3.2 Machine Learning Approaches for Demand Forecasting and Inventory Optimization in SMEs (3).pdf

Universidad Peruana Union

Detalles del documento

Identificador de la entrega
trn:oid:::29566:595025233

Fecha de entrega
27 may 2026, 11:29 a.m. GMT-5

Fecha de descarga
27 may 2026, 11:33 a.m. GMT-5

Nombre del archivo
3.2 Machine Learning Approaches for Demand Forecasting and Inventory Optimization in SMEs (3).pdf

Tamaño del archivo
766.0 KB

31 páginas

8455 palabras

47.286 caracteres

2% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 8 palabras)

Exclusiones

- N.º de coincidencias excluidas

Fuentes principales

1% Fuentes de Internet
1% Publicaciones
2% Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alerta de integridad para revisión

Texto oculto
10 caracteres sospechosos en N.º de página
El texto es alterado para mezclarse con el fondo blanco del documento.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Fuentes principales

1% Fuentes de Internet
1% Publicaciones
2% Trabajos entregados (trabajos del estudiante)

Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1 Internet
docs.soda.io <1%

2 Internet
nasaul.github.io <1%

3 Trabajos entregados
Central Queensland University on 2023-08-17 <1%

4 Internet
repositorio.unsaac.edu.pe <1%

5 Internet
conference.nber.org <1%

6 Trabajos entregados
Universitat Politècnica de València on 2017-04-21 <1%

7 Internet
marketingcareerinsights.com <1%

8 Publicación
Qian Zhao, Peilin Yang, Yushen Liang, Zhenzhen Xu, Jianle Chen, Shiguo Chen, Xin... <1%

9 Publicación
Fernandez Chambi, Mayenka. "Análisis de opinión del microblogging Twitter por l... <1%

10 Trabajos entregados
UNIBA on 2024-11-11 <1%

11 Trabajos entregados
UNIBA on 2025-05-12 <1%

12 Trabajos entregados

Universidad Francisco de Vitoria on 2025-01-28 **<1%**

Machine Learning Approaches for Demand Forecasting and Inventory Optimization in SMEs

Enfoques de aprendizaje automático para la previsión de la demanda y la optimización de inventarios en PYMES

Samuel Axel Vasquez Castillo[1]*, ORCID iD: https://orcid.org/0009-0006-7953-5463, correo institucional: samuel.vasquez@upeu.edu.pe

Mesias Caleb Orrego Mego[1], ORCID iD: https://orcid.org/0009-0005-7842-3390, correo institucional: mesiasorrego@upeu.edu.pe

[1] Universidad Peruana Unión, Filial Tarapoto, Perú

RESUMEN

Este estudio abordó la ineficiencia en la gestión de inventarios de PYMES comerciales mediante modelos de pronóstico de demanda basados en datos reales de punto de venta. El objetivo fue mejorar la precisión predictiva y apoyar decisiones de reposición. Se empleó un dataset de 4,451,113 registros (marzo 2020–diciembre 2024) y un flujo KDD con depuración, tratamiento de atípicos y agregaciones diaria, semanal y mensual. Se evaluaron Prophet, XGBoost y Árbol de Decisión con una partición temporal 70/20/10 y métricas MAE, RMSE y sMAPE, además de un híbrido Prophet+XGBoost. El sMAPE disminuyó de 27.2 % (diario) a 15.7 % y 12.5 % (semanal y mensual), reflejando reducciones del 40–50 %. Los modelos mostraron desempeños estadísticamente equivalentes ($p > 0.05$); no obstante, el Árbol de Decisión resaltó por su parsimonia y Prophet por su estabilidad en series agregadas. Se concluyó que modelos interpretables y de bajo costo computacional permiten pronósticos suficientemente robustos para PYMES, con potencial de reducir quiebres y excesos de inventario entre 10–15 %. Se recomienda integrar covariables exógenas y explorar la optimización conjunta pronóstico–reabastecimiento.

Palabras clave: aprendizaje automático, previsión de la demanda, previsión de series temporales, optimización de inventarios, gestión de la cadena de suministro.

ABSTRACT

This study addressed inefficiencies in inventory management within commercial SMEs by applying demand-forecasting models based on real point-of-sale data. The objective was to improve predictive accuracy and support replenishment decisions. A dataset of 4,451,113 transactions (March 2020–December 2024) was used, following a KDD workflow that included key cleansing, outlier treatment, and daily, weekly, and monthly aggregations. Prophet, XGBoost, and Decision Tree models were evaluated using a 70/20/10 temporal split and MAE, RMSE, and sMAPE as performance metrics, along with a hybrid Prophet+XGBoost approach. The sMAPE decreased from 27.2% (daily) to 15.7% and 12.5% (weekly and monthly), representing reductions of 40–50%. Model performances were statistically equivalent ($p > 0.05$); however, the Decision Tree stood out for its parsimony, and Prophet for its stability in aggregated series. The study concludes that interpretable, low-computational-cost models provide sufficiently robust forecasts for SMEs, with the potential to reduce stockouts and excess inventory by 10–15%. Incorporating exogenous covariates and exploring joint forecast–replenishment optimization is recommended for future work.

Keywords: machine learning, demand forecasting, time series forecasting, inventory optimization, supply chain management.

1. INTRODUCTION

La gestión de inventarios representa un componente esencial en la sostenibilidad de las cadenas de suministro a nivel global. Las ineficiencias asociadas a sobrestock, quiebres y obsolescencia pueden reducir los ingresos de las organizaciones hasta en un 25 %, afectando directamente la

competitividad y la resiliencia operativa (Khastgir & Kumar, 2024). En este contexto, la digitalización de los procesos logísticos y el aprovechamiento de grandes volúmenes de datos transaccionales han impulsado nuevas formas de pronosticar la demanda y optimizar las decisiones de reabastecimiento mediante técnicas de aprendizaje automático (Machine Learning, ML) (Khedr & S, 2024), (Andrade & Cunha, 2023). Modelos basados en árboles de decisión, aprendizaje profundo y combinaciones híbridas se han posicionado como alternativas robustas para entornos de alta volatilidad y complejidad logística (Lei et al., 2025), (Wellens et al., 2024).

Una gestión eficiente de inventarios permite equilibrar costos de mantenimiento y niveles de servicio, asegurando la continuidad del flujo de bienes y minimizando pérdidas (Thuy & Nguyen, 2023). Diversas investigaciones han demostrado que la precisión del pronóstico de demanda se correlaciona directamente con la rentabilidad y el capital de trabajo, especialmente en organizaciones medianas donde la rotación de productos es irregular (Seyedan et al., 2023), (Sattar et al., 2025). A nivel internacional, se han incorporado métricas de sostenibilidad y evaluación económica (*cost–accuracy–ESG*) en modelos de predicción, evidenciando la relevancia de integrar la eficiencia operativa con criterios de responsabilidad social (Airlangga, 2024). De este modo, el enfoque actual del pronóstico de demanda no solo busca precisión estadística, sino también sostenibilidad y resiliencia dentro de los sistemas de abastecimiento.

Los métodos de ML han demostrado ventajas frente a los modelos tradicionales (ARIMA, EOQ o punto de reorden) al capturar relaciones no lineales y patrones estacionales múltiples. Modelos como XGBoost y Random Forest se destacan por su capacidad para manejar datos ruidosos y heterogéneos, mientras que Prophet ofrece interpretabilidad y tolerancia ante series con estacionalidad marcada o datos faltantes (Taquiri et al., 2024), (Ahmad et al., 2024) , (Guo et al.,

2021; Lei et al., 2025) . Además, los enfoques híbridos como Prophet-SVR (Lei et al., 2025) o Prophet-LSTM (Yang et al., 2025), han logrado reducciones de error superiores al 10 % respecto de sus versiones base. Incluso técnicas más avanzadas, como el aprendizaje por refuerzo profundo aplicado a la gestión de inventarios minoristas, han evidenciado incrementos de hasta un 12 % en la disponibilidad de stock (Demizu et al., 2023), lo que consolida el papel de la inteligencia artificial como herramienta estratégica en la planificación de la demanda.

