



PRONÓSTICOS DE PLANES DE DEMANDA

Autores:

Juan David Borja Marulanda
Mitchel Esteban Collazos Agredo

Trabajo de grado II

Asesor metodológico:

Ph.D. Yesid Ospitia Medina

Tutor:

Ms. Diego Fernando Agudelo Burbano

Universidad ICESI

Maestría en Ciencia de Datos
Cali, Valle del Cauca, Colombia
2025

TABLA DE CONTENIDO

| | |
|--------------------------------------|-----------|
| 1. Justificación del Problema | 3 |
| 2. Objetivo general | 5 |
| 3. Objetivos específicos | 5 |
| 4. Metodología | 6 |
| 5. Marco teórico | 10 |
| 6. Estado del arte | 14 |
| 7. Resultados | 21 |
| 8. Conclusiones | 47 |
| 9. Limitaciones | 48 |
| 10. Consideraciones Futuras | 49 |
| 11. Referencias | 50 |

1. Justificación del Problema

En los últimos años, la planificación de la demanda de productos de consumo masivo e industriales en la empresa se ha llevado a cabo mediante un enfoque colaborativo. Aunque este método tiene ciertos beneficios tales como un compromiso y alineación entre las áreas relacionadas (producción, marketing y ventas), ha mostrado limitaciones significativas debido a la inexactitud en las proyecciones y la aparición de demandas inesperadas y negociaciones especiales. En cada categoría los refuerzos de los clientes frente a una demanda oficial establecida durante el mismo mes genera un impacto en toda la cadena de suministro. Estos problemas han impactado principalmente en los siguientes puntos, (ver *Figura 1. Árbol de problemas*):

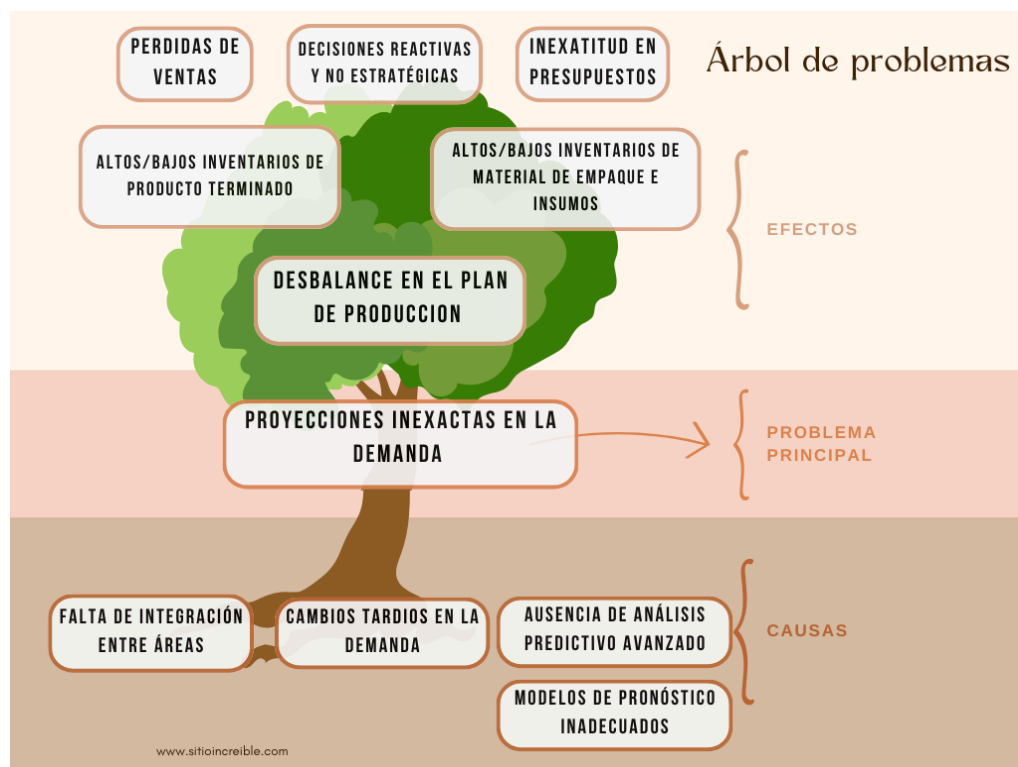


Figura.1. Árbol de problemas

En actividades comerciales regulares, una dificultad constante es igualar las acciones planificadas con la situación comercial real. La planificación de la producción, basada en estimaciones de ventas futuras, a menudo no refleja con precisión los requisitos reales del mercado. Los elementos como la demanda impredecible, los cambios inesperados de negociaciones especiales o pronósticos imprecisos han resultado en que las instalaciones de producción no se utilicen de la manera más eficiente. Esto conlleva a gastos y espacios adicionales ; Y en otros casos, la producción es demasiado baja, impactando directamente la capacidad de la compañía para satisfacer las necesidades de sus clientes.

Esta discrepancia también afecta directamente al control de acciones. Para prevenir la escasez de existencias, la empresa ha implementado una regla de mantener medidas de seguridad, pero esta acción, aunque es de precaución, no está libre de riesgos. Demasiado stock, particularmente de bienes completos, aumenta los gastos de transporte y almacenamiento, además de que trae el peligro de expirar, lo que significa un claro golpe monetario. Por el contrario, mantener un inventario demasiado bajo pone en riesgo la disponibilidad de productos, especialmente en el mercado nacional, reduciendo la capacidad de cumplir con la demanda de forma oportuna.

Además, esta imprecisión en la programación no solo influye en el resultado final, sino que también afecta en gran medida la logística inversa. Cuando la planificación no tiene en cuenta los tiempos y cantidades necesarios, el suministro de materiales de insumos y de material de empaque se ve afectado. Esto puede causar demoras en la fabricación, interrupciones en el flujo de trabajo y, en el peor de los casos, fallas en los envíos. Por lo tanto, la empresa encuentra una situación de imprevisibilidad continua que afecta su rendimiento laboral y su posición con los clientes.

2. Objetivo general

Desarrollar un modelo integrado de predicción de demanda que permita predecir el mes siguiente de los 2 productos de mayor venta, optimizando la precisión en la planificación y mejorando la gestión de inventarios y recursos en la empresa.

3. Objetivos específicos

- Identificar y priorizar los 2 productos de mayor facturación utilizando análisis exploratorio de la serie de tiempo basados en datos semanales.
- Establecer un protocolo de evaluación mediante ventanas móviles utilizando la métrica de error RMSE, con el fin de comparar el desempeño de los modelos aplicados sobre datos originales y datos limpios en diferentes horizontes de predicción y niveles de agregación temporal.
- Evaluar modelos de promedio móvil, suavización exponencial, ARIMA y vectores autorregresivos (VAR) en múltiples series de tiempo de acuerdo a las variables exógenas, para una serie de tiempo con datos originales y limpios.
- Realizar un análisis comparativo entre la proyección obtenida del mejor modelo mensual y semanal vs la demanda colaborativa y las ventas reales para determinar cuál ofrece la mayor precisión y confiabilidad en la proyección de la demanda.

4. Metodología

4.1. Proyecto de dos fases

Este proyecto utiliza como base los resultados de PRONÓSTICOS DE PLANES DE DEMANDA realizado en Trabajo de Grado I. En aquella ocasión, los objetivos específicos fueron los siguientes:

- Identificar y priorizar el producto de mayor facturación utilizando análisis exploratorio de la serie de tiempo.
- Evaluar modelos de promedio móvil, suavización exponencial y ARIMA para una serie de tiempo con datos sin imputar y con imputación.
- Realizar un análisis comparativo entre la proyección obtenida del mejor modelo vs la demanda colaborativa y las ventas reales para determinar cuál ofrece la mayor precisión y confiabilidad en la proyección de la demanda.

Los resultados de este trabajo son llamados fase 1 y los resultados del presente trabajo son llamados fase 2.

Se decidió profundizar para esta segunda fase en el uso de modelos no tradicionales de entrenamiento multivariado como vectores autorregresivos (VAR). De igual forma, se consiguió ampliar el número de filas del dataset al utilizar semanas en lugar de meses, y se toman en cuenta variables adicionales como si el producto tuvo promoción o productos bonificados en esas semanas (“Promoción (SI/NO)” y “Bonificado” en el dataset, respectivamente). Con esto, se espera mejorar los resultados obtenidos a fin de mejorar los pronósticos de planificación, gestión de inventarios y de recursos en la empresa.

Finalmente, se considera que para ambas fases, la metodología de CRISP-DM. En el marco teórico y el estado del arte se incluye lo relacionado con vectores autorregresivos (VAR) .

4.2. CRISP-DM

La metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) es una de las metodologías más utilizadas en proyectos de análisis de datos y minería de datos, debido a su flexibilidad y enfoque estructurado. Adaptar CRISP-DM a un proyecto de series de tiempo, especialmente para el pronóstico de demanda y producción, permite llevar a cabo un

análisis sistemático y orientado a resolver problemas específicos del negocio [1] (ver Figura 2 Esquema del ciclo CRISP-DM estándar).

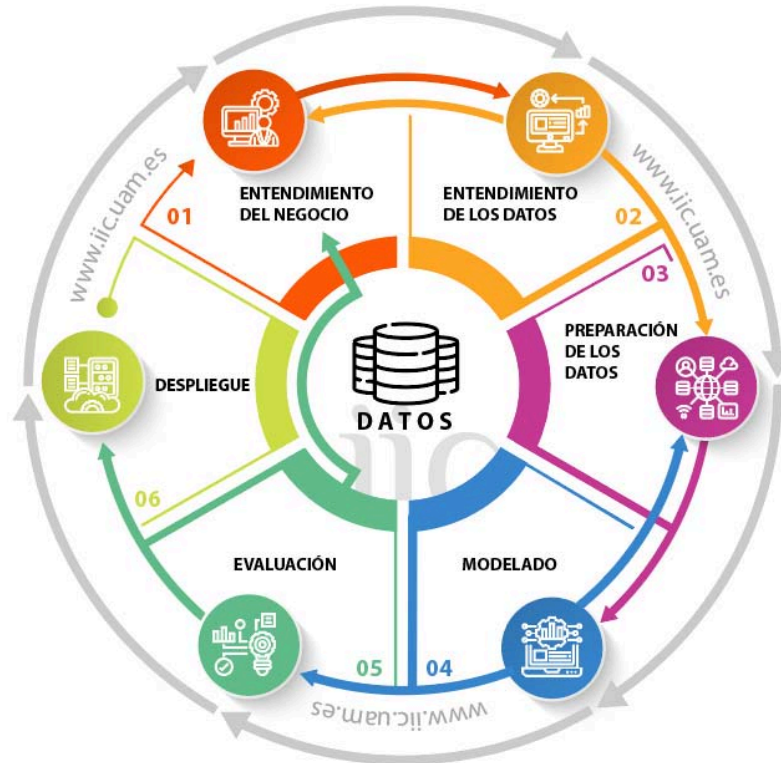


Figura 2. Esquema del ciclo CRISP-DM estándar[2].

Como se ilustra en la **Figura 2**, la metodología se presenta como un proceso secuencial, es fundamental destacar su naturaleza iterativa. Las fases 1 y 2 pueden repetirse si, a partir de los resultados obtenidos en los análisis exploratorios (fase 2), surgen nuevos elementos que requieren redefinir los objetivos de negocio (fase 1). De igual forma, la fase de modelado (fase 4) puede provocar la necesidad de realizar nuevos pre-procesamientos de los datos (fase 3) para mejorar los análisis previos. Además, los resultados de la fase de evaluación (fase 5) pueden generar nuevas demandas en los objetivos de negocio (fase 1).

