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 22.*Detección de outliers (z-score) Retailer 11*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	47	1.49 %
LEAD_TIME_DAYS	84	2.66 %
MEDIAN_STOCK_LEVEL	42	1.33 %
ORDERS	59	1.87 %
ORDERS_STOCKOUT	59	1.87 %
SALES	38	1.20 %
STDDEV_DEMAND_FINAL	57	1.80 %
UNITS	63	1.99 %
UNITS_DEMAND	64	2.02 %
UNITS_STOCKOUT	38	1.20 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Retailer 12

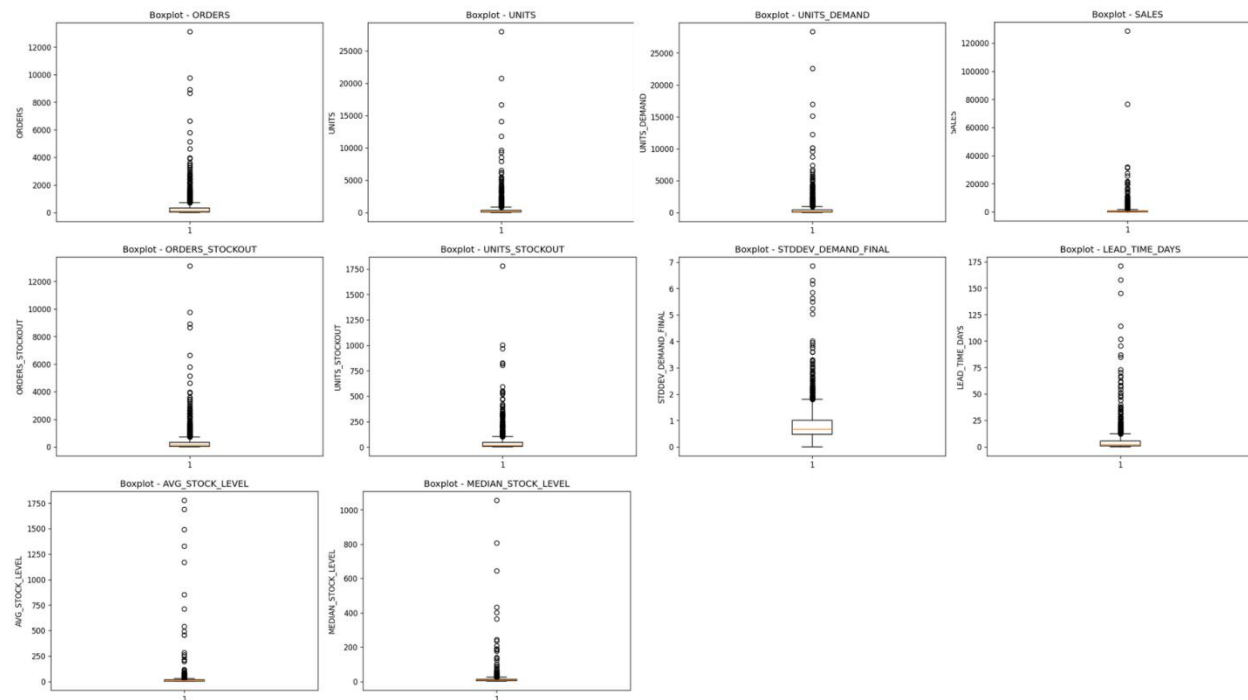
Tabla 23.

Estadísticas Descriptivas Retailer 12

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	1467	480.93	1400.61	1	53	148	375	28010
ORDERS	1467	354.77	774.38	2	55	138	328	13119
SALES	1467	1166.19	4713.15	1	85	250	763	128640
UNITS_STOCKOUT	1467	45.27	96.84	1	6	17	45	1780
ORDERS_STOCKOUT	1467	354.77	774.38	2	55	138	328	13119
UNITS_DEMAND	1467	526.19	1475.58	2	65	169	421	28340
STDDEV_DEMAND_FINAL	1467	0.86	0.66	0	0	1	1	7
AVG_STOCK_LEVEL	1467	20.16	96.80	2	5	8	14	1778
MEDIAN_STOCK_LEVEL	1467	13.39	45.59	1	4	7	12	1056
LEAD_TIME_DAYS	1467	5.53	11.94	0	1	2	6	171

Figura 35.