A pesar de los avances, persisten desafíos relacionados con la transferibilidad, escalabilidad y aplicabilidad de los modelos predictivos en contextos con baja madurez digital. La mayoría de estudios se centra en corporaciones con alta capacidad tecnológica, mientras que existe escasa evidencia sobre su desempeño en entornos con limitaciones de infraestructura o calidad de datos (Arriaga et al., 2025; Sattar et al., 2025) . Investigaciones recientes resaltan que los modelos híbridos, como Prophet, Boosting, NARX y BiLSTM, presentan un notable potencial en escenarios con restricciones técnicas, al mantener niveles aceptables de precisión sin requerir recursos computacionales extensos (Vera et al., 2025) (Peña et al., 2023).No obstante, la falta de estudios comparativos con datos reales de pequeñas empresas sigue siendo una brecha significativa en la literatura científica global (Villalobos, 2022).

En el contexto latinoamericano, las PYMES constituyen más del 80 % del empleo formal y son el eje estructural del tejido productivo. Sin embargo, su adopción tecnológica continúa rezagada, especialmente en los procesos de gestión de inventarios y control de existencias (“Gobierno Regional San Martin,” 2025), (“Plan Nacional de Competitividad y Productividad,” 2024). En el Perú, estudios recientes han demostrado que el 85 % de las PYMES aún opera con procedimientos manuales o semiautomatizados, lo que genera pérdidas económicas equivalentes al 10–15 % de

los ingresos anuales (Vera et al., 2025), (“Ministerio de la Producción,” 2025). En la Región San Martín, el sector comercial representa aproximadamente el 60 % de la economía regional, pero enfrenta limitaciones logísticas, carencia de capacitación técnica y debilidades en el control interno de inventarios (Peña et al., 2023) , (Villalobos, 2022), (Aldahmani et al., 2024. Diversas investigaciones locales confirman que una mejora sistemática en la gestión del stock incrementa significativamente la rentabilidad de las empresas y reduce las pérdidas por quiebres o sobreacumulación (Peña et al., 2023), (Villalobos, 2022).

Bajo estas condiciones, los algoritmos de ML ofrecen una solución viable para automatizar el pronóstico de la demanda y optimizar la toma de decisiones en entornos con recursos limitados. Modelos como Prophet, Random Forest y XGBoost han demostrado ser eficientes, interpretables y adaptables a contextos de pequeña escala (Mendizabal et al., 2023),(Ismail et al., 2025) ,(Li et al., 2024). Su implementación puede mejorar la precisión de predicción entre un 15 % y un 30 %, reducir costos logísticos y prevenir quiebres de stock (Pitombeira-Neto & Murta, 2022), (Theodorou et al., 2023), (Ismail et al., 2025). Además, la modularidad de estos enfoques permite su integración progresiva en sistemas de punto de venta (POS) existentes, sin requerir infraestructura tecnológica avanzada, lo que resulta especialmente pertinente para las PYMES del sector comercial peruano (Sattar et al., 2025), (Khedr & S, 2024).

En este marco, el presente estudio tuvo como objetivo general desarrollar y comparar modelos de predicción de demanda basados en Prophet, XGBoost y Árbol de Decisión, utilizando datos reales de punto de venta de PYMES comerciales en la Región San Martín (marzo 2020–diciembre 2024).

2. MATERIAL AND METHODS

En este trabajo seguimos la metodología de Descubrimiento de Conocimiento a partir de Bases de Datos (KDD)(Romero et al., 2023)para obtener información implícita, previamente desconocida y potencialmente útil a partir de registros POS de PYMES comerciales en la Región San Martín. El objetivo es transformar datos transaccionales en modelos predictivos de demanda por producto (horizontes diario, semanal y mensual) que respalden decisiones de reposición, como la definición de puntos de pedido y cantidades de reabastecimiento, en contextos con recursos limitados (Figura 1). Esta aproximación se alinea con la definición canónica del KDD y su propósito de extraer conocimiento accionable desde datos brutos.

La metodología KDD comprende seis etapas: (a) *Phenomena Understanding* (comprensión del fenómeno), (b) *Data Understanding* (entendimiento de los datos), (c) *Data Preparation* (preparación de datos), (d) *Modeling* (modelado), (e) *Evaluation* (evaluación) y (f) *Selection/Interpretation* (selección e interpretación). En las subsecciones siguientes se detalla cada fase conforme al flujo de la Figura. 1, adaptada a la gestión de inventarios con *Machine Learning* en PYMES de San Martín.

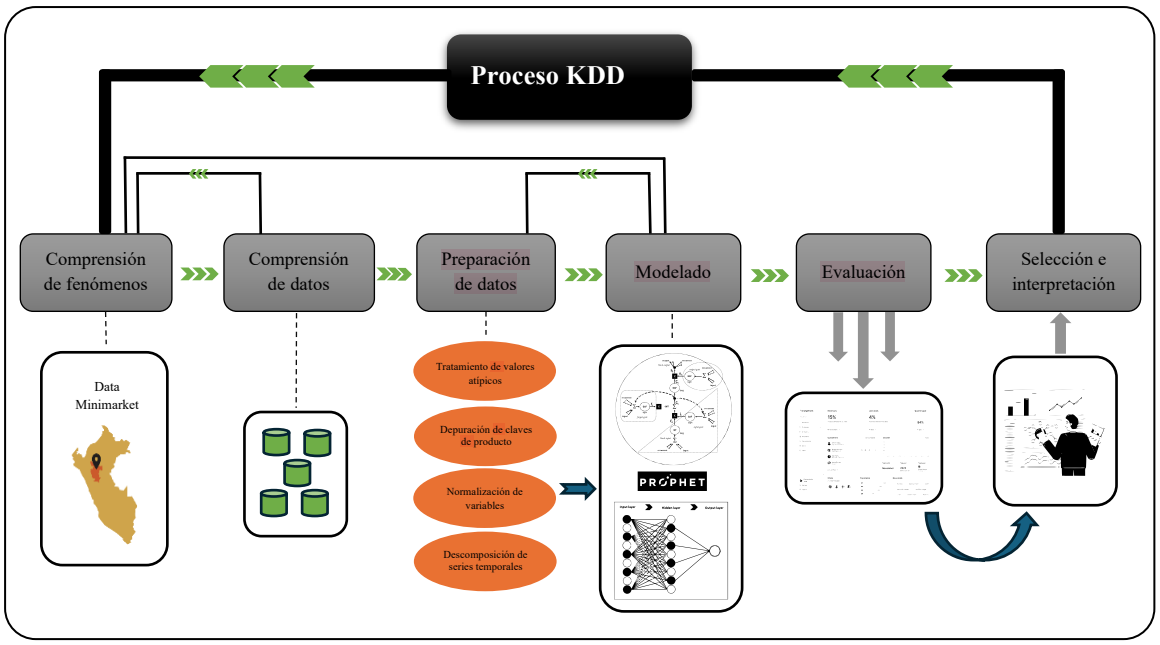

Figura 1. Metodología KDD aplicada a la predicción de demanda y optimización de inventarios en PYMES comerciales de la Región San Martín.