1. Entendimiento del Negocio:

- **Reunión con stakeholders:** definir los objetivos del proyecto y las expectativas.
- **KPI:** Identificación de indicadores clave de rendimiento (KPI) relevantes, como accuracy, plan demanda oficial vs el pronosticado. Plan de producción vs presupuesto.
- **Definición del alcance del proyecto:** ¿Se centrará solo en pronósticos de demanda? ¿debería incluir la producción? ¿Cuál es el horizonte temporal?

2. Entendimiento de los Datos:

Recolectar, explorar y analizar los datos históricos para obtener un entendimiento inicial de los patrones y estructuras presentes.

- **Realización del EDA (Análisis exploratorio de Datos).**
- **Datos:** Que clases de datos se requieren recolectar (Históricos de producción y facturación).

3. Preparación de los Datos

Limpiar, transformar y preparar los datos para el modelado. Este es uno de los pasos más críticos en un proyecto de series de tiempo.

- **Limpieza de Datos:** Tratar valores atípicos, imputar o eliminar valores faltantes y corregir errores en los registros.
- **Transformación de Series de Tiempo:** Hacer transformaciones como diferenciación para eliminar tendencias y hacer la serie estacionaria, escalado de datos, y ajustes de estacionalidad si es necesario.
- **Feature Engineering:** Creación de nuevas variables como indicadores de estacionalidad, efectos de calendario (fines de semana, festivos), y variables de eventos especiales (promociones, lanzamientos).

4. Modelado

Seleccionar y aplicar modelos de pronóstico que sean adecuados para los patrones observados en los datos.

- **Selección de Modelos:** Considerar varios modelos, como promedio móvil, ARIMA, modelos de suavización exponencial (Holt-Winters) u otros.
- **Protocolos de Evaluación:** Separar los datos en conjuntos de entrenamiento y prueba para evaluar la capacidad predictiva de los modelos.
- **Optimización de Parámetros:** Ajustar los parámetros del modelo para mejorar su rendimiento. Esto incluye la selección de parámetros utilizando optimización bayesiana.

5. Evaluación

Validar los modelos seleccionados y asegurarse de que cumplen con los requisitos del negocio:

- **Evaluación de Desempeño:** Utilizar métricas como RMSE (Root Mean Square Error), MAE (Mean Absolute Error), o MAPE (Mean Absolute Percentage Error) para medir la precisión.
- **Pruebas de Robustez:** Evaluar la estabilidad del modelo utilizando validación cruzada o pruebas en diferentes ventanas de tiempo.

- **Análisis de Resultados:** Comparar los resultados del modelo con las expectativas del negocio y ajustar si es necesario.

6. Despliegue

Implementar un entregable como un excel con la predicción obtenida, que permita la rápida toma de decisiones y monitorear su desempeño en la práctica.

- **Implementación:** Integrar el modelo en sistemas de planificación de la empresa como Excel y/o SAP. Esto puede incluir la automatización de pronósticos y la generación de informes periódicos.
- **Monitoreo Continuo:** Establecer un sistema de monitoreo para ajustar el modelo conforme cambien las condiciones del negocio (nuevos datos, cambios estacionales, etc.).
- **Documentación y Comunicación:** Documentar el proceso, las decisiones tomadas, los resultados obtenidos y cualquier recomendación futura.

Con la adaptación de esta metodología se busca no solo pronosticar sino también integrar las predicciones en procesos críticos del negocio (POV y Producción), asegurando un impacto directo en la precisión del plan de demanda y ajuste en el plan de producción de acuerdo a las variables más significativas garantizando una alineación entre los dos procesos.

5. Marco teórico

El presente estudio se enfoca en la evaluación de modelos de pronóstico para mejorar la precisión en la planificación de demanda en una empresa manufacturera. La precisión en los pronósticos de demanda es crucial para mayor exactitud en la elaboración de planes maestros de producción (MPS), reducir sobrecostos de inventario y evitar el desabastecimiento, dos problemas comunes en la gestión de la cadena de suministro.

“**La proyección de la demanda** es el proceso de estimar la demanda futura de un producto o servicio. Es esencial comprender la demanda pasada y presente, así como los factores que pueden influir en la demanda futura, como las tendencias del mercado, el comportamiento del consumidor y los cambios en la economía” [3].

Una proyección de demanda exacta nos ayuda a prevenir el desabastecimiento y a reducir los gastos vinculados con la sobreproducción. Además, facilita una planificación más efectiva de la mano con la producción, asegurando que los productos estén disponibles para su distribución y cumplan con las políticas de almacenamiento establecidas con la compañía. Esto impacta directamente todo el proceso de **Supply Chain**, (ver Figura 3. Cadena de suministros):



Figura 3. Cadena de suministros [4]

Supply Chain (Cadena de Suministro) (Figura 3):

Se compone de todas las partes involucradas, directa o indirectamente, para satisfacer la petición de un cliente. El proceso se inicia con una demanda que se traduce en una necesidad de compra de materiales de empaques e insumos para la programación de un plan maestro de producción (MPS).

Demanda:

La demanda se refiere a la cantidad de bienes o servicios que los consumidores están dispuestos a comprar en un mercado particular durante un periodo de tiempo definido. En términos económicos, la demanda representa el número de unidades de un producto o servicio que los compradores estarían dispuestos a adquirir a distintos niveles de precios. Según la ley de la demanda, cuando el precio de un bien baja, la cantidad demandada tiende a aumentar.

MPS:

Es una herramienta fundamental para iniciar el proceso productivo en una empresa. Es un documento que detalla qué se va a producir, en qué cantidad y en qué plazo. Define los pasos necesarios para fabricar un producto y los recursos requeridos, proporcionando un nivel de detalle crucial para una toma de decisiones informada y efectiva [5].

Ventajas:

- Optimizar la producción, ya que mejora la capacidad productiva a corto y medio plazo y permite gestionar grandes cantidades de pedidos sin problemas. También puedes priorizar los pedidos de los clientes más importantes. Además, ganas en eficiencia al poder alinear tiempos, empleados y materiales, y esto repercute también en una mejor rentabilidad.
- Prevenir o reducir las pérdidas: permite planificar todo bien y controlar cada etapa. Aplicando el MPS puedes abordar cualquier imprevisto con mejores resultados.
- Toma de decisiones: gracias a los datos seguros y precisos que se reflejan en el documento, que ofrecen una visión inmejorable de los procesos.

Producto Terminado:

El producto terminado es el resultado final del proceso de producción y está listo para ser distribuido o vendido. La planificación eficiente de la producción y gestión de inventarios asegura que los productos terminados estén disponibles en el momento adecuado, evitando tanto el exceso como la escasez de stock.

Material de Empaque:

El material de empaque incluye todos los componentes necesarios para contener, proteger y presentar los productos. Su gestión es vital, ya que afecta no solo la presentación y protección del producto, sino también los costos logísticos y la percepción del cliente.

Insumos:

Los insumos, como materias primas e ingredientes, son fundamentales en el proceso de producción. La calidad y disponibilidad de estos insumos impactan directamente en la eficiencia del proceso productivo y la calidad del producto final, siendo un elemento clave en la planificación de la producción.

Gestión de Inventarios:

La gestión de inventarios requiere una visibilidad clara de las existencias, saber cuándo reabastecer, cuánto pedir y dónde almacenar los productos. Las operaciones de cumplimiento de pedidos multicanal suelen distribuir el inventario en múltiples ubicaciones a lo largo de la cadena de suministro. Las empresas necesitan una visión precisa del inventario para asegurar el cumplimiento de los pedidos de los clientes, reducir los tiempos de envío y minimizar los desabastecimientos, sobreventas y descuentos [6].

Obtener el control de los inventarios garantizando el cumplimiento de una venta se debe garantizar desde las variaciones de la producción y de la demanda sean mínimas, por ello al realizar una comparación entre ambos históricos (Ventas y Producción) se obtendrá un pronóstico más acertado.

Series de Tiempo:

Una serie de tiempo es una secuencia de observaciones de una variable registrada en intervalos de tiempo equidistantes. Las series de tiempo se utilizan para analizar la evolución temporal de los datos y realizar predicciones futuras basadas en patrones pasados [7].

Modelos de Series de Tiempo: Los modelos más comunes para el análisis de series de tiempo incluyen los modelos ARIMA (Autoregressive Integrated Moving Average), SARIMA (Seasonal ARIMA), y modelos de suavizamiento exponencial como Holt-Winters. Estos modelos permiten capturar tanto la componente autorregresiva como las tendencias y estacionalidades presentes en los datos [8].

Descomposición de Series de Tiempo: La descomposición de series de tiempo es una técnica que permite separar una serie en sus componentes fundamentales: tendencia, estacionalidad y componente aleatorio. Esto facilita la comprensión de los patrones subyacentes y mejora la precisión de los modelos de pronóstico [9].

Modelos de suavizamiento exponencial:

Los modelos de suavizamiento exponencial, como el método de Holt-Winters, se utilizan para capturar patrones de tendencia y estacionalidad. Estos modelos asignan más peso a los

datos recientes, haciendo que las predicciones sean más sensibles a los cambios recientes en la serie [10].

Modelos ARIMA y SARIMA:

Los modelos ARIMA (Autoregressive Integrated Moving Average) son utilizados para modelar series de tiempo estacionarias mediante la combinación de componentes autoregresivos, diferencias para la integración y medias móviles. SARIMA extiende este concepto para capturar patrones estacionales en las series [11].

Medición de la Precisión de los Modelos:

Para evaluar la precisión de los modelos de series de tiempo, se utilizan métricas como el Error Absoluto Medio (MAE), el Error Cuadrático Medio (MSE) y el Error Absoluto Porcentual Medio (MAPE). Estas métricas ayudan a comparar el rendimiento de diferentes modelos y seleccionar el más adecuado [12].

Vectores Autorregresivos (VAR):

Los modelos autorregresivos conocidos como AR son aquellos en los que una variable o conjunto de variables se explican en función de los valores pasados de esa misma variable o conjunto de variables. Así, los modelos de vectores autorregresivos (VAR) hacen referencia a la generalización de los AR de un vector con n variables “ y ” a través del tiempo (t) [13].

Protocolo de evaluación de ventana recursiva:

Se simula el contexto de uso, donde a medida que se predice, el modelo que se va a evaluar se entrena con nueva información (ver Figura 4. Evaluación de pronósticos en series de tiempo, usando el protocolo ventana recursiva.). Es un método que resulta muy interesante cuando se pretende examinar cómo un modelo es capaz de adaptarse a la nueva información en un entorno de predicción continuo [14].

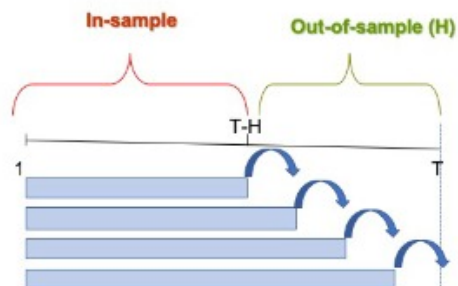


Figura 4. Evaluación de pronósticos en series de tiempo, usando el protocolo ventana recursiva.