Boxplots Retailer 12



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 24.*Detección de outliers (z-score) Retailer 12*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	11	0.75 %
LEAD_TIME_DAYS	24	1.64 %
MEDIAN_STOCK_LEVEL	13	0.89 %
ORDERS	23	1.57 %
ORDERS_STOCKOUT	23	1.57 %
SALES	16	1.09 %
STDDEV_DEMAND_FINAL	24	1.64 %
UNITS	20	1.36 %
UNITS_DEMAND	19	1.30 %
UNITS_STOCKOUT	23	1.57 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Retailer 13

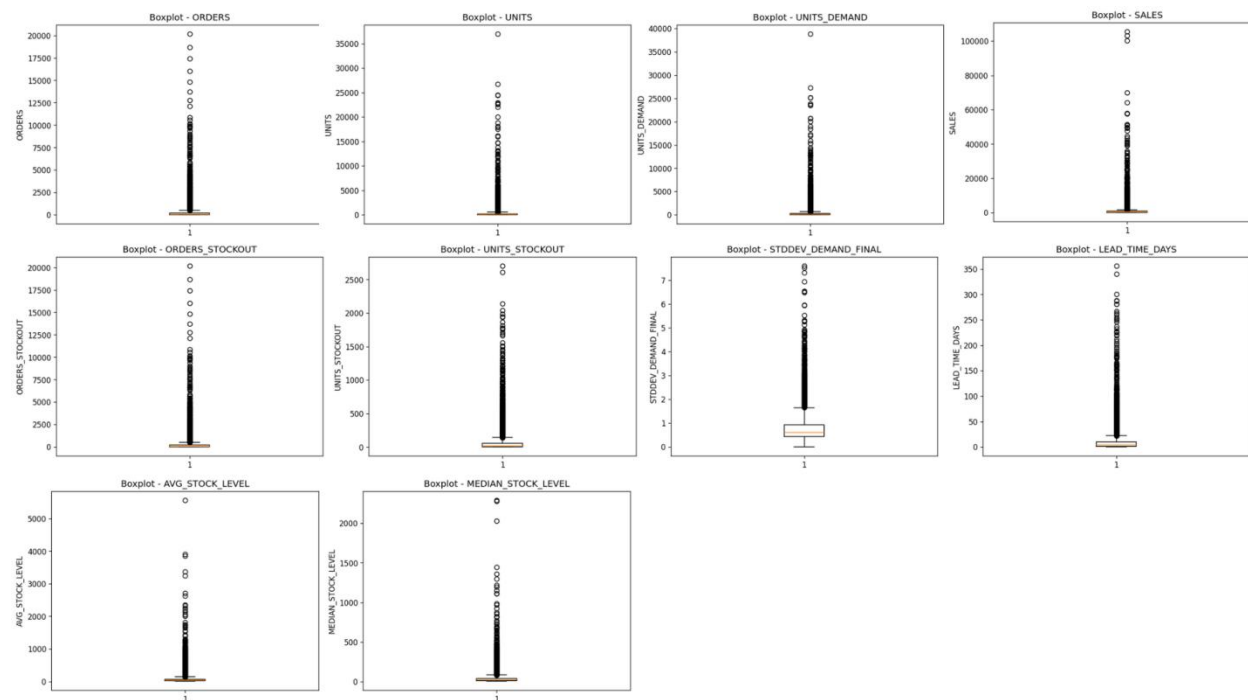
Tabla 25.

Estadísticas Descriptivas Retailer 13

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	11693	354.20	1203.25	1	21	69	240	36992
ORDERS	11693	284.25	811.14	2	24	72	227	20181
SALES	11693	1024.89	3522.75	1	78	230	720	105538
UNITS_STOCKOUT	11693	62.56	144.02	1	5	17	59	2704
ORDERS_STOCKOUT	11693	284.25	811.14	2	24	72	227	20181
UNITS_DEMAND	11693	416.76	1311.87	2	29	88	304	38860
STDDEV_DEMAND_FINAL	11693	0.76	0.59	0	0	1	1	8
AVG_STOCK_LEVEL	11693	75.69	166.73	5	18	33	68	5565
MEDIAN_STOCK_LEVEL	11693	43.19	79.77	5	13	22	42	2290
LEAD_TIME_DAYS	11693	9.91	21.14	0	1	3	10	356

Figura 38.

Boxplots Retailer 13



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 39.