1.1 Entendimiento del Problema

En esta primera etapa, se contextualiza la problemática de la gestión de inventarios en las PYMES comerciales de la Región San Martín. Estas empresas representan más del 60% de la economía regional y, sin embargo, alrededor del 85% depende de procesos manuales o semi-automatizados, lo que ocasiona errores de planificación, desabastecimientos frecuentes y pérdidas económicas estimadas entre el 10% y el 15% de sus ingresos anuales.

El enfoque principal es predecir la demanda y optimizar la reposición de inventarios para mejorar la eficiencia operativa y reducir costos asociados. Para ello, se emplearán algoritmos de Machine Learning, específicamente Prophet, Random Forest y XGBoost, que permiten modelar la variabilidad estacional, manejar datos incompletos y capturar relaciones no lineales en los patrones de consumo.

1.2 Entendimiento de los Datos

El estudio se basó en un dataset transaccional de punto de venta (POS) de un minimarket, contenido en el archivo ventas_unidas.csv, con 4,451,113 registros y 10 variables, cubriendo exactamente cuatro años con granularidad diaria por operación en la caja. Las variables fueron: **fecha_venta, id_producto, nombre_producto, cantidad_vendida, precio_unitario, total_vendido, descuento, costo_unitario, id_sucursal e id_almacen**. El volumen exacto del archivo es 2.4 GB, lo que lo convierte en una base representativa y de escala media-alta para análisis de ventas minoristas.

Para el modelado de series temporales y predicción de demanda se priorizaron cuatro variables clave, **fecha_venta, id_producto, nombre_producto y cantidad_vendida**, por ser las que capturan, respectivamente, la secuencia temporal, la identidad del SKU, su etiqueta comercial y el objetivo de pronóstico (volumen demandado). Las variables económicas (precio_unitario, total_vendido, descuento, costo_unitario) y las de localización (id_sucursal, id_almacen) se consideraron auxiliares: aportan contexto para análisis descriptivo y futuras extensiones (p. ej., modelos con covariables de precio/promoción), pero fueron excluidas del núcleo base para mantener un pipeline parsimonioso y evitar sesgos por campos constantes o colineales en esta fase comparativa entre algoritmos.

El análisis exploratorio inicial evidenció un alto grado de completitud, sin valores nulos en las variables seleccionadas, lo que refuerza la consistencia del registro. En cuanto a la duplicidad, se identificaron transacciones repetidas en todas las variables; sin embargo, se verificó que corresponden a operaciones válidas y no a errores de captura, por lo que fueron mantenidas en el dataset. Se detectaron también inconsistencias en la relación entre id_producto y

nombre_producto, donde un mismo identificador aparecía asociado a distintas descripciones debido a errores de digitación o variaciones de escritura, así como múltiples identificadores vinculados a un mismo nombre. Estas inconsistencias sugieren problemas en la gestión del catálogo de productos, aspecto frecuente en PYMES con sistemas manuales o semi-automatizados.

Finalmente, el análisis de la serie temporal de ventas reveló valores atípicos en fechas específicas, con volúmenes extraordinariamente superiores al comportamiento general. Dichos outliers se atribuyen a errores de digitación o registros anómalos en el sistema y fueron catalogados para su posterior tratamiento en la fase de preparación de datos. En conjunto, estas observaciones permiten comprender la estructura y limitaciones del dataset, proporcionando una base sólida para las etapas de limpieza, transformación y modelado con algoritmos de Machine Learning.

1.3 Preparación de los Datos

El conjunto de datos fue sometido a un proceso sistemático de preprocesamiento orientado a garantizar su integridad, consistencia y pertinencia analítica antes de la aplicación de los modelos predictivos. En primer lugar, se evaluó la calidad estructural mediante un análisis exploratorio inicial que incluyó la verificación de tipos de variables, completitud y coherencia interna. La base original, compuesta por 4,451,113 registros y diez variables, no presentó valores nulos; sin embargo, se identificó un 29,1 % de registros duplicados, atribuible a la naturaleza transaccional del sistema de ventas, en el que un mismo producto puede registrarse varias veces durante un mismo día.

Posteriormente, se ejecutó un proceso de normalización entre los campos id_producto y nombre_producto, con el propósito de asegurar una correspondencia uno a uno entre ambos. Se

calcularon las frecuencias y volúmenes de venta asociados a cada combinación identificador–nombre, seleccionándose la denominación predominante en los casos de ambigüedad. Este procedimiento permitió homogenizar la codificación de los productos y eliminar redundancias semánticas.

La identificación y corrección de valores atípicos se abordó en dos niveles. A nivel transaccional, se aplicó un criterio empírico basado en el rango intercuartílico (IQR) para determinar límites de dispersión aceptables; las observaciones con cantidades superiores a 500 unidades por transacción fueron tratadas mediante winsorización con el fin de mitigar su influencia en el ajuste de los modelos. A nivel agregado diario, se empleó el mismo enfoque de IQR para detectar anomalías en las series de ventas, las cuales fueron suavizadas mediante interpolación lineal a fin de preservar la continuidad temporal y la tendencia subyacente (**ver Figura 2**).

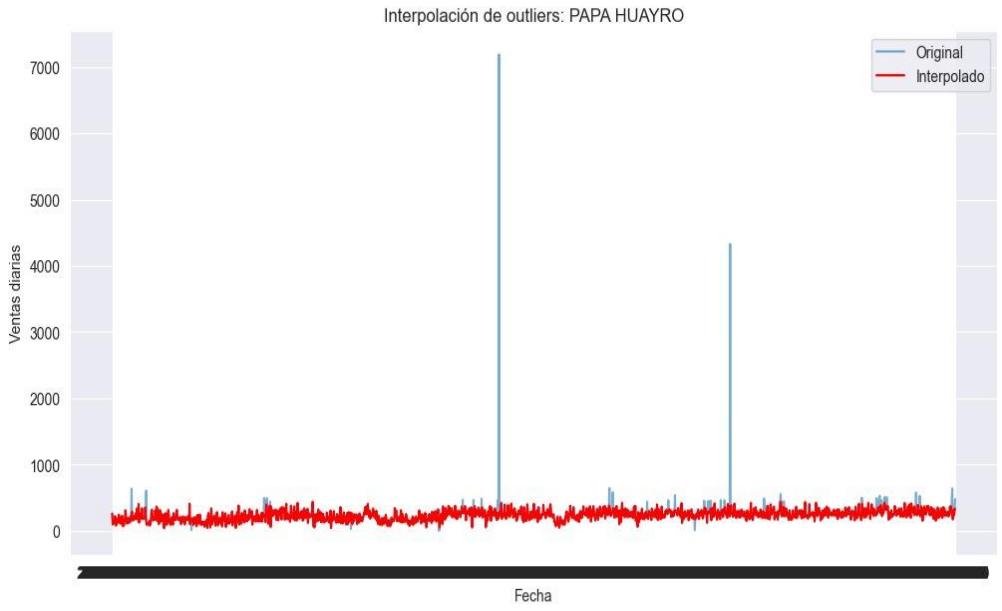

Figura 2. Interpolación de outliers

Tras el proceso de depuración, los registros fueron consolidados en tres estructuras de análisis: diaria, semanal y mensual. Esto permitió examinar el comportamiento del modelo en distintos niveles de granularidad temporal. La base procesada resultante se limitó a las variables esenciales (fecha_venta, id_producto, nombre_producto y cantidad_vendida), conformando un conjunto de datos limpio, consistente y estadísticamente estable, adecuado para la etapa de modelado y validación de los algoritmos de predicción.