6. Estado del arte

A continuación se indican los textos seleccionados que son relevantes para el desarrollo de este trabajo de grado:

1. **Fernández Muñoz, C. A., Rodríguez Morachis, M. A., Terrazas Mata, L. E., Zorrilla Briones, F. (2022). *Determinación De Demandas De Clientes Anticipadas Con La Utilización De Modelos De Pronósticos Para La Planeación De La Producción De Producto Médico. Revista Ipsumtec, 5(5), 82–89. Recuperado a partir de <https://revistas.milpaalta.tecnm.mx/index.php/IPSUMTEC/article/view/154>.*** El enfoque es referente a mejorar la planeación de producción en una planta de manufactura ubicada en Ciudad Juárez, Chihuahua, México, especializada en la fabricación de productos electrónicos. La empresa enfrenta problemas significativos en la entrega de órdenes a tiempo, con menos del 50% de cumplimiento, debido a una planeación de producción ineficiente y no sistematizada, basada principalmente en la experiencia del planeador de producción y sin una fuente confiable de requerimientos de los clientes. A través de modelos de pronósticos (Holt, Holt-Winters y promedio móvil) desean encontrar el mejor modelo para un producto (PDM) basado en datos históricos de venta. Con este enfoque resalta como una mayor precisión de los modelos tendría un mayor tiempo de respuestas a nuestros clientes.
2. **Cancino, S. E., Cancino-Escalante, G. E. Cancino-Ricketts, D. F. (2023). *Modeling and forecasting blackberry production in Colombia using a Box Jenkins ARIMA approach. Económicas CUC, 44(1), 69–82. DOI: <https://doi.org/10.17981/econcuc.44.1.2023.Econ.4>***. El artículo resalta la importancia económica de la producción de mora de Castilla en Colombia, destacando su contribución al PIB, al empleo y al bienestar social de los agricultores. La inestabilidad en la producción afecta la rentabilidad económica, por lo que pronosticar la producción es crucial para la asignación de recursos y la toma de decisiones. Basándose en el modelo Arima buscan pronosticar la producción de los años 2021 y 2023 teniendo en cuenta los históricos del 1999 - 2020. El artículo muestra como el no tener una demanda estable genera una variación en producción, pero teniendo un buen modelo se puede ajustar la programación en planta y disponer de manera óptima los insumos y material de empaque.

3. **Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive analytics for demand forecasting—a comparison of SARIMA and LSTM in retail SCM. *Procedia Computer Science*, 200, 993-1003. <https://www.sciencedirect.com/science/article/pii/S1877050922003076>.** Este artículo busca pronosticar la demanda de frutas y verduras en el comercio minorista de comestibles. Para ello, utiliza datos de cuatro productos perecederos (tomate, papa, ensalada, pepino) vendidos en más de 90 tiendas de un minorista austríaco. Explica el proceso de preparación y limpieza de datos utilizando promedio aritmético simple. Posteriormente obtiene pronósticos para enero de 2020 utilizando los modelos SARIMA, LSTM y SARIMAX, comparándolos con los datos de ventas reales. Concluye que LSTM funcionó mejor para productos con demanda estable (papa), mientras que SARIMA mostró mejores resultados para productos con comportamiento estacional (pepino). Si se incorporan días de descuento como un factor externo en SARIMAX, existen mejoras significativas en la precisión de los pronósticos, particularmente para productos con fuertes ventas impulsadas por descuentos (ensalada, tomate).
4. **Sohrabpour, V., Oghazi, P., Toorajipour, R., & Nazarpour, A. (2021). Export sales forecasting using artificial intelligence. *Technological Forecasting and Social Change*, 163, 120480. <https://www.sciencedirect.com/science/article/pii/S0040162520313068>.** Presenta un marco para modelar y pronosticar las ventas de exportación utilizando algoritmos genéticos (GP). El marco desarrollado abarca cuatro pasos: recopilación de datos, preprocesamiento de datos, modelado y producción y uso. Adicionalmente, se explica la lógica para elegir variables según la disponibilidad de datos y el respaldo de la literatura. Se presentan las métricas de error utilizadas para evaluar la calidad del modelo de pronóstico causal: error absoluto medio (MAE), error cuadrático medio (MSE), bondad de ajuste R-cuadrado (R^2) y coeficiente de correlación. Se presentan los resultados de un modelo de ventas de exportación basado en programación genética y analiza el resultado de las métricas de error para un período de seis semanas.
5. **Van Steenbergen, R. M., & Mes, M. R. (2020). Forecasting demand profiles of new products. *Decision support systems*, 139, 113401. <https://www.sciencedirect.com/science/article/pii/S0167923620301561>.** Este artículo examina la previsión de la demanda de nuevos productos, que a diferencia de productos ya establecidos, no disponen de datos históricos de ventas como indicador de las ventas futuras. Se presenta un nuevo método de previsión de la demanda denominado DemandForest, que combina K-means, Random Forest y

Quantile Regression Forest orientado a cuantificar la incertidumbre de la demanda. DemandForest agrupa y predice patrones de demanda, y predice los cuantiles de la demanda total durante un período de introducción.

6. ***Pinedo Chapa, Joely Mireilli.(2018).Propuesta de un modelo de pronósticos de demanda y gestión de inventarios para la planeación de demanda en prendas de vestir juvenil.Universidad Peruana de Ciencias Aplicadas (UPC) .<https://repositorioacademico.upc.edu.pe/handle/10757/623528>.*** La tesis identifica oportunidades económicas y estratégicas en la gestión de la planeación que actualmente no se aplican en empresas de comercialización de prendas de vestir. En Gestión de Inventarios, se propone la clasificación ABC, que revela que el 20% de las existencias concentra el 80% del valor del inventario, clasificando los productos según impacto en valor, inventario y ventas. sugiere implementar modelos de pronóstico de demanda para mejorar la gestión de inventarios y recursos. La tesis enlaza conceptos principales de la cadena de suministro abarcando desde la demanda, producción, compras y gestión de inventarios agregando un concepto de clasificación ABC el cual define los SKU'S principales que permitiría al planeador identificar prioridades de producción.

7. ***Feizabadi, J. (2022). Machine learning demand forecasting and supply chain performance. International Journal of Logistics Research and Applications, 25(2), 119-142.. <https://www.tandfonline.com/doi/abs/10.1080/13675567.2020.1803246>.*** En este artículo, se desarrolla un método híbrido de pronóstico de la demanda basados en ARIMAX y redes neuronales. Esto permitió que tanto las series temporales como los factores explicativos se incorporarán al método desarrollado. Se evaluó el método en el contexto de un fabricante de acero. El autor indica que la técnica ARIMAX es mejor para predecir los picos de demanda, mientras que las redes neuronales generan predicciones más “suavizadas” con mayor precisión.

8. ***Granja Brito, P. A. (2018). Diseño De Un Plan Maestro De Producción De La Familia De Productos René Chardón Para La Empresa Distribuciones Y Logística J&J, Ubicada En El Cantón De Ambato, Provincia De Tungurahua. <https://repositorio.uti.edu.ec/handle/123456789/628>.*** (Bachelor's thesis, Ambato: Universidad Tecnológica Indoamérica). Destaca la importancia del Plan Maestro de Producción (MPS) en la optimización de procesos en un entorno competitivo. La creación del MPS se complica con el aumento de la producción y recursos limitados, enfrentando objetivos como maximizar el servicio, usar eficientemente los recursos y minimizar inventarios. Las metaheurísticas, especialmente los Algoritmos Genéticos, superan las limitaciones de las heurísticas simples y ofrecen mejores resultados en la

optimización del MPS. En un caso de estudio, su aplicación mejoró significativamente la planificación de la producción. El trabajo aporta conocimiento de los inputs necesarios para el desarrollo de un MPS, a su vez amplía el conocimiento sobre el significado y cómo se elabora un MPS.

9. Huber, J., Gossmann, A., & Stuckenschmidt, H. (2017). *Cluster-based hierarchical demand forecasting for perishable goods. Expert systems with applications*, 76, 140-151.

<https://www.sciencedirect.com/science/article/pii/S0957417417300313>. Presenta un enfoque para identificar grupos de artículos en función del patrón de ventas durante el día. Se aplicaron modelos ARIMA multivariados para pronosticar la demanda diaria con el fin de respaldar las decisiones operativas de una cadena de panadería industrializada. La ventaja de pronosticar sobre grupos de artículos en lugar de artículos individuales es que los grupos suelen tener un patrón de ventas durante el día similar, lo que hace que sea razonable pronosticar la demanda a nivel agregado. Esto además, reduce los costos computacionales de los pronósticos.

10. Camacho, H., Gómez, K., & Monroy, C. A. (2012, July). *Importancia de la cadena de suministros en las organizaciones. In Tenth Laccei Lat. Am. Caribb. Conf.* <https://www.laccei.org/LACCEI2012-Panama/RefereedPapers/RP200.pdf>

El artículo destaca la importancia de que las empresas comprendan y apliquen el concepto de cadena de suministros para ser más competitivas, mejorar procesos y aumentar la satisfacción del cliente. Subraya que las empresas no deben verse como entidades aisladas, sino como parte de un entorno más amplio. Además, resalta la necesidad de que las áreas dentro de una empresa sean interdependientes para lograr éxito en los procesos y el producto final. Este artículo explica la importancia de la comunicación entre áreas y como todo funciona de manera integrada en una compañía el cual soporta la relación que deben tener las áreas (demanda y producción).

11. Rahmer, B., Garzón, H., & Garzón, J. (2019). *Estimación de un modelo de vectores autorregresivos para pronóstico de demanda. Un estudio empírico.* https://www.academia.edu/download/60022001/Rahmer_et_al_WP20190716-70684-17gpldp.pdf

El artículo trata sobre el diseño, construcción y validación de un modelo de Vectores Autorregresivos (VAR) para la estimación cuantitativa de las ventas en una empresa manufacturera del clúster petroquímico de Cartagena de Indias. Se justifica el uso de este modelo porque permite capturar la dinámica de múltiples variables y establecer relaciones de causalidad en el sentido de Granger. Además, se comparan sus predicciones con otros modelos tradicionales de series

temporales para evaluar su eficiencia predictiva. Los resultados muestran que variables como la cuota de mercado, las estrategias de marketing, la inversión en capital de los consumidores y el precio unitario influyen significativamente en la estimación de las ventas mensuales. Finalmente, se propone realizar estudios comparativos con otras técnicas de pronóstico y aplicar estos enfoques en el sector microempresarial.