Histogramas Retailer 13

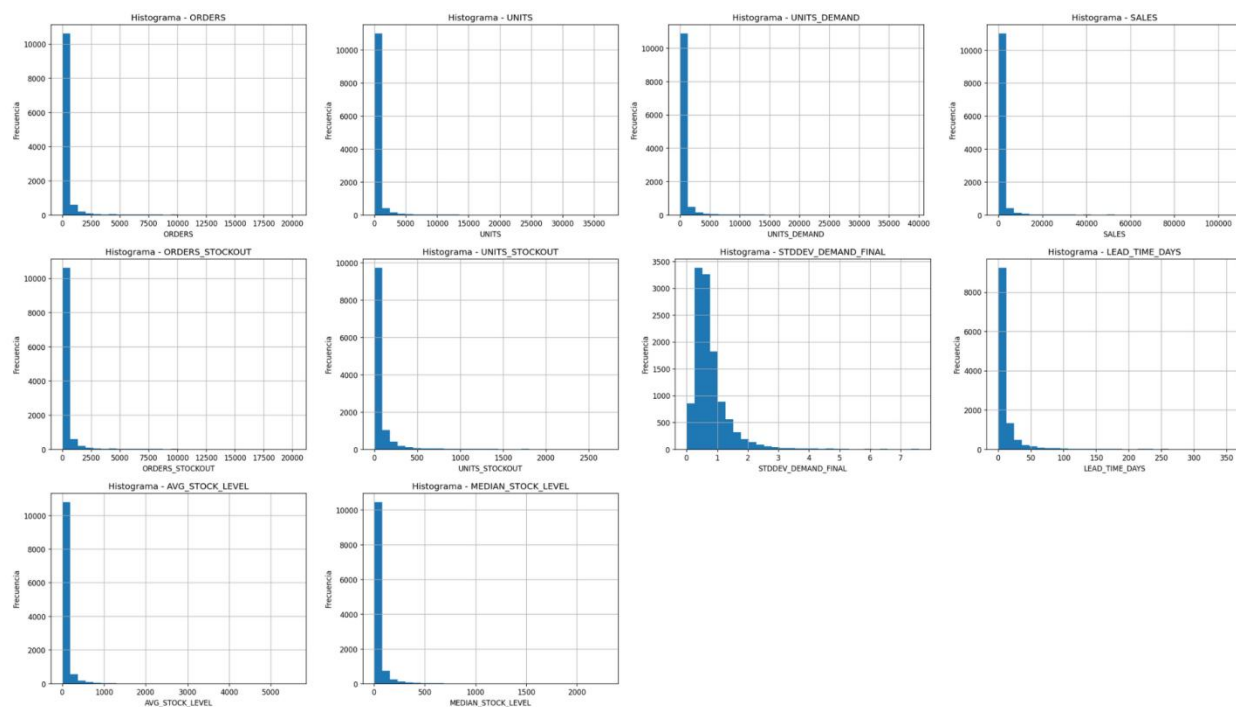
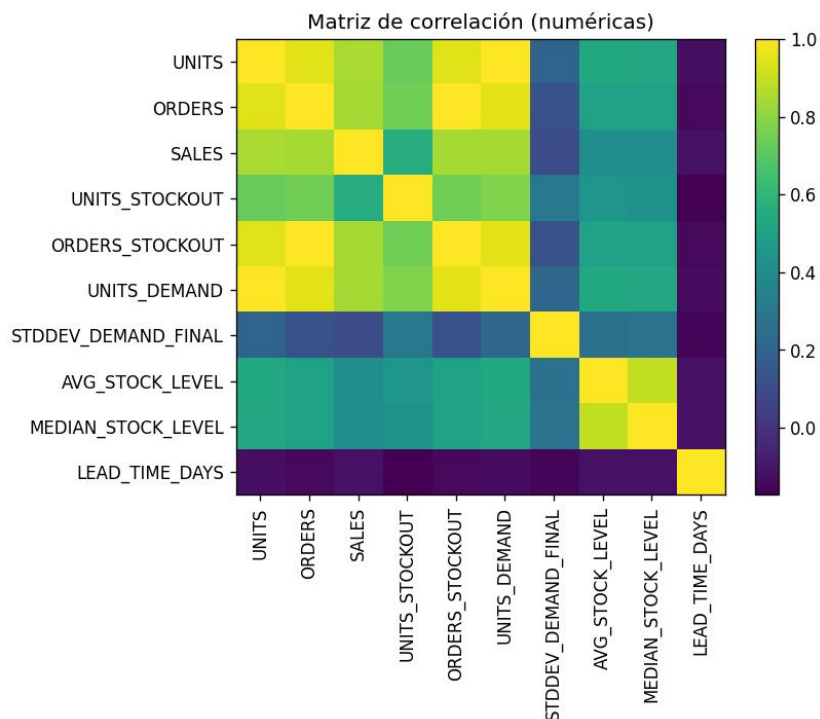


Figura 40.