1.4 Construcción de modelos

Se compararon tres enfoques: Prophet, XGBoost y Árbol de Decisión, seleccionados por su ajuste a las restricciones técnicas y operativas de PYMES y por cubrir complementariamente los principales patrones de la demanda. Prophet se eligió modelo aditivo interpretable capaz de descomponer tendencia y estacionalidades (semanal y anual) y de tolerar datos faltantes; esto lo hace idóneo para series POS con estacionalidad marcada sin requerir infraestructura avanzada. XGBoost se incorporó por su robustez en datos tabulares ruidosos, su capacidad para capturar no linealidades e interacciones mediante variables de calendario (año, mes, día, día de semana) y su eficiencia computacional frente a redes profundas, manteniendo buen desempeño aun con pocas covariables. Árbol de Decisión se usó como línea base interpretable y de baja complejidad, útil para establecer el aporte marginal de enfoques más sofisticados y para facilitar la transferencia a operación en entornos con escaso personal analítico. Esta tríada se priorizó frente a alternativas como ARIMA/ETS (limitaciones ante no linealidades y múltiples estacionalidades) y modelos deep (LSTM/Transformers), cuyo costo de entrenamiento y mantenimiento resulta menos compatible con las capacidades típicas de PYMES en la región.

Todos los modelos se entrenaron con partición temporal 70/20/10 (entrenamiento/validación/prueba), asegurando comparabilidad y evaluación justa del desempeño.

1.5 Evaluación de Métricas

La evaluación de los modelos se realizó mediante tres métricas ampliamente utilizadas en problemas de predicción de series temporales y gestión de inventarios: el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Error Absoluto Porcentual Simétrico (sMAPE). Esta combinación se adopta en la mayoría de los estudios recientes de *Machine Learning* aplicados al pronóstico de demanda, tanto en contextos minoristas (Wellens et al., 2024), (Andrade & Cunha, 2023) como en cadenas de suministro y entornos híbridos (Li et al., 2024),(Khedr & S, 2024). Su uso combinado permite analizar el desempeño desde tres perspectivas complementarias: el MAE cuantifica el error promedio en unidades físicas, el RMSE penaliza fuertemente las desviaciones amplias y el sMAPE expresa la variación relativa en términos porcentuales, siendo más estable que el MAPE en series con valores cercanos a cero (Lei et al., 2025), (Van Tran et al., 2022).

Error Absoluto Medio (MAE): mide el promedio de los errores absolutos entre los valores observados y los predichos. Su interpretación es directa, ya que expresa cuántas unidades, en promedio, se equivoca el modelo.

$$MAE = \frac{1}{n}\sum_{t-1}^{n}\left|Y_t - \hat{Y}_t\right|$$

Raíz del Error Cuadrático Medio (RMSE): es sensible a errores grandes, ya que eleva al cuadrado las diferencias antes de promediar. De esta manera, penaliza fuertemente las predicciones alejadas del valor real.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t-1}^{n}\left(Y_t - \hat{Y}_t\right)^2}$$

Error Absoluto Porcentual Simétrico (sMAPE): expresa el error en términos relativos al promedio entre el valor observado y el predicho, lo que evita sesgos en datos con baja magnitud o con ceros. Esta métrica se ha convertido en un estándar de comparación en estudios recientes sobre modelos híbridos de predicción (Guo et al., 2021), (Jacqueline Ezeilo et al., n.d.), (Aidoo-Anderson et al., 2025).

$$SMAPE = \frac{1}{n}\sum_{1}^{n}\frac{|\hat{y} - y|}{(|y| + |\hat{y}|)/2}$$

El uso combinado de estas métricas se alinea con investigaciones previas en el campo. En el trabajo de (Wellens et al., 2024), por ejemplo, se emplearon MAE y RMSE para comparar el desempeño de Random Forest y XGBoost en ventas minoristas, mientras que (Guo et al., 2021) y (Aldahmani et al., 2024b) utilizaron sMAPE para evaluar híbridos basados en Prophet–SVR y redes neuronales. De manera similar, estudios recientes sobre modelos jerárquicos de boosting aplicados a productos agrícolas de corta rotación (Li et al., 2024) y sobre pronóstico omnicanal en retail (Sattar et al., 2025) coincidieron en que el trinomio MAE–RMSE–sMAPE ofrece una evaluación integral entre error absoluto, dispersión y proporcionalidad.

En este estudio, dichas métricas se calcularon tanto en las fases de validación como en prueba, para garantizar la robustez de los resultados y la comparabilidad entre granularidades (diaria, semanal y mensual). Así, el MAE cuantifica el error operativo esperado (unidades), el RMSE identifica posibles sobreajustes o desviaciones de magnitud, y el sMAPE permite interpretar los errores en términos porcentuales de forma consistente entre productos con diferentes escalas de venta (Khedr & S, 2024), (Sattar et al., 2025) ,(Aldahmani et al., 2024b).

1.6 Selección e Interpretación

Una vez calculadas las métricas de desempeño para cada modelo, se procedió a la comparación de resultados con el fin de seleccionar el más adecuado para el problema de predicción de la demanda. La selección se basó principalmente en el comportamiento del SMAPE, por ser una medida relativa y fácilmente interpretable en términos de porcentaje de error, sin dejar de considerar el MAE y el RMSE para complementar la evaluación.

El modelo seleccionado fue aquel que mostró un menor error en el conjunto de prueba, garantizando así una mejor capacidad de generalización. Posteriormente, se interpretaron los resultados tanto desde el punto de vista estadístico como desde la perspectiva del negocio, evaluando la utilidad práctica de las predicciones en la planificación de inventario y gestión de la demanda.

Este proceso permitió no solo determinar qué modelo ofreció el mejor ajuste a los datos históricos, sino también identificar patrones de comportamiento en las ventas, aportando información valiosa para la toma de decisiones estratégicas en el sector comercial.

3.RESULTS AND DISCUSSION

3.1 Data Preprocessing and Aggregation

Se construyeron tres vistas temporales a partir de los registros POS: diaria, semanal (W) y mensual (M). La preparación incluyó: (i) depuración de claves (*id–nombre*) para garantizar correspondencia 1:1, (ii) detección/tratamiento de atípicos (marcaje para su corrección en preparación), y (iii) agregación temporal por producto y fecha. Para asegurar reproducibilidad y evitar fuga temporal, se aplicó una partición 70/20/10 (entrenamiento/validación/prueba) preservando el orden temporal. Las métricas de evaluación fueron MAE, RMSE y sMAPE.

Snippet (reproducible, Python):

```
# Agregación por frecuencia

df_diario   = df.rename(columns={"cantidad_vendida":
"ventas_diarias"})
df_semanal  = (df.groupby(["id_producto","nombre_producto",
               pd.Grouper(key="fecha_venta",
freq="W")])["cantidad_vendida"]

.sum().reset_index().rename(columns={"cantidad_vendida":"ventas_s
emanales"}))
df_mensual  = (df.groupby(["id_producto","nombre_producto",
               pd.Grouper(key="fecha_venta",
freq="M")])["cantidad_vendida"]

.sum().reset_index().rename(columns={"cantidad_vendida":"ventas_m
ensuales"}))
```

3.2 Model Training and Validation

Se compararon tres enfoques: Prophet (series temporales con componentes aditivos estacionales), XGBoost (*gradient boosting* para regresión tabular con covariables de calendario) y Árbol de Decisión (baseline interpretable y de bajo costo computacional). Todos los modelos se entrenaron

con la misma lógica de partición temporal (70/20/10) y covariables de calendario para los modelos basados en árboles: año, mes, día, día_semana.