A continuación se presenta una tabla comparativa de los artículos anteriormente mencionados, (ver *Tabla I Comparación de artículos científicos con la propuesta de trabajo de grado*):

Tabla I. Comparación de artículos científicos con la propuesta de trabajo de grado.

| # | Nombre | Autores | Fecha de publicación | País | Idioma | Algoritmo de predicción | ¿Realiza análisis exploratorio? | ¿Pronostica serie de tiempo multivariada? | ¿Evalúa modelo con métricas de evaluación? |
|---|--|--|----------------------|----------------------------|---------|---|---------------------------------|---|--|
| 1 | Determinación De Demandas De Clientes Anticipadas Con La Utilización De Modelos De Pronósticos Para La Planeación De La Producción De Producto Médico. | Fernández Muñoz, C. A., Rodríguez Morachis, M. A., Terrazas Mata, L. E., & Zorrilla Briones, F | 2022 | México | Español | Suavizamiento exponencial | No | No | No |
| 2 | Modeling and forecasting blackberry production in Colombia using a Box Jenkins ARIMA approach. | Cancino, S. E., Cancino-Escalante, G. E. Cancino-Ricketts, D. F | 2023 | Colombia | Inglés | ARIMA, VAR | No | No | Sí |
| 3 | Predictive analytics for demand forecasting – A comparison of SARIMA and LSTM in retail SCM. | Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. | 2022 | Austria | Inglés | SARIMA, SARIMAX, LSTM | Sí | Si | Sí |
| 4 | Export sales forecasting using artificial intelligence. | Sohrabpour, V., Oghazi, P., Toorajipour, R., & Nazarpour, A. | 2021 | Dinamarca, Suecia, Irlanda | Inglés | Algoritmos genéticos | Sí | No | No |
| 5 | Forecasting demand profiles of new products. | Van Steenberghe, R. M., & Mes, M. R. | 2020 | Países bajos | Inglés | Demand Forest, Quantile Regression Forest, Promedio móvil | No | Si | Sí |

| | | | | | | | | | |
|----|--|---|-------------|-----------------|----------------|---------------------------------------|-----------|-----------|-----------|
| 6 | Propuesta de un modelo de pronósticos de demanda y gestión de inventarios para la planeación de demanda en prendas de vestir juvenil. | Pinedo Chapa, Joely Mireilli. | 2018 | Perú | Español | Suavizamiento exponencial | No | No | No |
| 7 | Machine learning demand forecasting and supply chain performance. | Feizabadi, J. | 2022 | Malasia | Inglés | ARIMAX, Redes Neuronales | Sí | Sí | Sí |
| 8 | Diseño De Un Plan Maestro De Producción De La Familia De Productos René Chardón Para La Empresa Distribuciones Y Logística J&J, Ubicada En El Cantón De Ambato, Provincia De Tungurahua. | Granja Brito, P. A. | 2018 | Ecuador | Español | Promedio móvil | No | No | No |
| 9 | Cluster-based hierarchical demand forecasting for perishable goods. | Huber, J., Gossmann, A., & Stuckenschmidt, H. | 2017 | Alemania | Inglés | ARIMA + Clusterización | No | Sí | No |
| 10 | Importancia de la cadena de suministros en las organizaciones. | Camacho, H., Gómez, K., Monroy, C. A | 2012 | Colombia | Español | No aplica | No | No | No |
| 11 | Estimación de un modelo de vectores autorregresivos para pronóstico de demanda. Un estudio empírico. | Rahmer, B., Garzón, H., & Garzón, J. | 2019 | Colombia | Español | Vectores Autorregresivos (VAR) | Sí | Sí | Sí |
| 12 | Propuesta de trabajo propia | Borja J., Collazos M. | 2025 | Colombia | Español | Vectores Autorregresivos (VAR) | Sí | Sí | Sí |

En general, los artículos generan pronósticos de series de tiempo en los cuales menos de la mitad constan de análisis exploratorio o evaluación de distintos modelos. Aquellos artículos que trataron el análisis exploratorio tuvieron como objetivo:

- Encontrar relaciones entre la variable objetivo y variables predictoras
- Corregir valores anómalos y nulos

Lo cual son aspectos que tienen en común con la propuesta de este trabajo.

En el entrenamiento de modelos, todos realizan un protocolo de evaluación holdout (dividir en set de entrenamiento y de prueba 75-25%). Mientras que en esta propuesta, se utilizó un protocolo de evaluación de ventana recursiva.

Otro aspecto a mencionar es que los artículos en español aplicados a casos específicos de demanda en producción tienen un enfoque a métodos de pronóstico tradicionales (suavización exponencial y promedio móvil) en oposición a los artículos en inglés los cuales exploran métodos más sofisticados (principalmente variaciones de ARIMA o redes neuronales). Esta propuesta se enfoca en los métodos de pronóstico tradicionales con los cuales se apunta a mejorar el pronóstico por consenso realizado en la compañía.

En definitiva, el enfoque de este trabajo de grado tiene los siguientes diferenciadores:

- Entrenamiento con series multivariadas
- Entrenamiento con protocolo de evaluación de ventana recursiva

7. Resultados

7.1. Conclusiones de la primera fase

Mediante el desarrollo del análisis exploratorio de la serie de tiempo, en el análisis exploratorio se identificó que el producto con la mayor facturación es el 1102 HARINA DE TRIGO HAZ DE OROS X 50 KILOS. Este producto fue priorizado en la planeación de la demanda.

Se determinó que entre los tres modelos utilizados de promedio móvil, suavización exponencial y ARIMA, fue el modelo de suavización exponencial ejecutado bajo optimización bayesiana quien demostró un mejor desempeño en términos del menor error RMSE. Adicionalmente, el modelo tuvo mejores resultados con datos sin imputación que con datos imputados por la mediana.

El mejor modelo obtenido en este proyecto logró tener un comportamiento más ajustado a la venta real en comparación al POV planteado mensual por consenso (pronóstico por consenso) en un 50.8% durante los últimos seis meses usados de validación. Con estos resultados, finalmente se obtuvo un pronóstico para el mes de diciembre de 2024 de **9,547,326.20 kilos (9,547 Ton)**.

7.2 Segunda Fase.

En esta fase, se incorpora el modelo de Vectores Autorregresivos (VAR) con el propósito de evaluar el impacto de variables exógenas en la precisión del pronóstico. Para facilitar este análisis y adaptarlo al comportamiento de las variables involucradas, se realiza un cambio en la granularidad temporal del dataset, pasando de una frecuencia mensual a una semanal.

7.2.1 Análisis exploratorio

7.2.1.1 Consolidación del dataset

Para el desarrollo del proyecto, se cuenta con un dataset en formato Excel elaborado por el equipo de planeación, el cual contempla las proyecciones desde la semana 52 del año 2021 hasta la semana 01 de enero de 2025. Adicionalmente, se integraron las variables exógenas correspondientes para el mismo periodo, asegurando así la unificación de toda la información relevante en un único archivo para su posterior análisis.

- **Data Comercial Ventas KG.**
Tamaño del Dataset: (632, 970)

Diccionario de Datos:

- *Producto:* Códigos de materiales
- *Tipo:* Variables exógenas
- *Serie de tiempo:* Semana 52. 2021- Semana Enero 01.2025

Cantidad de semanas registradas: 158.

Se busca estructurar adecuadamente la información, por lo que se realiza un proceso de pivotado con el objetivo de obtener el formato más adecuado para modelar cada serie de tiempo (*ver Tabla II. Paso 1 del proceso de pivotado y Tabla III. Paso 2 del proceso de pivotado*)

Tabla II. Paso 1 del proceso de pivotado.

| Fecha | Material | Tipo | Valor |
|------------|----------|-------------|-----------|
| 27/12/2021 | 1 | Ventas \$\$ | \$ - |
| 3/01/2022 | 1 | Ventas \$\$ | \$ 40,000 |
| 10/01/2022 | 1 | Ventas \$\$ | \$ 36,000 |

Tabla III. Paso 2 del proceso de pivotado.

| Fecha | Material | Bonificado | Promoción (SI/NO) | Ventas Kg | Ventas \$\$ |
|------------|----------|------------|-------------------|-----------|-------------|
| 27/12/2021 | 1 | 0 | 0 | 0 | 0 |
| 27/12/2021 | 2 | 0 | 0 | 0 | 0 |
| 27/12/2021 | 3 | 0 | 0 | 0 | 0 |

Al trabajar con un dataset de frecuencia semanal y con el objetivo de asignar una fecha específica a cada registro, se estableció como referencia el día lunes de cada semana. De esta forma, a cada semana se le asigna la fecha correspondiente al lunes, utilizando el formato dd/mm/aaaa. Esto permite contar con una variable temporal homogénea y adecuada para el modelado de series de tiempo.

Posteriormente, se filtran los materiales del dataset con los siguientes criterios con el fin de encontrar los 2 materiales con mayor venta (*ver Figura 5. Ventas promedio anual por material y año*).

Filtro:

- Al menos 52 semanas de datos
- Al menos 5 valores distintos en Ventas _ Pesos
- Pocos nulos, menor al 20%

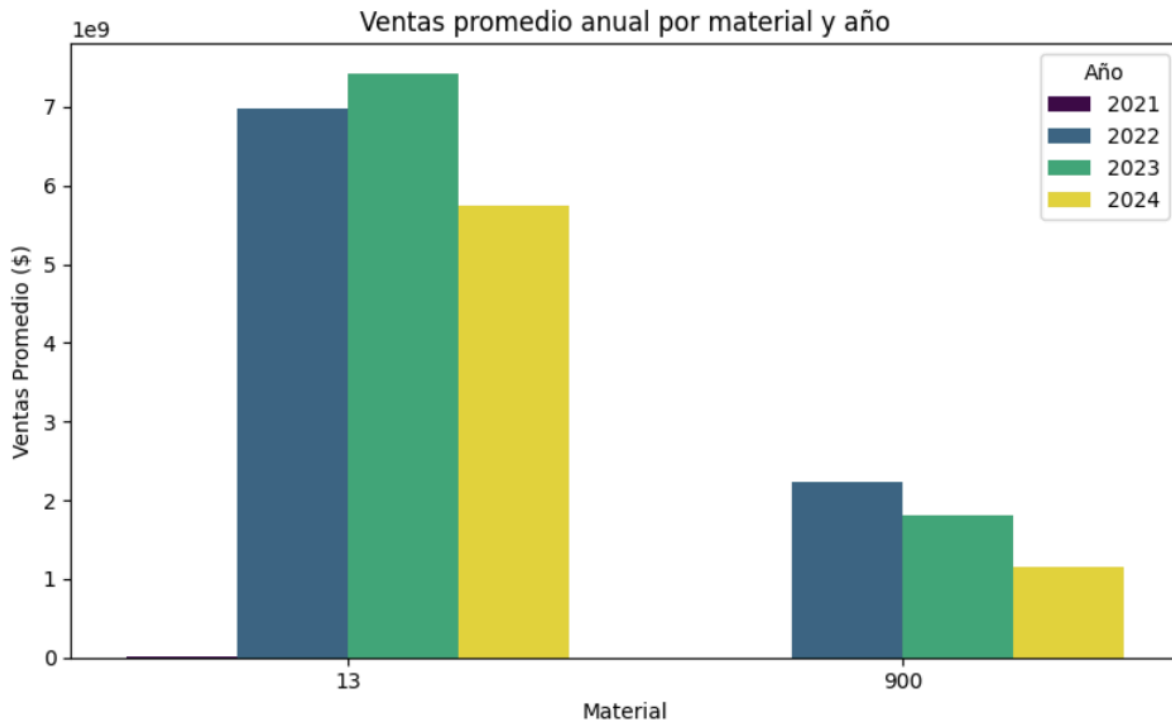


Figura 5. Ventas promedio anual por material y año

7.2.1.2 Outliers

A continuación se pueden observar los outliers de cada material y el gráfico de box plot de la venta en pesos (ver *Figura 6. Boxplot de las ventas en pesos para los dos materiales con mayor facturación.*).