Mapa de calor: Matriz de Correlación Retailer 13



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 26.*Detección de outliers (z-score) Retailer 13*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	185	1.58 %
LEAD_TIME_DAYS	235	2.01 %
MEDIAN_STOCK_LEVEL	201	1.72 %
ORDERS	175	1.50 %
ORDERS_STOCKOUT	175	1.50 %
SALES	139	1.19 %
STDDEV_DEMAND_FINAL	212	1.81 %
UNITS	152	1.30 %
UNITS_DEMAND	157	1.34 %
UNITS_STOCKOUT	219	1.87 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Retailer 14

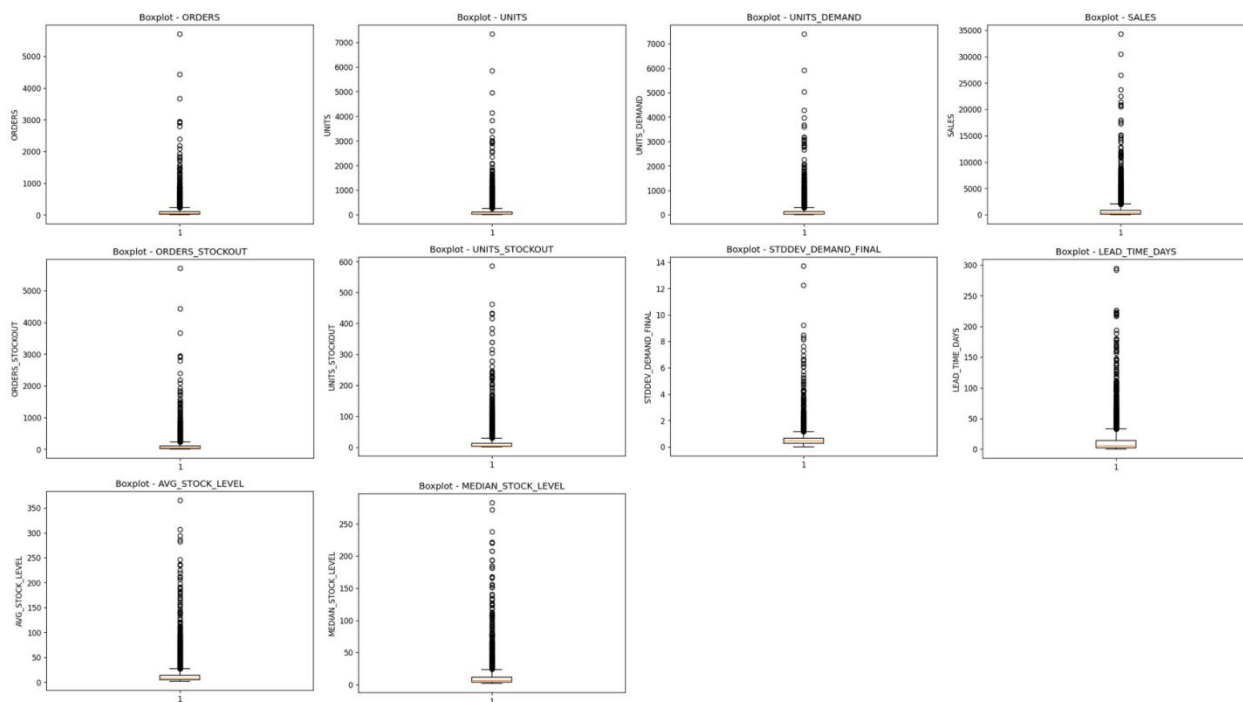
Tabla 27.