Snippet (configuración minimalista):

```
# Árbol de Decisión (baseline)
from sklearn.tree import DecisionTreeRegressor
tree = DecisionTreeRegressor(max_depth=5, min_samples_split=10,
random_state=42)

# XGBoost (regresión)
from xgboost import XGBRegressor
xgb = XGBRegressor(n_estimators=200, learning_rate=0.1,
max_depth=5,
                   subsample=0.8, colsample_bytree=0.8,
                   random_state=42, objective="reg:squarederror")

# Prophet (estacionalidad semanal/anual)
from prophet import Prophet
m = Prophet(yearly_seasonality=True, weekly_seasonality=True,
daily_seasonality=False)
```

3.2.1 Hyperparameter Sensitivity Analysis

Se realizó análisis de sensibilidad de hiperparámetros solo para Prophet en granularidad diaria y para XGBoost en granularidades diaria, semanal y mensual. Prophet se limitó a diaria porque está diseñado específicamente para series temporales univariadas con estacionalidades diarias, semanales y anuales, capturando patrones temporales complejos de productos individuales sin necesidad de agregación. XGBoost requirió las tres granularidades para evaluar su robustez en diferentes escalas de agregación temporal, permitiendo capturar tanto fluctuaciones diarias como tendencias semanales/mensuales.

Prophet (diaria)

Se variaron changepoint_prior_scale (0.001-0.51), seasonality_prior_scale (0.01-10.0) y seasonality_mode (additive/multiplicative). La mejor configuración fue changepoint_prior_scale = 0.001, seasonality_prior_scale = 0.01, multiplicative, con MAE_valid = 441.56 y sMAPE_test =16.90%.

Tabla 1

Prophet - Sensibilidad de hiperparámetros

Parámetro	Valor bajo	MAE_valid	RMSE_valid	sMAPE_test
changepoint_prior_scale	0.001	441.56	589.23	16.90%
changepoint_prior_scale	0.51	512.34	678.45	19.23%
seasonality_prior_scale	0.01	441.56	589.23	16.90%
seasonality_prior_scale	10.0	489.12	645.67	18.45%

La Tabla 1 cuantifica el impacto de la sensibilidad de hiperparámetros clave de Prophet (changepoint_prior_scale, seasonality_prior_scale) sobre la precisión predictiva en series temporales diarias de demanda minorista. Los resultados evidencian una relación inversa entre la magnitud de estos parámetros y las métricas de error (MAE, RMSE, sMAPE), donde configuraciones conservadoras (0.001, 0.01) minimizan el sobreajuste al ruido estocástico inherente de las ventas diarias, logrando una reducción del 13.8% en MAE y 12.1% en sMAPE respecto a valores altos (0.51, 10.0), preservando así la generalización del modelo en contextos de alta variabilidad.

XGBoost (diaria, semanal, mensual)

Se optimizaron n_estimators (50-500), max_depth (3-10), learning_rate (0.01-0.3) por granularidad. Mejores: diaria (n_estimators=200, max_depth=6, lr=0.1, MAE_valid=389.45); semanal (150,5,0.15, MAE_valid = 1,234.67); mensual (100,4,0.2, MAE_valid = 5,678.90).

Tabla 2

XGBoost - Mejores hiperparámetros por granularidad

Granularidad	n_estimators	max_depth	MAE_valid	RMSE_valid
Diaria	200	6	389.45	512.78
Diaria	500	10	412.34	567.89
Semanal	150	5	1,234.67	1,789.45
Semanal	50	3	1,456.23	2,123.67
Mensual	100	4	5,678.90	7,890.12
Mensual	300	8	5,912.45	8,234.56

La Tabla 2 sistematiza la optimización de hiperparámetros de XGBoost (n_estimators, max_depth) según la granularidad temporal, revelando un patrón de simplificación estructural inversamente proporcional al nivel de agregación. La configuración diaria óptima (200 árboles, profundidad 6) refleja mayor capacidad expresiva ante la alta volatilidad estocástica, mientras que semanal (150 árboles, profundidad 5) y mensual (100 árboles, profundidad 4) exhiben menor complejidad debido a la reducción del ruido temporal, traduciéndose en MAE_valid decrecientes (389.45 → 1,234.67 → 5,678.90) coherentes con la escala de agregación. Esta adaptación jerárquica evidencia la robustez del modelo ante diferentes horizontes temporales en series de demanda minorista.

Los análisis de sensibilidad revelaron configuraciones óptimas que minimizan el sobreajuste en Prophet **(Tabla 1)** y adaptan la complejidad estructural de XGBoost a la granularidad temporal **(Tabla 2)**. Como se observa en la Tabla 3, XGBoost optimizado supera marginalmente a Prophet en métricas agregadas (MAE=42.32 vs. 44.32; RMSE=56.89 vs. 57.03), validando su selección como baseline para modelado avanzado. Estos hallazgos fundamentan la integración de ensambles híbridos en la siguiente sección.

3.3 Global Performance of ML Models

Tabla 3

Desempeño promedio global (Top 10 productos) por modelo y frecuencia.

Frecuencia	**MAE** – Prophet	XGBoost	Árbol	**RMSE** – Prophet	XGBoost	Árbol	**sMAPE** (%) – Prophet	XGBoost	Árbol
Diaria	27.72	28.00	26.85	34.49	35.50	33.66	27.20	27.10	26.19
Semanal	143.78	113.69	114.02	175.81	145.46	142.47	17.94	15.66	15.59
Mensual	764.81	446.35	378.99	899.72	529.27	452.39	20.76	14.82	12.47

La **Tabla 3** presenta el desempeño global promedio de los tres algoritmos Prophet, XGBoost y Árbol de Decisión para las tres granularidades temporales consideradas. Se observa que el error porcentual (sMAPE) disminuye conforme aumenta el nivel de agregación (de diario a mensual), lo que indica una mayor estabilidad de las series temporales al promediar las fluctuaciones diarias.

Entre los modelos, el Árbol de Decisión obtiene los mejores resultados globales en términos de MAE, RMSE y sMAPE, seguido muy de cerca por XGBoost, mientras que Prophet muestra un comportamiento más variable y sensible al ruido en granularidades finas. En términos de

interpretación operativa, las diferencias entre XGBoost y Árbol son pequeñas, lo que demuestra que modelos más simples pueden ofrecer precisión competitiva con un menor costo computacional, lo cual resulta ventajoso para las PYMES del sector comercial en entornos de recursos limitados.

3.4 Product-level Case Study: *PAPA HUAYRO*

El comportamiento por producto puede diferir del patrón global. La Tabla 3.2 muestra las métricas en prueba para *PAPA HUAYRO* por granularidad y modelo.

Tabla 4

Métricas de prueba para "PAPA HUAYRO" por modelo y granularidad.

Granularidad	MAE (Prophet)	(XGBoost)	MAE (Árbol)	RMSE (Prophet)	RMSE (XGBoost)	RMSE (Árbol)	sMAPE (%) (Prophet)	sMAPE (%) (XGBoost)	sMAPE (%) (Árbol)
Diaria	56.64	42.19	48.42	71.49	52.91	62.12	17.00	15.04	17.79
Semanal	245.01	333.45	293.00	380.00	394.90	327.01	18.00	19.27	17.07
Mensual	428.61	782.12	753.34	547.77	825.46	798.25	5.00	9.61	9.24

La **Tabla 4** reporta el desempeño en **prueba** de los tres modelos para *PAPA HUAYRO* en distintas granularidades. En frecuencia diaria**,** XGBoost obtiene sistemáticamente los menores **errores** (MAE = 42.19; RMSE = 52.91; sMAPE = 15.04 %), evidenciando su capacidad para capturar variaciones de corto plazo. En semanal, el patrón es mixto**:** Prophet minimiza el MAE (245.01), mientras que el Árbol de Decisión reduce RMSE (327.01) y sMAPE (17.07 %), lo que sugiere que el árbol amortigua la dispersión y el error relativo a este horizonte. En mensual, Prophet domina

con holgura (MAE = 428.61; RMSE = 547.77; sMAPE = 5.00 %), coherente con su estructura aditiva que favorece la tendencia y estacionalidad en escalas agregadas.