Cantidad de outlier (material 13): 5

2021-12-27 5487846
 2022-12-19 11223609227
 2023-03-27 11503109586
 2023-04-24 11450449921
 2024-12-30 119341377

Cantidad de outlier (material 900): 6

2022-05-16 5440897451
 2022-06-13 5600379492
 2022-07-11 6260175790
 2022-11-14 5409327000

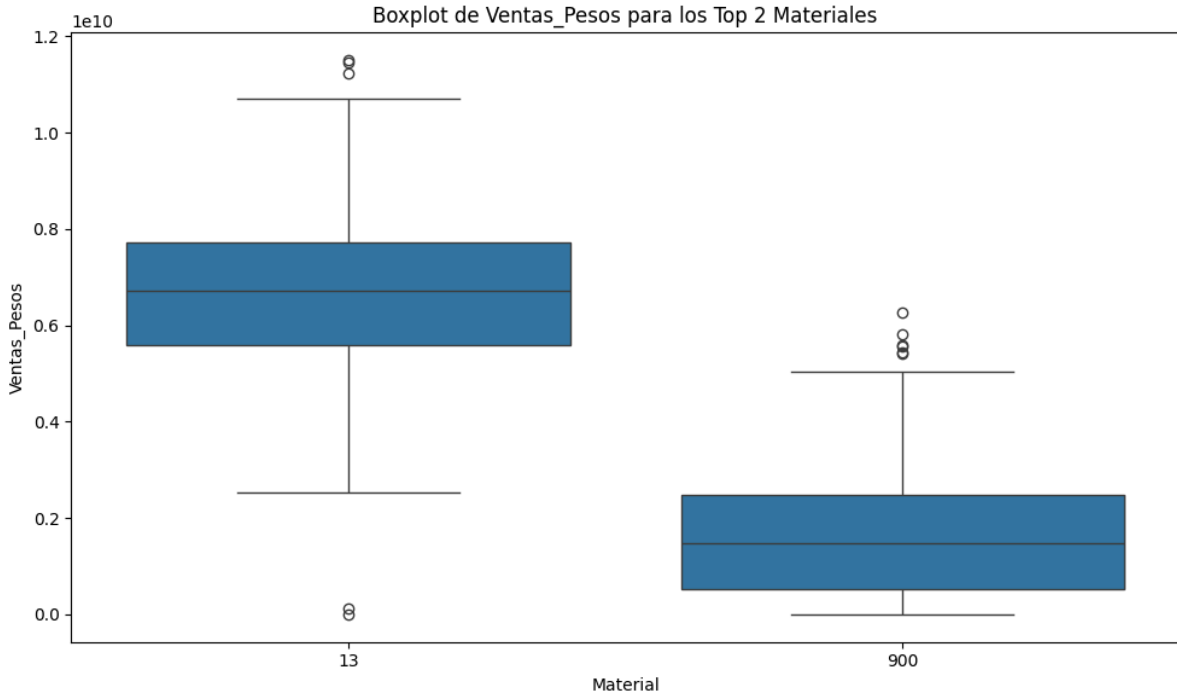


Figura 6. Boxplot de las ventas en pesos para los dos materiales con mayor facturación.

Se observa que el material 13 presenta una mediana y un rango intercuartílico más altos en comparación con el material 900, indicando una mayor dispersión y valores más altos en sus ventas. También se identifican varios valores atípicos en ambos casos.

Cálculo de IQR:

El IQR se determina, descrito como la brecha entre Q3 y Q1 para la variable

Definición de límites:

El rango considerado bien para los datos es $[Q1 - 1.5IQR, Q3 + 1.5iqr]$ se establece Los valores fuera de este rango se consideran atípicos

Reemplazo de valores atípicos:

Los valores inusuales no se eliminan; En cambio, se sustituyen con la media de sus puntos adyacentes en la serie temporal (la anterior y la después) Esto mantiene el orden de tiempo y disminuye el impacto de números inusuales Para valores extremos al comienzo o al final, se utiliza el único valor adyacente

7.2.1.3 Series temporales y correlaciones

Para cada variable corregida, se gráfica la evolución temporal tras el tratamiento de outliers, permitiendo una visualización clara del comportamiento real de la serie y facilitando la detección de tendencias o patrones tanto para el material 13 (ver Figura 7. Serie de tiempo imputadas de Ventas Kg para el material 13 y Figura 8. Serie de tiempo imputadas de Ventas Pesos para el material 13) como para el material 900 (ver Figura 9. Serie de tiempo imputadas de Ventas Kg para el material 900 y Figura 10. Serie de tiempo imputadas de Ventas Pesos para el material 900)

SERIE TEMPORAL MATERIAL 13:

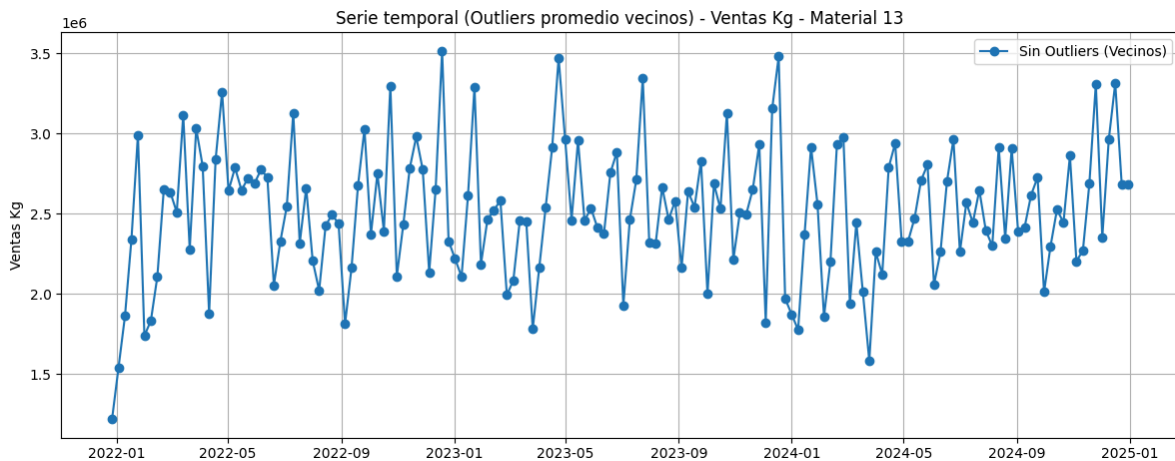


Figura 7. Serie de tiempo imputadas de Ventas Kg para el material 13.

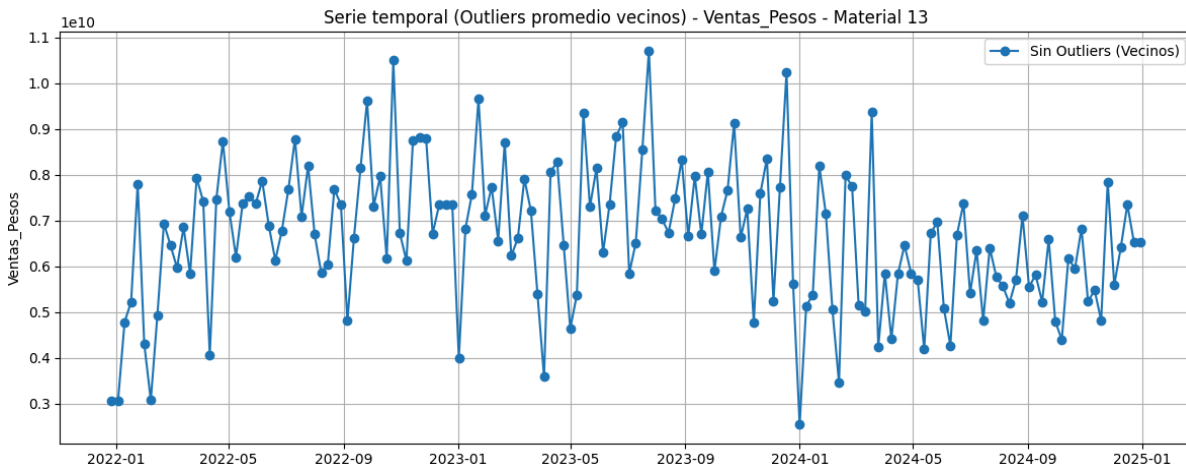


Figura 8. Serie de tiempo imputadas de Ventas Pesos para el material 13.

SERIE TEMPORAL MATERIAL 900:

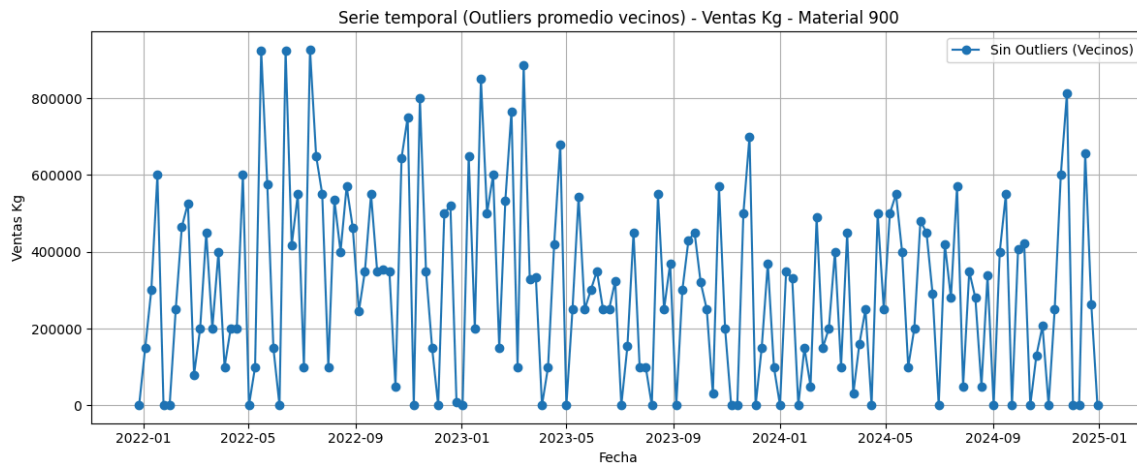


Figura 9. Serie de tiempo imputadas de Ventas_Kg para el material 900.

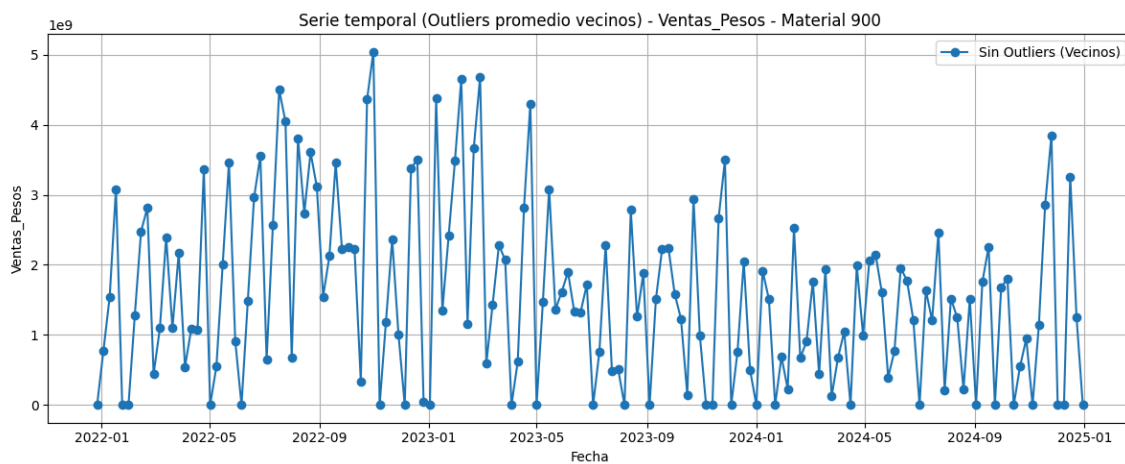


Figura 10. Serie de tiempo imputadas de Ventas_Pesos para el material 900.

Finalmente, se calcula la matriz de correlación entre las tres variables principales para cada material, empleando los datos ya corregidos. La matriz de correlación se visualiza mediante un mapa de calor (heatmap), facilitando la interpretación visual de la relación entre las variables después de la limpieza de outliers (ver Figura 11. *Matrices de correlaciones entre Ventas Kg, Ventas Pesos y Bonificado de los materiales 13 y 900.*).