Estadísticas Descriptivas Retailer 14

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	5557	121.72	284.87	1	17	44	117	7339
ORDERS	5557	104.93	215.33	2	19	44	106	5709
SALES	5557	852.71	1698.42	2	131	352	894	34317
UNITS_STOCKOUT	5557	13.69	30.29	1	2	4	13	586
ORDERS_STOCKOUT	5557	104.93	215.33	2	19	44	106	5709
UNITS_DEMAND	5557	135.40	303.31	2	21	50	132	7407
STDDEV_DEMAND_FINAL	5557	0.56	0.65	0	0	0	1	14
AVG_STOCK_LEVEL	5557	13.63	21.28	2	5	8	14	365
MEDIAN_STOCK_LEVEL	5557	11.42	17.56	2	4	7	12	283
LEAD_TIME_DAYS	5557	11.94	21.17	0	2	5	14	295

Figura 41.

Boxplots Retailer 14



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Figura 42.

Histogramas Retailer 14

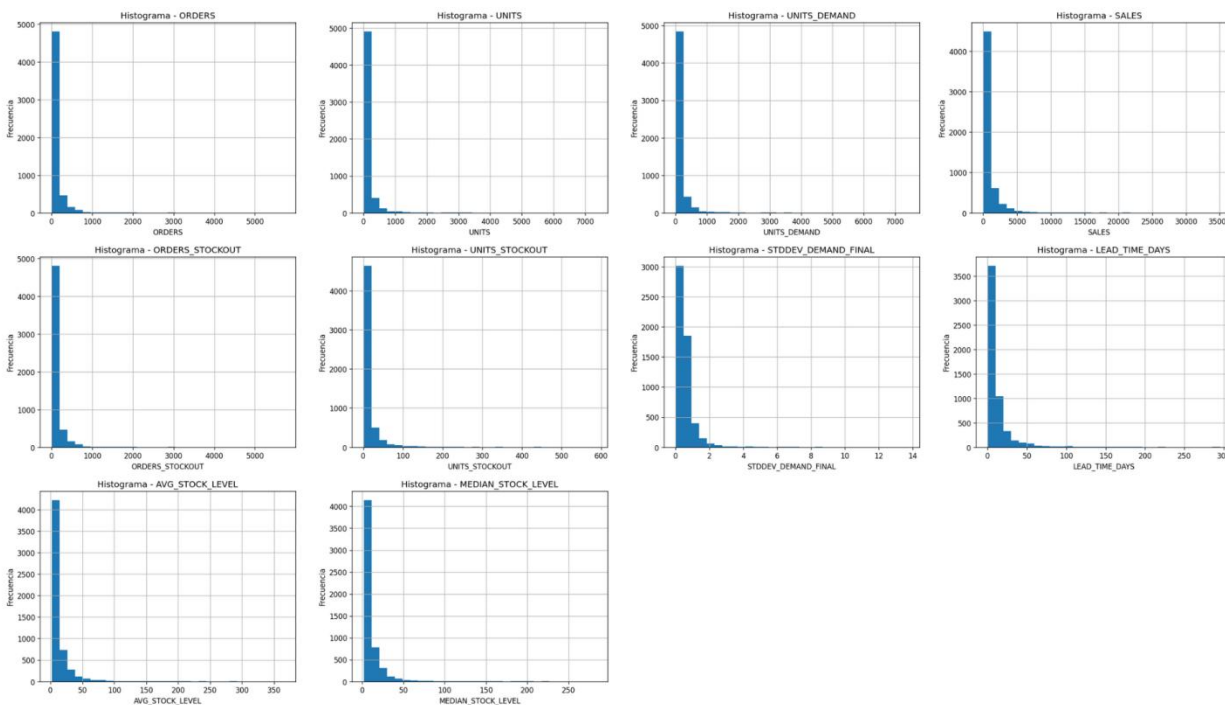
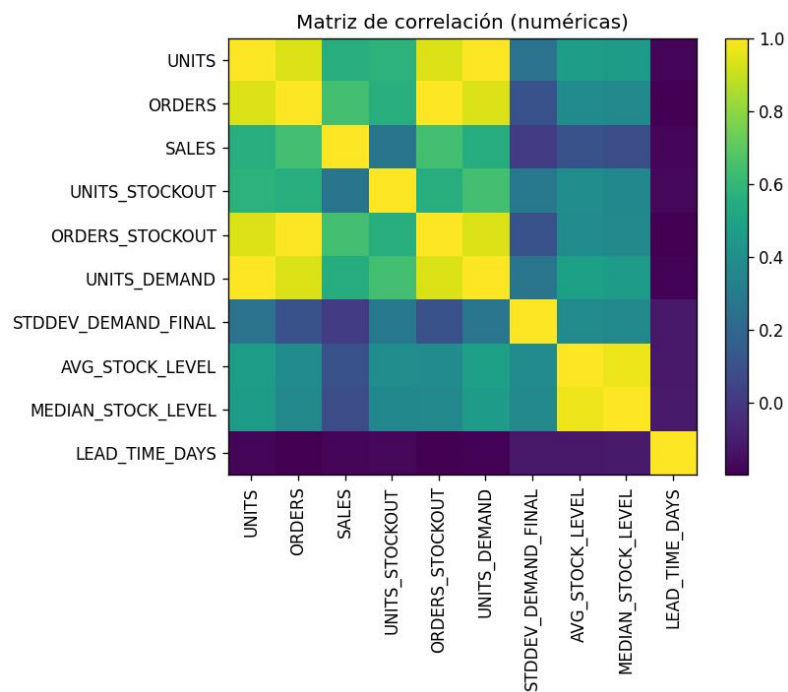