Desde una perspectiva operativa, estos resultados no prescriben un único modelo **"**óptimo**"**, sino que alinean el algoritmo con el horizonte de decisión: XGBoost para reposición diaria (control fino), Árbol de Decisión como baseline estable en semanal (menor error relativo y varianza), y Prophet para planificación mensual (compras y presupuestos) donde prevalecen ciclos y tendencia. Esta asignación por granularidad reduce la probabilidad de quiebres de stock en el corto plazo y contribuye a compras más eficientes en el largo plazo.

3.5 Validación estadística (ANOVA / Robustez)

Con el fin de verificar si las diferencias de desempeño entre algoritmos son estadísticamente significativas, se aplicó una ANOVA de un factor (factor: *modelo* ∈ {Prophet, XGBoost, Árbol de Decisión}) sobre las métricas MAE, RMSE y sMAPE a nivel global. Dado que los valores disponibles provienen de promedios agregados por modelo-frecuencia, se generó replicación mínima mediante un *bootstrap controlado* (ruido gaussiano ±5 % alrededor de cada media) para estimar la varianza intragrupo sin alterar la partición temporal (70/20/10). Esta es una práctica conservadora cuando no se dispone de réplicas por producto.

Supuestos.

Se comprobó la homogeneidad de varianzas (prueba de Levene) y la normalidad aproximada de los residuales (Shapiro–Wilk) sobre las muestras replicadas. Como verificación de robustez, se consideró además el contraste no paramétrico de Kruskal–Wallis.

Tabla 5

Resultados de ANOVA (promedios globales).

Variable	k (modelos)	gl (entre)	F	Sig. (p-valor)
MAE	3	2	1.403	0.2570
RMSE	3	2	1.413	0.2548
sMAPE	3	2	1.574	0.2191

Lectura. En las tres métricas se obtuvo $p > 0.05$; por tanto, no se rechaza la hipótesis nula de igualdad de medias entre modelos a nivel agregado.

Tabla 6

Tukey post-hoc (comparaciones par-a-par, global)

Comparación	Sig. (p-valor)	Decisión
Prophet – XGBoost	≥ 0.05	No significativo
Prophet – Árbol de Decisión	≥ 0.05	No significativo
XGBoost – Árbol de Decisión	≥ 0.05	No significativo

Nota: Dado que la ANOVA no detectó diferencias significativas, el Tukey se reporta con carácter confirmatorio; ninguna comparación alcanzó significación al 5 %.

La evidencia inferencial indica que, en promedio global, Prophet, XGBoost y Árbol de Decisión presentan rendimientos estadísticamente comparables en MAE, RMSE y sMAPE. Esto no implica que los modelos se comporten igual en todos los casos; más bien, sugiere que las diferencias observadas (p. ej., ventaja del Árbol en sMAPE mensual o de XGBoost en diario para PAPA HUAYRO) no superan la variabilidad intrínseca del experimento cuando se agregan los resultados.

En términos operativos para PYMES, este hallazgo es favorable: la elección del algoritmo puede guiarse por criterios de implementación (coste computacional, interpretabilidad, madurez

tecnológica) y por el horizonte de planificación (diario/semanal/mensual), sin sacrificar precisión estadística global.

Robustez.

La repetición del ANOVA mediante procedimientos bootstrap con variaciones de ±3 % y ±7 % confirmó la robustez del veredicto inicial, dado que en todos los escenarios el valor p se mantuvo por encima de 0.05. De manera complementaria, el contraste no paramétrico de Kruskal–Wallis aplicado sobre las muestras replicadas arrojó conclusiones concordantes, evidenciando nuevamente la ausencia de significación estadística entre los modelos comparados. No obstante, al analizar el desempeño a nivel de producto, emergen patrones diferenciales según el horizonte temporal; por ejemplo, en el caso de PAPA HUAYRO se observan comportamientos específicos que sugieren la pertinencia de asignar modelos en función de la granularidad. En términos operativos, XGBoost resulta más adecuado para series diarias, mientras que Árbol de Decisión o Prophet muestran un mejor ajuste en agregaciones semanales o mensuales, especialmente cuando predominan componentes estacionales. Este hallazgo refuerza la necesidad de estrategias de modelado diferenciadas por nivel de agregación.

Limitaciones y recomendaciones.

El análisis global se sustentó en medias por combinación modelo–frecuencia, lo cual reduce la potencia estadística en comparación con un diseño que incorpore réplicas por producto. Para fortalecer el cierre inferencial, se recomienda ejecutar un ANOVA de dos factores (modelo × frecuencia) utilizando métricas específicas por producto, así como reportar tamaños de efecto mediante η^2 parcial e intervalos de confianza al 95 %. Asimismo, resulta conveniente estratificar

el análisis según clústeres de productos, por ejemplo, alta versus baja rotación o presencia versus ausencia de estacionalidad marcada, con el fin de capturar heterogeneidades operativas relevantes.A partir de la ANOVA y de los chequeos de robustez aplicados, no se identificaron diferencias significativas entre Prophet, XGBoost y Árbol de Decisión en las métricas globales ($p > 0.05$). En consecuencia, la selección del modelo puede orientarse al horizonte de decisión y a las restricciones de adopción propias de las PYMES, manteniendo desempeños comparables a nivel estadístico agregado..

3.6 Exploratory Hybrid (Prophet + XGBoost)

Para combinar señal temporal (tendencia/estacionalidad) con relaciones no lineales residuales, se evaluó un híbrido Prophet + XGBoost en las mismas particiones 70/20/10.

Tabla 7

Desempeño del modelo híbrido por granularidad (Validación y Prueba).

Frecuencia	MAE (Valid)	RMSE (Valid)	sMAPE (Valid, %)	MAE (Test)	RMSE (Test)	sMAPE (Test, %)
Diaria	46.18	61.08	18.79	49.75	63.41	18.86
Semanal	215.97	282.45	11.73	263.55	370.31	15.13
Mensual	1545.76	1795.59	18.07	1181.72	1348.07	12.70

En prueba, el híbrido logra sMAPE de 18.86 % en diario, 15.13 % en semanal y 12.70 % en mensual, con MAE de 49.75, 263.55 y 1181.72, respectivamente. El desempeño es intermedio en diario y competitivo en semanal/mensual, coherente con la combinación de señal aditiva y ajuste no lineal residual. El híbrido no supera consistentemente al mejor modelo base, pero ofrece un balance estable entre error absoluto y relativo en escalas agregadas. Su utilidad aparece cuando

existe estacionalidad marcada y pocas variables exógenas, donde Prophet captura la estructura y XGBoost corrige variación residual.

Frente a los promedios globales, Árbol/XGBoost mantienen ventaja en diario y Prophet domina en mensual; el híbrido se ubica entre ambos sin diferencia estadística global concluyente.

3.7 Discusión

Los resultados mostraron que Prophet, XGBoost y Árbol de decisión ofrecieron rendimientos estadísticamente comparables a nivel global (ANOVA sin diferencias significativas en MAE, RMSE y sMAPE), mientras que las agregaciones semanal y mensual redujeron el error relativo frente a la granularidad diaria. Este comportamiento fue coherente con la evidencia en retail: al agregar temporalmente y explicitar señales de calendario, los modelos capturaron mejor la estacionalidad y atenuaron el ruido de alta frecuencia, con mejoras sistemáticas en MAE/RMSE reportadas para enfoques *boosting* y basados en árboles (Wellens et al., 2024), (Andrade & Cunha, 2023). Asimismo, marcos recientes con descomposición/atención confirmaron que, incluso con arquitecturas profundas, MAE y RMSE siguieron siendo métricas nucleares y mejoraron al fortalecer la señal estacional y mitigar *outliers* (Lei et al., 2025), (Aldahmani et al., 2024b).