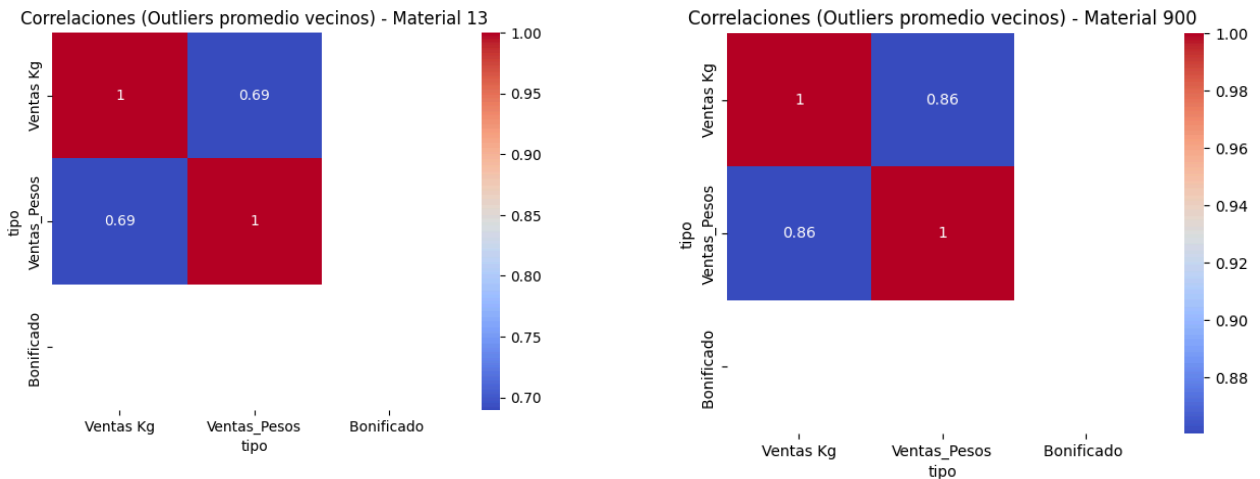


Figura 11. Matrices de correlaciones entre Ventas Kg, Ventas Pesos y Bonificado de los materiales 13 y 900.

El uso de la media de puntos cercanos para intercambiar valores atípicos proporciona un conjunto de datos a corto plazo más preciso, lo que mejora la confiabilidad de los exámenes futuros y evita los resultados sesgados en los cálculos estadísticos

Los gráficos separados para cada factor revelan el patrón verdadero de los datos basados en el tiempo después de los ajustes, ayudando a detectar desarrollos auténticos sin la interferencia de las figuras atípicas

El estudio de las conexiones entre los puntos de datos limpios ofrece una comprensión más clara de cómo los factores importantes se relacionan con cada sustancia, crucial para el pronóstico y las elecciones.

7.2.2 Modelos Predictivos

Se propone el uso del protocolo de evaluación con ventana recursiva con ventanas de entrenamiento de 40 y 20 semanas. También se varía el número de periodos hacia atrás con los cuales se entrena el modelo de 79 y 158 semanas (siendo 158 el número de semanas máximo del dataset). Para todos los modelos se utilizó una búsqueda de parámetros utilizando optimización bayesiana.

A continuación se explica el entrenamiento de cada modelo:

- **Promedio móvil:** Se probaron valores de la ventana del promedio móvil en un rango de 2 hasta 30 semanas realizando 15 iteraciones en la optimización bayesiana.
- **Suavizamiento exponencial:** Los hiper parámetros de error, trend and seasonal variaron entre ‘add’, ‘mul’ y ‘None’, mientras que los coeficientes de alpha, beta y gamma variaron entre 0.01 y 0.99. Todo esto con 35 iteraciones en la optimización bayesiana.
- **ARIMA:** Se encontraron los mejores hiper parámetros p (términos autorregresivos), q (rezagos de errores) y d (diferencias no estacionarias) cada uno en un rango de 0 a 10 que lograron el menor RMSE dentro de 15 iteraciones.
- **VAR:** Para hacer el pronóstico de la variable Ventas Kg se utilizó la variable de Ventas Pesos. Se probó dentro de un rango de 1 a 20 el mejor hiper parámetro p (términos autorregresivos) con 15 iteraciones.

A continuación se presentan los resultados del entrenamiento de cada modelo utilizando optimización bayesiana. Cada fila en las Tabla IV a la Tabla XI, representa el mejor resultado después de la ejecución del algoritmo de optimización con un valor específico de semanas seleccionadas, de ventana de entrenamiento y de un set de datos originales o limpios. Finalmente, cada tabla está ordenada de mayor a menor RMSE (el valor de menor RMSE está indicado en amarillo).

7.2.2.1 Material 13

El mejor modelo encontrado fue un ARIMA con parámetros $p = 2$, $d = 0$ y $q = 4$.

Tabla IV. Resultado del entrenamiento con promedio móvil para el material 13.

| RMSE | Semanas seleccionadas | Ventana de entrenamiento | Datos usados | Ventana del promedio móvil |
|-----------|-----------------------|--------------------------|--------------|----------------------------|
| 292922.33 | 79 | 40 | Originales | 14 |
| 295289.14 | 79 | 40 | Limpios | 30 |
| 367942.23 | 79 | 20 | Limpios | 18 |
| 380202.48 | 158 | 20 | Limpios | 30 |
| 383703.96 | 158 | 40 | Limpios | 30 |
| 409568.55 | 79 | 20 | Originales | 14 |
| 442282.56 | 158 | 20 | Originales | 30 |
| 453681.98 | 158 | 40 | Originales | 18 |

Tabla V. Resultado del entrenamiento con suavizamiento exponencial para el material 13.

| RMSE | Semanas seleccionadas | Ventana de entrenamiento | Datos usados | alfa | beta | gamma | error | seasonal | trend |
|-----------|-----------------------|--------------------------|--------------|------|------|-------|-------|----------|-------|
| 283095.73 | 79 | 40 | Limpios | 0.01 | 0.99 | 0.01 | mul | None | None |
| 297018.95 | 79 | 40 | Originales | 0.01 | 0.99 | 0.01 | mul | None | None |
| 360142.1 | 79 | 20 | Limpios | 0.01 | 0.99 | 0.01 | mul | None | None |
| 410767.78 | 79 | 20 | Originales | 0.01 | 0.99 | 0.01 | mul | None | None |
| 471809.18 | 158 | 20 | Originales | 0.32 | 0.7 | 0.61 | add | None | None |
| 482913.23 | 158 | 40 | Originales | 0.32 | 0.7 | 0.61 | add | None | None |
| 512879.72 | 158 | 20 | Limpios | 0.99 | 0.01 | 0.01 | add | None | None |
| 522156.32 | 158 | 40 | Limpios | 0.99 | 0.01 | 0.01 | add | None | None |

Tabla VI. Resultado del entrenamiento con ARIMA para el material 13.

| RMSE | Semanas seleccionadas | Ventana de entrenamiento | Datos Usados | p | d | q |
|-----------|-----------------------|--------------------------|--------------|---|---|---|
| 249481.69 | 79 | 40 | Limpios | 2 | 0 | 4 |
| 283507.7 | 79 | 40 | Originales | 2 | 0 | 4 |
| 348765.66 | 79 | 20 | Limpios | 3 | 0 | 4 |
| 378200.62 | 158 | 20 | Limpios | 2 | 0 | 7 |
| 412565.91 | 158 | 40 | Limpios | 3 | 1 | 1 |
| 435881.69 | 79 | 20 | Originales | 3 | 1 | 1 |
| 457918.35 | 158 | 40 | Originales | 2 | 0 | 4 |
| 468925.9 | 158 | 20 | Originales | 4 | 1 | 1 |

Tabla VII. Resultado del entrenamiento con VAR para el material 13.

| Producto | Semanas seleccionadas | Ventana de entrenamiento | RMSE | p |
|-----------|-----------------------|--------------------------|------------|---|
| 275799.02 | 79 | 40 | Originales | 4 |
| 281897.95 | 79 | 40 | Limpios | 2 |
| 366960.48 | 79 | 20 | Limpios | 1 |
| 378303.43 | 158 | 40 | Limpios | 4 |
| 378879.67 | 158 | 20 | Limpios | 1 |
| 414246.32 | 79 | 20 | Originales | 1 |
| 434877.86 | 158 | 40 | Originales | 5 |
| 435976.13 | 158 | 20 | Originales | 4 |

7.2.2.2 Material 900

El mejor modelo encontrado fue de promedio móvil con una ventana de 30.

Tabla VIII. Resultado del entrenamiento con promedio móvil para el material 900.

| RMSE | Semanas seleccionadas | Ventana de entrenamiento | Datos usados | Ventana del promedio móvil |
|-----------|-----------------------|--------------------------|--------------|----------------------------|
| 198243.5 | 79 | 20 | Originales | 30 |
| 198243.5 | 79 | 20 | Limpios | 30 |
| 210706.56 | 79 | 40 | Limpios | 30 |
| 210706.56 | 79 | 40 | Originales | 30 |
| 223804.62 | 158 | 40 | Limpios | 18 |
| 223804.62 | 158 | 40 | Originales | 18 |
| 228228.82 | 158 | 20 | Originales | 18 |
| 228228.82 | 158 | 20 | Limpios | 18 |

Tabla IX. Resultado del entrenamiento con suavizamiento exponencial para el material 900.

| RMSE | Semanas seleccionadas | Ventana de entrenamiento | Datos usados | alfa | beta | gamma | error | seasonal | trend |
|-----------|-----------------------|--------------------------|--------------|------|------|-------|-------|----------|-------|
| 213746.2 | 79 | 20 | Originales | 0.12 | 0.01 | 0.01 | add | add | None |
| 213746.2 | 79 | 20 | Limpios | 0.12 | 0.01 | 0.01 | add | add | None |
| 227016.32 | 79 | 40 | Limpios | 0.28 | 0.01 | 0.01 | add | add | None |
| 227016.32 | 79 | 40 | Originales | 0.28 | 0.01 | 0.01 | add | add | None |
| 269944.31 | 158 | 40 | Limpios | 0.77 | 0.59 | 0.5 | add | None | None |
| 269944.31 | 158 | 40 | Originales | 0.77 | 0.59 | 0.5 | add | None | None |
| 296952.88 | 158 | 20 | Originales | 0.79 | 0.01 | 0.47 | add | add | None |
| 296952.88 | 158 | 20 | Limpios | 0.79 | 0.01 | 0.47 | add | add | None |

Tabla X. Resultado del entrenamiento con ARIMA para el material 900.

| RMSE | Semanas seleccionadas | Ventana de entrenamiento | Datos Usados | p | d | q |
|-----------|-----------------------|--------------------------|--------------|---|---|---|
| 203203.65 | 79 | 20 | Originales | 0 | 1 | 3 |
| 203203.65 | 79 | 20 | Limpios | 0 | 1 | 3 |
| 216197.9 | 79 | 40 | Limpios | 4 | 1 | 1 |
| 216197.9 | 79 | 40 | Originales | 4 | 1 | 1 |
| 229907.78 | 158 | 40 | Limpios | 3 | 1 | 1 |
| 229907.78 | 158 | 40 | Originales | 3 | 1 | 1 |
| 231546.68 | 158 | 20 | Originales | 3 | 1 | 1 |

| | | | | | | |
|-----------|-----|----|---------|---|---|---|
| 231546.68 | 158 | 20 | Limpios | 3 | 1 | 1 |
|-----------|-----|----|---------|---|---|---|

Tabla XI. Resultado del entrenamiento con VAR para el material 900.

| Producto | Semanas seleccionadas | Ventana de entrenamiento | RMSE | p |
|-----------|-----------------------|--------------------------|------------|---|
| 198519.79 | 79 | 20 | Originales | 1 |
| 198519.79 | 79 | 20 | Limpios | 1 |
| 211449.68 | 79 | 40 | Limpios | 1 |
| 211449.68 | 79 | 40 | Originales | 1 |
| 226140.38 | 158 | 40 | Originales | 2 |
| 226722.19 | 158 | 40 | Limpios | 3 |
| 231093.16 | 158 | 20 | Originales | 1 |
| 235187.89 | 158 | 20 | Limpios | 1 |

7.2.2.3 Observaciones

En ambos protocolos de evaluación (material 13 y 900) se obtuvieron mejores resultados entrenando con la mitad de los datos (79 semanas) que utilizando todos los datos (158 semanas). Adicionalmente, para el material 13 la mejor ventana de entrenamiento fue de 40 datos mientras que para el material 900 fue de 20 datos. En cuanto al tipo de datos usados, el material 13 logró buenos resultados con datos limpios y originales mientras que con el material 900, no hubo diferencia en el RMSE resultante debido a que muy pocos puntos fueron imputados en los datos limpios, y por tanto, aún son muy similares a los originales.