Figura 43.

Mapa de calor: Matriz de Correlación Retailer 14



MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 28.*Detección de outliers (z-score) Retailer 14*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	100	1.80 %
LEAD_TIME_DAYS	115	2.07 %
MEDIAN_STOCK_LEVEL	97	1.75 %
ORDERS	76	1.37 %
ORDERS_STOCKOUT	76	1.37 %
SALES	84	1.51 %
STDDEV_DEMAND_FINAL	80	1.44 %
UNITS	91	1.64 %
UNITS_DEMAND	91	1.64 %
UNITS_STOCKOUT	101	1.82 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

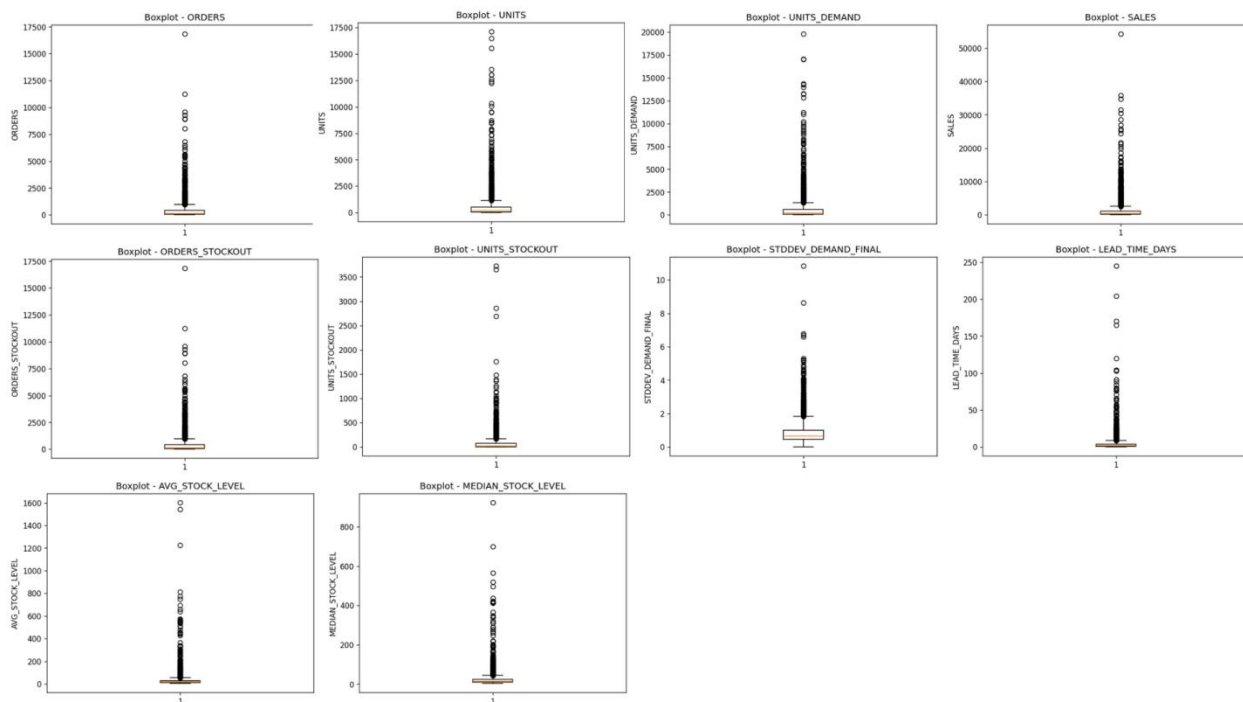
Retailer 15

Tabla 29.