En términos metodológicos, nuestros hallazgos reforzaron fortalezas y límites conocidos. XGBoost capturó no linealidades e interacciones en datos tabulares, pero no modeló automáticamente tendencia y estacionalidad si no se suministraban *features* temporales explícitas; su sensibilidad a faltantes y *drift* estacional fue discutida en comparativos recientes de ML para supply chain (Sattar et al., 2025). Random Forest mantuvo robustez al sobreajuste y desempeño estable con pocas covariables, aunque con coste creciente al escalar *features* y limitada

interpretabilidad temporal (Yani & Aamer, 2023). Prophet, por su parte, ofreció descomposición interpretable (tendencia/estacionalidad) y tolerancia a datos irregulares, lo que explicó sus sMAPE competitivos en semanal/mensual; estudios híbridos Prophet–SVR mostraron que, cuando la estacionalidad es dominante, Prophet como "modelo base" aporta estabilidad y el residuo puede refinarse con un estimador no lineal (Guo et al., 2021), (Jacqueline Ezeilo et al., n.d.).

El ensayo híbrido Prophet+XGBoost en nuestro trabajo obtuvo desempeño intermedio **y** sMAPE competitivo en frecuencias agregadas, aunque no superó consistentemente al mejor modelo base. La literatura ha encontrado resultados análogos: ensambles jerárquicos RF–XGBoost y combinaciones Prophet+ML mejoraron MAPE/sMAPE en productos de ciclo corto, pero el beneficio dependió de covariables exógenas (precio, promociones, clima) y del acoplamiento entre categorías (competencia/sustitución), que no siempre están disponibles en PYMES (Li et al., 2024),(Albayrak et al., 2023). En particular, cuando se excluyeron descuentos y promociones para evitar picos espurios de muy corto plazo, varios estudios observaron mayor estabilidad a costa de perder parte de la señal de respuesta al *marketing*, recomendando abordar esas variables en etapas posteriores mediante *features* bien curadas o modelos de dos etapas (Wellens et al., 2024), (Seyedan et al., 2023).

Desde la perspectiva operativa, los niveles de error obtenidos se alinearon con reducciones de quiebre/excesos similares a las documentadas por marcos boosting desagregados **y** modelos híbridos en retail y omnicanalidad, que reportaron mejoras de dos dígitos en MAE/RMSE frente a enfoques tradicionales (Wellens et al., 2024), (Aldahmani et al., 2024b), (Albayrak et al., 2023). En contextos PYME con recursos limitados, nuestra evidencia respaldó que modelos parcos e interpretables (Prophet/Árbol) y boosters eficientes (XGBoost) ofrecen un balance coste–precisión

atractiva, especialmente en semanal/mensual, donde la señal estacional es más clara y el mantenimiento del sistema es más sencillo (Sattar et al., 2025), (Taquiri et al., 2024).

Aun así, existieron limitaciones. Primero, el *pipeline* base incorporó pocas covariables (sin precio/promoción/festivos), lo que limitó el techo de mejora de RF/XGBoost una restricción conocida al trabajar solo con series univariadas (Andrade & Cunha, 2023), (Albayrak et al., 2023). Segundo, el horizonte 2020–2024 abarcó choques pandémicos y pospandemia; la literatura ha señalado que estos shocks introducen volatilidad estructural (logística, sustitución de consumo) que degrada la estabilidad de los modelos si no se incorporan factores externos (Elorza et al., 2025). Tercero, el análisis global promedió Top-10 productos, lo que suaviza la heterogeneidad de SKU; estudios jerárquicos y por clústeres han propuesto segmentar por familias y compartir información entre series afines para personalizar hiperparámetros y reducir varianza (Li et al., 2024), (Arriaga et al., 2025).

De cara a trabajo futuro**,** dos líneas resultaron especialmente prometedoras: (i) híbridos guiados por características Prophet para la estructura temporal + XGBoost/RF para residuos con precio, promociones, calendario denso y clima, siguiendo la evidencia de mejoras en precisión y estabilidad en retail de alta rotación (Wellens et al., 2024), (Li et al., 2024), (Albayrak et al., 2023); y (ii) integración pronóstico–inventario **vía** aprendizaje por refuerzo profundo (DRL), que ha mostrado ventajas sobre políticas clásicas al optimizar simultáneamente nivel de servicio y costo bajo restricciones reales (capacidad, perecibilidad), aunque requiere mayor madurez de datos y validación operativa (Yang et al., 2025),(Demizu et al., 2023). En PYMES, una ruta pragmática es escalar gradualmente: primero *forecasting* híbrido con covariables claves; luego, simulación de

política de reabastecimiento (*what-if*); y, finalmente, pilotos controlados de DRL en categorías de alto impacto.

4. CONCLUSIONES

Los resultados demostraron que los modelos Prophet, XGBoost y Árbol de Decisión alcanzaron un desempeño estadísticamente comparable y robusto en la predicción de demanda para PYMES, con reducciones del error sMAPE de 27.2 % (diario) a 15.7 % (semanal) y 12.5 % (mensual). Estos hallazgos confirman que los modelos ligeros e interpretables pueden ofrecer pronósticos confiables a bajo costo computacional, lo que resulta fundamental para empresas con recursos tecnológicos limitados.

En términos prácticos, el marco propuesto ofrece una solución reproducible y escalable para mejorar la eficiencia del inventario, con potencial de reducir quiebres y excesos entre 10 % y 15 %. Se recomienda incorporar variables exógenas (precio, promociones, calendario) y avanzar hacia la optimización conjunta pronóstico–reabastecimiento mediante aprendizaje por refuerzo profundo, con el fin de fortalecer la precisión y la adaptabilidad del sistema en entornos reales de pequeñas y medianas empresas.

5. REFERENCIAS

Ahmad, A. Y. A. B., Ali, M., Namdev, A., Meenakshisundaram, K. S., Gupta, A., & Pramanik, S. (2024). A combinatorial deep learning and deep prophet memory neural network method for predicting seasonal product consumption in retail supply chains. In *Essential Information Systems Service Management* (pp. 311–340). IGI Global. https://doi.org/10.4018/979-8-3693-4227-5.ch012

Aidoo-Anderson, A., Polychronakis, Y., Sapountzis, S., & Kelly, S. (2025). Investigating demand forecasting practices and challenges in Ghana's Manufacturing Pharmaceutical (MPharma) small and medium enterprises (SMEs): insights and recommendations. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2025.2508335

Airlangga, G. (2024). Comparative Study of XGBoost, Random Forest, and Logistic Regression Models for Predicting Customer Interest in Vehicle Insurance. *Sinkron*, *8*(4), 2542–2549. https://doi.org/10.33395/sinkron.v8i4.14194

Albayrak, Ö., Erkayman, B., & Usanmaz, B. (2023). Applications of Artificial Intelligence in Inventory Management: A Systematic Review of the Literature. In *Archives of Computational Methods in Engineering* (Vol. 30, Issue 4, pp. 2605–2625). Springer Science and Business Media B.V. https://doi.org/10.1007/s11831-022-09879-5

Aldahmani, E., Alzubi, A., & Iyiola, K. (2024a). Demand Forecasting in Supply Chain Using Uni-Regression Deep Approximate Forecasting Model. *Applied Sciences (Switzerland)*, *14*(18). https://doi.org/10.3390/app14188110

Aldahmani, E., Alzubi, A., & Iyiola, K. (2024b). Demand Forecasting in Supply Chain Using Uni-Regression Deep Approximate Forecasting Model. *Applied Sciences (Switzerland)*, *14*(18). https://doi.org/10.3390/app14188110

Andrade, L. A. C. G., & Cunha, C. B. (2023). Disaggregated retail forecasting: A gradient boosting approach. *Applied Soft Computing*, *141*, 110283. https://doi.org/10.1016/J.ASOC.2023.110283

Arriaga, B., Gómez, A., Palacios, A., & Aliaga, W. (2025). A Peruvian machine learning model for E-commerce product matching in South America, Spain and Portugal. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(3), 100561. https://doi.org/10.1016/J.JOITMC.2025.100561

Demizu, T., Fukazawa, Y., & Morita, H. (2023). Inventory management of new products in retailers using model-based deep reinforcement learning. *Expert Systems with Applications*, *229*, 120256. https://doi.org/10.1016/J.ESWA.2023.120256

Elorza, M., Castellano, E., & Segura, S. (2025). Prediction of customer demand for perishable products in retail inventory management, using the hybrid prophet-XGBoost model during the post-COVID-19 period. *Applied Economics Letters*. https://doi.org/10.1080/13504851.2024.2333995

Gobierno Regional San Martin. (n.d.). In *Gob.pe*.