Sobre los modelos entrenados, para el material 13 los mejores resultados fueron los siguientes:

- Promedio Móvil: Con ventana de 14 o de 30 datos.
- Suavizamiento exponencial: Sin componente de tendencia ni estacionalidad y con error multiplicativo.
- ARIMA: Sin componente diferencial.
- VAR: En general, con menos de 5 rezagos.

Para el material 900 los mejores resultados fueron los siguientes:

- Promedio Móvil: Con ventana de 30 datos.
- Suavizamiento exponencial: Sin componente de tendencia, con estacionalidad y con error aditivo
- ARIMA: Con componente diferencial igual a 1.
- VAR: Con un valor de $p = 1$.

7.3. Análisis de residuales

A continuación se muestra un resumen del análisis de residuales de los modelos evaluados. Se evalúan los residuales para cada producto el mejor modelo, de no cumplirse los supuestos del modelo, se continúa con el mejor modelo de la lista.

Para verificar la distribución normal de los residuales se aplican 3 pruebas: Shapiro-Wilk, Kolmogorov-Smirnov y Kolmogorov-Smirnov.

7.3.1 Residuales del modelo ARIMA del material 13

7.3.1.1. Esperanza de los residuales es cero

H0: Media de las observaciones es igual a cero

H1: Media de las observaciones es diferente a cero

Como p-valor = 0.9853166715676274 > 0.05, con un nivel de significancia alfa = 0.05, NO existe evidencia muestral suficiente para rechazar la hipótesis nula.

Por lo tanto SÍ suponemos esperanza de los residuales igual a cero.

7.3.1.2. Distribución normal de los residuales

Prueba de Shapiro-Wilk.

H0: Los errores se distribuyen normal

H1: Los errores no se distribuyen normal

Como p-valor = 0.0517813859458704 > 0.05, con un nivel de significancia alfa = 0.05, NO existe evidencia muestral suficiente para rechazar la hipótesis nula.

Por lo tanto SÍ suponemos distribución normal de los residuales por Shapiro-Wilk.

Distribución normal de los residuales:

Prueba de Kolmogorov-Smirnov.

H0: Los errores se distribuyen normal

H1: Los errores no se distribuyen normal

Como p-valor = 1.3493114002851118e-23 <= 0.05, con un nivel de significancia alfa = 0.05, SÍ existe evidencia muestral suficiente para rechazar la hipótesis nula.

Por lo tanto NO suponemos distribución normal de los residuales por Kolmogorov-Smirnov.

Distribución normal de los residuales:

Prueba de Anderson-Darling.

H0: Los errores se distribuyen normal

H1: Los errores no se distribuyen normal

Como $p\text{-valor} = 0.752 > 0.05$, con un nivel de significancia $\alpha = 0.05$, NO existe evidencia muestral suficiente para rechazar la hipótesis nula.

Por lo tanto SÍ suponemos distribución normal de los residuales por Anderson-Darling.

7.3.1.3. Independencia en los residuales

Tabla XII. Test de autocorrelación Box-Pierce y la modificación de Ljung-Box de los residuales del modelo del material 13.

| iter | lb_stat | lb_pvalue | bp_stat | bp_pvalue | lb_ho | bp_ho |
|------|---------|-----------|---------|-----------|-----------------------------------|-----------------------------------|
| 1 | 0 | 1 | 0 | 1 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 2 | 0 | 1 | 0 | 1 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 3 | 0.12 | 0.99 | 0.11 | 0.99 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 4 | 1.69 | 0.79 | 1.56 | 0.82 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 5 | 1.73 | 0.89 | 1.6 | 0.9 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 6 | 2.77 | 0.84 | 2.54 | 0.86 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 7 | 2.92 | 0.89 | 2.68 | 0.91 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 8 | 4.52 | 0.81 | 4.08 | 0.85 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 9 | 4.63 | 0.87 | 4.17 | 0.9 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 10 | 4.63 | 0.91 | 4.17 | 0.94 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 11 | 5.75 | 0.89 | 5.11 | 0.93 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 12 | 6.09 | 0.91 | 5.39 | 0.94 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 13 | 7.27 | 0.89 | 6.35 | 0.93 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 14 | 13.29 | 0.5 | 11.19 | 0.67 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 15 | 13.36 | 0.57 | 11.24 | 0.74 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |
| 16 | 13.43 | 0.64 | 11.3 | 0.79 | no_rechazo_ho/no_hay_aut ocorr | no_rechazo_ho/no_hay_a utocorr |

| | | | | | | |
|----|-------|------|-------|------|-------------------------------|-------------------------------|
| 17 | 13.65 | 0.69 | 11.46 | 0.83 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 18 | 16.63 | 0.55 | 13.71 | 0.75 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 19 | 17.35 | 0.57 | 14.24 | 0.77 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 20 | 17.37 | 0.63 | 14.25 | 0.82 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |

7.3.1.4. Homocedasticidad

Tabla XIII. Prueba de Box-Pierce y la modificación de Ljung-Box para evaluación de homocedasticidad para los residuales del modelo del material 13.

| iter | lb_stat | lb_pvalue | bp_stat | bp_pvalue | lb_ho | bp_ho |
|------|---------|-----------|---------|-----------|-------------------------------|-------------------------------|
| 1 | 4.77 | 0.03 | 4.59 | 0.03 | rechazo_ho/hay_autocorr | rechazo_ho/hay_autocorr |
| 2 | 4.81 | 0.09 | 4.63 | 0.1 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 3 | 5.49 | 0.14 | 5.27 | 0.15 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 4 | 5.55 | 0.24 | 5.32 | 0.26 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 5 | 6.84 | 0.23 | 6.51 | 0.26 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 6 | 7.46 | 0.28 | 7.06 | 0.31 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 7 | 7.67 | 0.36 | 7.25 | 0.4 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 8 | 8.43 | 0.39 | 7.92 | 0.44 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 9 | 8.99 | 0.44 | 8.4 | 0.49 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 10 | 9.01 | 0.53 | 8.41 | 0.59 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 11 | 9.01 | 0.62 | 8.41 | 0.68 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 12 | 9.03 | 0.7 | 8.43 | 0.75 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 13 | 12.28 | 0.5 | 11.08 | 0.6 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 14 | 14.85 | 0.39 | 13.14 | 0.52 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |

| | | | | | autocorr | corr |
|----|-------|------|-------|------|-------------------------------|-------------------------------|
| 15 | 15.76 | 0.4 | 13.86 | 0.54 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 16 | 16.01 | 0.45 | 14.05 | 0.59 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 17 | 16.82 | 0.47 | 14.67 | 0.62 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 18 | 18.73 | 0.41 | 16.12 | 0.58 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 19 | 21.92 | 0.29 | 18.47 | 0.49 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 20 | 22.01 | 0.34 | 18.55 | 0.55 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |

A continuación se presenta la gráfica de dispersión de los residuales (ver. *Figura 12. Residuales del modelo ARIMA del material 13*) la cual nos indica el valor de cada residual en función del tiempo, y su distribución en histograma (*Figura 13. Distribución en histograma de los residuales del modelo ARIMA del material 13*), a fin de complementar el análisis a partir de pruebas estadísticas. La distribución del histograma corrobora la presencia de una distribución normal anteriormente representada por los test de normalidad (2 de 3 pruebas fueron afirmativas).

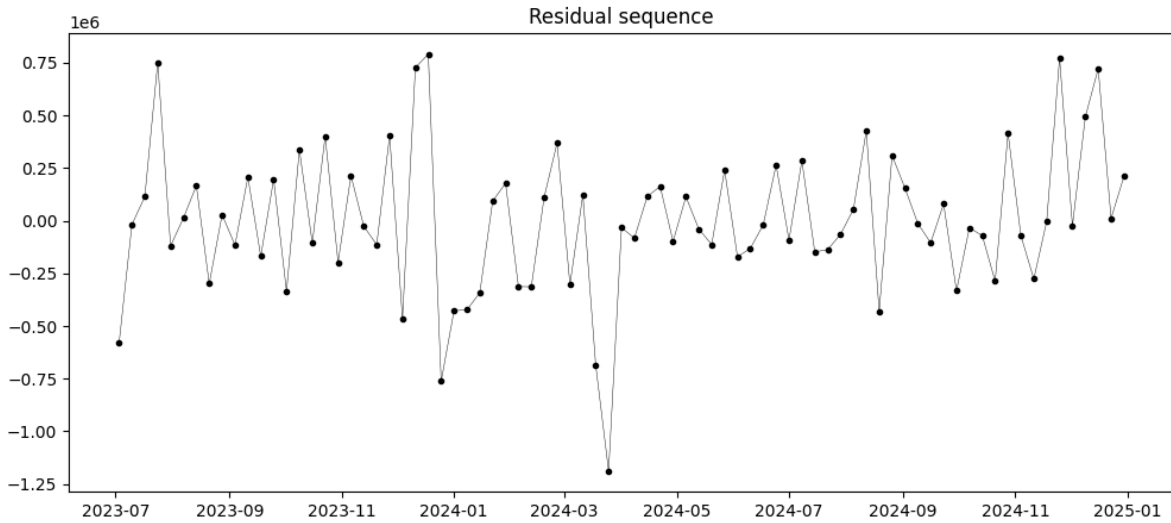


Figura 12. Residuales del modelo ARIMA del material 13.

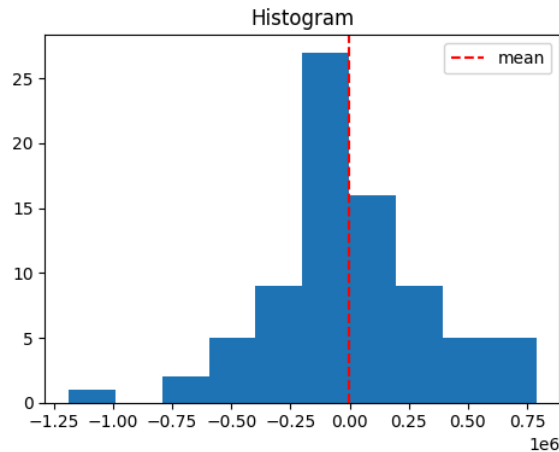


Figura 13. Distribución en histograma de los residuales del modelo ARIMA del material 13.

Finalmente, se corrobora la distribución normal a partir del gráfico q-q de los residuales (ver *Figura 14. Gráfico q-q de los residuales del modelo ARIMA del material 13.*), ya que este es un gráfico que se utiliza para comparar la distribución de una muestra de datos con una distribución teórica, que en este caso es la normal.