Estadísticas Descriptivas Retailer 15

Variable	Conteo	Promedio	Desviación estándar	Mínimo	25%	50%	75%	Máximo
UNITS	3512	511.00	1145.59	1	56	164	501	17117
ORDERS	3512	403.68	811.38	2	56	152	420	16848
SALES	3512	1201.85	2653.15	1	160	414	1127	54287
UNITS_STOCKOUT	3512	72.92	170.01	1	7	23	73	3726
ORDERS_STOCKOUT	3512	403.68	811.38	2	56	152	420	16848
UNITS_DEMAND	3512	583.91	1274.65	2	68	194	581	19807
STDDEV_DEMAND_FINAL	3512	0.86	0.70	0	0	1	1	11
AVG_STOCK_LEVEL	3512	32.27	70.29	3	12	18	29	1603
MEDIAN_STOCK_LEVEL	3512	23.72	41.10	3	10	15	24	923
LEAD_TIME_DAYS	3512	4.42	10.64	0	1	2	4	245

Figura 44.

Boxplots Retailer 15

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 30.*Detección de outliers (z-score) Retailer 15*

Variable	Conteo Outliers	Porcentaje Outliers
AVG_STOCK_LEVEL	51	1.45 %
LEAD_TIME_DAYS	49	1.40 %
MEDIAN_STOCK_LEVEL	44	1.25 %
ORDERS	70	1.99 %
ORDERS_STOCKOUT	70	1.99 %
SALES	61	1.74 %
STDDEV_DEMAND_FINAL	69	1.96 %
UNITS	61	1.74 %
UNITS_DEMAND	62	1.77 %
UNITS_STOCKOUT	47	1.34 %

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Apéndice D. Evaluación consolidada entre modelos

Tabla 1.

Evaluación consolidada entre modelos

ID	Modelo Ensamblado				ABC	K-means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)				
	Stacking based on Costs		Clustering performance based					Costo	Costo	Coeficiente de Silueta	Costo	Coeficiente de Silueta
	Costo	Coeficiente de Silueta	Costo	Coeficiente de Silueta								
1	\$83,174.26	69.64%	\$145,065.80	86.80%	\$243,551.66	\$267,964.43	72.50%	\$145,065.80	86.80%	\$103,995.57	17.46%	
2	\$20,691.45	76.95%	\$20,767.27	89.28%	\$30,215.96	\$20,776.29	70.00%	\$20,767.27	89.28%	\$26,922.58	15.26%	
3	\$11.63	11.59%	\$20.61	65.37%	\$20.40	\$20.61	65.37%	\$19.85	53.26%	\$14.89	28.24%	
4	\$339.62	49.51%	\$489.43	89.24%	\$912.08	\$556.98	85.65%	\$489.43	89.24%	\$395.30	42.41%	
5	\$77.16	39.35%	\$116.66	84.16%	\$199.70	\$128.53	81.99%	\$116.66	84.16%	\$92.71	45.98%	
6	\$158.30	70.33%	\$240.31	84.27%	\$453.65	\$158.64	68.25%	\$240.31	84.27%	\$184.81	37.01%	
7	\$9,287.64	-1.34%	\$16,385.88	73.42%	\$26,876.74	\$16,385.88	73.42%	\$18,025.08	58.92%	\$11,644.02	12.91%	
8	\$11,835.41	8.07%	\$25,101.61	73.83%	\$34,803.58	\$25,101.61	73.83%	\$24,355.01	59.57%	\$13,092.58	13.50%	
9	\$1,042.59	-5.24%	\$1,943.97	66.98%	\$2,830.27	\$1,943.97	66.98%	\$2,145.81	60.61%	\$1,373.45	10.90%	
10	\$463,253.33	73.69%	\$1,481,432.00	74.27%	\$1,739,147.00	\$1,519,116.00	73.35%	\$1,481,432.00	74.27%	\$506,643.20	16.25%	
11	\$946.71	44.38%	\$1,675.89	55.35%	\$2,555.19	\$3,285.42	52.80%	\$1,675.89	55.35%	\$1,315.58	17.90%	
12	\$23,680.50	87.17%	\$38,704.61	82.60%	\$59,077.57	\$83,387.40	73.19%	\$38,704.61	82.60%	\$27,165.41	23.31%	
13	\$252,444.23	67.47%	\$809,051.88	73.02%	\$882,187.03	\$809,051.88	73.02%	\$473,797.53	61.91%	\$276,824.56	14.12%	
14	\$9,153.94	16.11%	\$15,674.95	74.38%	\$25,402.83	\$14,593.82	68.03%	\$15,674.95	74.38%	\$12,064.71	13.04%	
15	\$72,653.53	74.00%	\$127,499.08	81.66%	\$205,926.67	\$226,138.32	58.71%	\$127,499.08	81.66%	\$91,066.83	21.08%	
Total	\$948,750.30	-	\$2,684,169.95	-	\$3,254,160.32	\$2,988,609.76	-	\$2,350,009.27	-	\$1,072,796.20	-	