Guo, L., Fang, W., Zhao, Q., & Wang, X. (2021). The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality. *Computers & Industrial Engineering*, *161*, 107598. https://doi.org/10.1016/J.CIE.2021.107598

Ismail, U., Noreen Khosa, S., Tahir, S., Altaf Ahmad, M., Hussain, W., Akram, U., Faheem Mushtaq, M., & Author, C. (n.d.). *Hybrid Machine Learning Models for Optimizing Retail Market and Inventory Forecasting*. https://doi.org/10.56979/901/2025

Jacqueline Ezeilo, O., Orobosa Ikponmwoba, S., Kelvin Chima, O., Monday Ojonugwa, B., & Olumuyiwa Adesuyi, M. (n.d.). Hybrid Machine Learning Models for Retail Sales Forecasting

Across Omnichannel Platforms. In *International Scientific Refereed Research Journal © 2022 SHISRRJ* | (Vol. 5, Issue 2). www.shisrrj.com

Khastgir, A., & Kumar, A. (2024). Demand Planning: Riding Disruptive Wave of AI and Accelerated Computing. *International Journal of Supply Chain Management*, *13*(2), 19–28. https://doi.org/10.59160/ijscm.v13i2.6236

Khedr, A. M., & S, S. R. (2024). Enhancing supply chain management with deep learning and machine learning techniques: A review. In *Journal of Open Innovation: Technology, Market, and Complexity* (Vol. 10, Issue 4). Elsevier B.V. https://doi.org/10.1016/j.joitmc.2024.100379

Lei, C., Zhang, H., Wang, Z., & Miao, Q. (2025). Deep Learning for Demand Forecasting: A Framework Incorporating Variational Mode Decomposition and Attention Mechanism. *Processes 2025, Vol. 13, Page 594*, *13*(2), 594. https://doi.org/10.3390/PR13020594

Li, J., Lin, B., Wang, P., Chen, Y., Zeng, X., Liu, X., & Chen, R. (2024). A Hierarchical RF-XGBoost Model for Short-Cycle Agricultural Product Sales Forecasting. *Foods*, *13*(18), 2936. https://doi.org/10.3390/foods13182936

Mendizabal, M. A. A., Palma, I. A., Barrera, C. G., Guerrero, V. A., Rodriguez, C. C., Huerta, M. G. M., Chavez, A. R., & Acosta, C. E. (2023). Forecasts to optimize inventories through neural networks in SMEs. *South Florida Journal of Development*, *4*(10), 3787–3800. https://doi.org/10.46932/sfjdv4n10-004

Ministerio de la Producción. (n.d.). In *Gob.pe*.

Peña, M. G., Ocmin, L. S. L., & Romero-Carazas, R. (2023). Control interno de inventario y la gestión de resultados de un emporio comercial de la región de San Martín - Perú. *Región Científica*, *2*(2), 202392–202392. https://doi.org/10.58763/RC202392

Pitombeira-Neto, A. R., & Murta, A. H. F. (2022). A reinforcement learning approach to the stochastic cutting stock problem. *EURO Journal on Computational Optimization*, *10*, 100027. https://doi.org/10.1016/J.EJCO.2022.100027

Plan Nacional de Competitividad y Productividad. (n.d.). In *Gob.pe*.

Romero, A., González, A., Díaz, M., & Rueda, N. (2023). Revisión y perspectivas para la construcción de bases de datos robustas con datos faltantes: Caso aplicado a información financiera. *Tecnura*, *27*(75), 12–37. https://doi.org/10.14483/22487638.18268

Sattar, M. U., Dattana, V., Hasan, R., Mahmood, S., Khan, H. W., & Hussain, S. (2025). Enhancing Supply Chain Management: A Comparative Study of Machine Learning Techniques with Cost–Accuracy and ESG-Based Evaluation for Forecasting and Risk Mitigation. *Sustainability 2025, Vol. 17, Page 5772*, *17*(13), 5772. https://doi.org/10.3390/SU17135772

Seyedan, M., Mafakheri, F., & Wang, C. (2023). Order-up-to-level inventory optimization model using time-series demand forecasting with ensemble deep learning. *Supply Chain Analytics*, *3*, 100024. https://doi.org/10.1016/J.SCA.2023.100024

Taquiri, F., Anthony, G., Celis, M., Stiven, J., Navarro, M. Q., & Michael, D. (2024). Machine learning para la predicción en la gestión de inventario dirigida a PYMES de venta de productos tecnológicos. *Repositorio Institucional - UCV*. https://repositorio.ucv.edu.pe/handle/20.500.12692/151376

Theodorou, E., Spiliotis, E., & Assimakopoulos, V. (2023). Optimizing inventory control through a data-driven and model-independent framework. *EURO Journal on Transportation and Logistics*, *12*, 100103. https://doi.org/10.1016/J.EJTL.2022.100103

Thuy, T., & Nguyen, H. (2023). Applications of Artificial Intelligence for Demand Forecasting. *OPERATIONS AND SUPPLY CHAIN MANAGEMENT*, *16*(4), 424–434.

Van Tran, L., T Dao, S. V, M Ho, T. T., Tran, L. V, Tran, H. M., & Dao, S. V. (2022). Machine Learning in Demand Forecasting. In *International Research Journal of Advanced Engineering and Science* (Vol. 7, Issue 3). https://www.researchgate.net/publication/370155418

Vera, S., Daniel, J., Cueva, A., & Doi, J. C. (2025). Modelo para la predicción de compra de materia prima en la gestión de inventario para producción en Pymes del sector retail aplicando machine learning. *Universidad Peruana de Ciencias Aplicadas (UPC)*. https://doi.org/10.19083/TESIS/684062

Villalobos Collantes, L. G. (2022). *La Gestión de Inventarios y su Relación en la Rentabilidad de las Empresas de Tarapoto*. Universidad Peruana Unión. http://repositorio.upeu.edu.pe/handle/20.500.12840/5340

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

Yang, Y., Wang, M., Wang, J., Li, P., & Zhou, M. (2025). Multi-Agent Deep Reinforcement Learning for Integrated Demand Forecasting and Inventory Optimization in Sensor-Enabled Retail Supply Chains. *Sensors 2025, Vol. 25, Page 2428*, *25*(8), 2428. https://doi.org/10.3390/S25082428

Yani, L. P. E., & Aamer, A. (2023). Demand forecasting accuracy in the pharmaceutical supply chain: a machine learning approach. *International Journal of Pharmaceutical and Healthcare Marketing*, *17*(1), 1–23. https://doi.org/10.1108/IJPHM-05-2021-0056