MODELO ENSAMBLADO PARA CLASIFICACIÓN DE INVENTARIO

Tabla 2.

Evaluación consolidada entre modelos aplicando PCA

ID	Modelo Ensamblado				ABC	K-means	Clustering Jerárquico	Modelo de Mezclas Gaussianas (GMM)			
	Stacking based on Costs		Clustering performance based					Costo	Coeficiente de Silueta	Costo	Coeficiente de Silueta
	Costo	Coeficiente de Silueta	Costo	Coeficiente de Silueta	Costo	Costo	Coeficiente de Silueta	Costo	Coeficiente de Silueta	Costo	Coeficiente de Silueta
1	\$83,174.26	17.06%	\$267,589.92	81.34%	\$243,551.66	\$267,589.92	81.34%	\$233,777.79	76.39%	\$187,853.58	63.95%
2	\$27,780.71	-12.80%	\$35,863.24	80.43%	\$30,215.96	\$35,863.24	80.43%	\$41,352.67	73.25%	\$40,414.33	64.74%
3	\$11.63	-8.72%	\$33.95	60.21%	\$20.40	\$30.18	57.36%	\$31.80	51.59%	\$33.95	60.21%
4	\$339.62	18.59%	\$561.46	90.24%	\$912.08	\$561.46	90.24%	\$590.15	87.81%	\$383.62	57.36%
5	\$77.16	18.59%	\$116.27	88.40%	\$199.70	\$133.22	85.18%	\$116.27	88.40%	\$89.10	49.69%
6	\$233.48	71.70%	\$302.33	84.88%	\$453.65	\$310.53	83.44%	\$302.33	84.88%	\$475.03	59.56%
7	\$9,270.74	11.52%	\$23,304.69	77.76%	\$26,876.74	\$23,304.69	77.76%	\$17,801.83	73.35%	\$19,351.05	61.20%
8	\$16,219.37	-16.98%	\$26,680.26	91.33%	\$34,803.58	\$28,390.02	89.09%	\$26,680.26	91.33%	\$17,410.55	72.73%
9	\$1,564.85	67.50%	\$2,210.57	83.76%	\$2,830.27	\$2,210.57	83.76%	\$3,502.84	71.15%	\$2,450.74	70.87%
10	\$625,879.26	82.74%	\$1,430,284.00	85.25%	\$1,739,147.00	\$1,537,917.00	82.42%	\$1,430,284.00	85.25%	\$977,171.60	65.10%
11	\$946.67	8.21%	\$3,223.07	70.88%	\$2,555.19	\$3,223.07	70.88%	\$1,988.58	54.22%	\$3,335.96	56.73%
12	\$72,011.81	-18.42%	\$84,828.11	79.94%	\$59,077.57	\$84,504.20	79.41%	\$84,828.11	79.94%	\$86,238.89	64.03%
13	\$327,436.43	45.72%	\$816,869.82	82.98%	\$882,187.03	\$816,869.82	82.98%	\$502,097.87	65.08%	\$558,766.54	62.54%
14	\$12,017.83	-21.68%	\$17,034.46	82.64%	\$25,402.83	\$23,869.13	73.48%	\$17,034.46	82.64%	\$24,863.84	60.91%
15	\$85,736.32	70.60%	\$257,107.03	73.78%	\$205,926.67	\$257,107.03	73.78%	\$271,473.59	72.44%	\$158,116.16	61.72%
Total	\$1,262,700.13	-	\$2,966,009.17	-	\$3,254,160.32	\$3,081,884.08	-	\$2,631,862.53	-	\$2,076,954.95	-