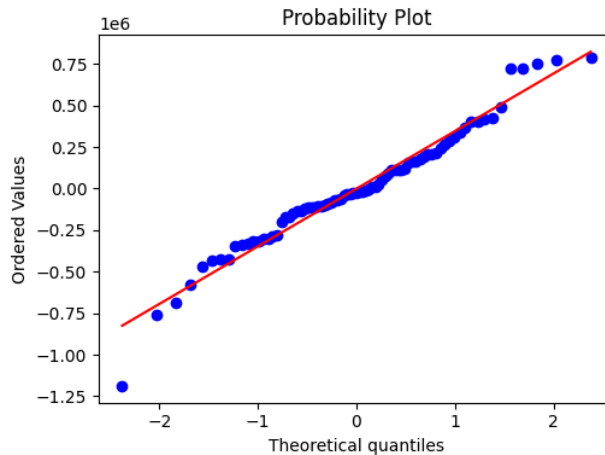


Figura 14. Gráfico q-q de los residuales del modelo ARIMA del material 13.

Ya que se confirma el comportamiento de ruido blanco de los residuales bajo una distribución normal, se puede utilizar el modelo para generar predicciones en el año 2025.

7.3.2 Residuales del modelo Promedio Móvil del material 900

7.3.2.1. Esperanza de los residuales es cero

H0: Media de las observaciones es igual a cero

H1: Media de las observaciones es diferente a cero

Como $p\text{-valor} = 0.760955050925302 > 0.05$, con un nivel de significancia $\alpha = 0.05$, NO existe evidencia muestral suficiente para rechazar la hipótesis nula.

Por lo tanto SÍ suponemos que la esperanza de los residuales es igual a cero.

7.3.2.2. Distribución normal de los residuales

Distribución normal de los residuales:

Prueba de Shapiro-Wilk.

H0: Los errores se distribuyen normal

H1: Los errores no se distribuyen normal

Como $p\text{-valor} = 0.053441007621326954 > 0.05$, con un nivel de significancia $\alpha = 0.05$, NO existe evidencia muestral suficiente para rechazar la hipótesis nula.

Por lo tanto SÍ suponemos distribución normal de los residuales por Shapiro-Wilk.

Distribución normal de los residuales:

Prueba de Kolmogorov-Smirnov.

H0: Los errores se distribuyen normal

H1: Los errores no se distribuyen normal

Como $p\text{-valor} = 2.3501115697123033e-12 \leq 0.05$, con un nivel de significancia $\alpha = 0.05$, SÍ existe evidencia muestral suficiente para rechazar la hipótesis nula.

Por lo tanto NO suponemos distribución normal de los residuales por Kolmogorov-Smirnov.

Distribución normal de los residuales:

Prueba de Anderson-Darling.

H0: Los errores se distribuyen normal

H1: Los errores no se distribuyen normal

Como $p\text{-valor} = 0.735 > 0.05$, con un nivel de significancia $\alpha = 0.05$, NO existe evidencia muestral suficiente para rechazar la hipótesis nula.

Por lo tanto SÍ suponemos distribución normal de los residuales por Anderson-Darling.

7.3.2.3. Independencia en los residuales (Test de autocorrelación Box-Pierce y la modificación de Ljung-Box)

Tabla XIV. Test de autocorrelación Box-Pierce y la modificación de Ljung-Box de los residuales del modelo del material 900.

| iter | lb_stat | lb_pvalue | bp_stat | bp_pvalue | lb_ho | bp_ho |
|------|---------|-----------|---------|-----------|-------------------------------|-------------------------------|
| 1 | 0.87 | 0.35 | 0.82 | 0.37 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 2 | 3.87 | 0.14 | 3.58 | 0.17 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |

| | | | | | | |
|----|-------|------|-------|------|-------------------------------|-------------------------------|
| 3 | 4.59 | 0.2 | 4.23 | 0.24 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 4 | 4.7 | 0.32 | 4.33 | 0.36 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 5 | 4.75 | 0.45 | 4.37 | 0.5 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 6 | 7.06 | 0.32 | 6.32 | 0.39 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 7 | 7.36 | 0.39 | 6.57 | 0.48 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 8 | 7.76 | 0.46 | 6.89 | 0.55 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 9 | 7.93 | 0.54 | 7.02 | 0.63 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 10 | 9.58 | 0.48 | 8.28 | 0.6 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 11 | 9.9 | 0.54 | 8.53 | 0.67 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 12 | 12.54 | 0.4 | 10.44 | 0.58 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 13 | 13 | 0.45 | 10.76 | 0.63 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 14 | 13.16 | 0.51 | 10.88 | 0.7 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 15 | 15.39 | 0.42 | 12.36 | 0.65 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 16 | 15.39 | 0.5 | 12.36 | 0.72 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 17 | 15.87 | 0.53 | 12.66 | 0.76 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 18 | 18.85 | 0.4 | 14.47 | 0.7 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 19 | 19.3 | 0.44 | 14.73 | 0.74 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 20 | 19.45 | 0.49 | 14.82 | 0.79 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |

7.3.2.4. Homocedasticidad

Tabla XV. Prueba de Box-Pierce y la modificación de Ljung-Box para evaluación de homocedasticidad para los residuales del modelo del material 900.

| iter | lb_stat | lb_pvalue | bp_stat | bp_pvalue | lb_ho | bp_ho |
|------|---------|-----------|---------|-----------|-------------------------------|-------------------------------|
| 1 | 0.91 | 0.34 | 0.85 | 0.36 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 2 | 0.93 | 0.63 | 0.88 | 0.64 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 3 | 8.88 | 0.03 | 8.05 | 0.05 | rechazo_ho/hay_autocorr | rechazo_ho/hay_autocorr |
| 4 | 9.31 | 0.05 | 8.43 | 0.08 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 5 | 9.44 | 0.09 | 8.54 | 0.13 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 6 | 11.38 | 0.08 | 10.17 | 0.12 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 7 | 12.18 | 0.09 | 10.83 | 0.15 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 8 | 12.18 | 0.14 | 10.83 | 0.21 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 9 | 13.58 | 0.14 | 11.93 | 0.22 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 10 | 13.66 | 0.19 | 11.99 | 0.29 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 11 | 13.98 | 0.23 | 12.23 | 0.35 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 12 | 14.74 | 0.26 | 12.78 | 0.39 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 13 | 15.69 | 0.27 | 13.45 | 0.41 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 14 | 15.81 | 0.33 | 13.53 | 0.48 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 15 | 15.86 | 0.39 | 13.57 | 0.56 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 16 | 19.93 | 0.22 | 16.2 | 0.44 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 17 | 20.63 | 0.24 | 16.64 | 0.48 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 18 | 21.25 | 0.27 | 17.01 | 0.52 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 19 | 21.6 | 0.3 | 17.22 | 0.57 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |
| 20 | 21.61 | 0.36 | 17.23 | 0.64 | no_rechazo_ho/no_hay_autocorr | no_rechazo_ho/no_hay_autocorr |

De igual manera al caso anterior, se presenta la gráfica de dispersión de los residuales (ver *Figura 14. Residuales del modelo ARIMA del material 900*) y su distribución (*Figura 15. Distribución en histograma de los residuales del modelo ARIMA del material 900*), a fin de complementar el análisis a partir de pruebas estadísticas. En este caso, la distribución de los residuales no se considera normal a pesar de que 2 de las 3 pruebas de normalidad fueron positivas.

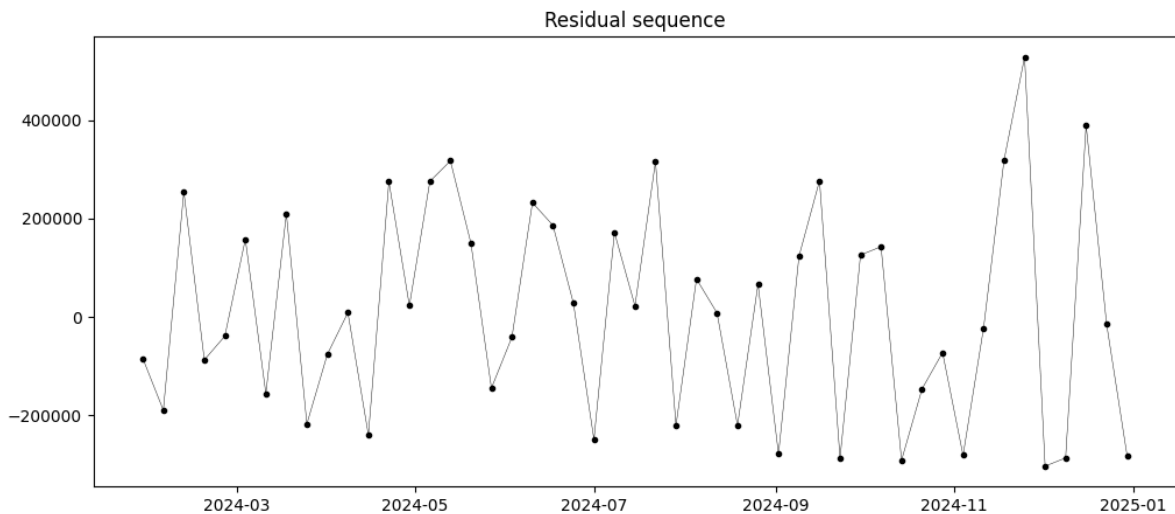


Figura 14. Residuales del modelo de Promedio Móvil del material 900.

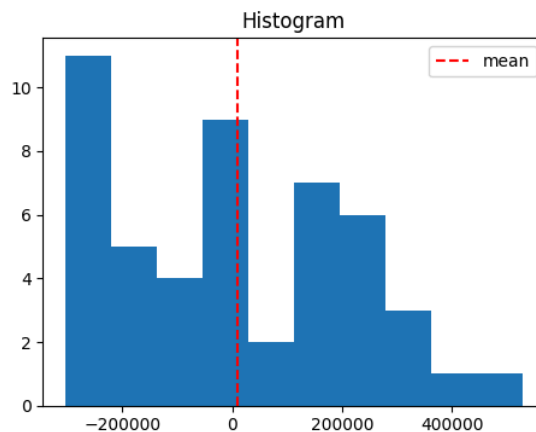


Figura 15. Distribución en histograma de los residuales del modelo de Promedio Móvil del material 900.

Al revisar el gráfico q-q (ver *Figura 16. Gráfico q-q de los residuales del modelo Promedio Móvil del material 900*) se corrobora el hecho de que este no tiene una distribución de tipo normal.

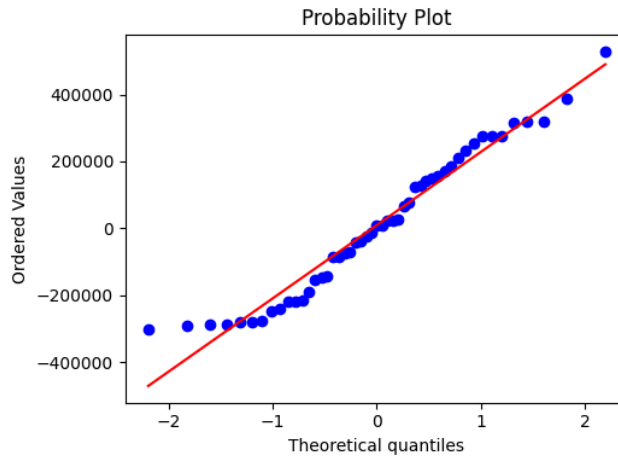


Figura 16. Gráfico q-q de los residuales del modelo Promedio Móvil del material 900.

Como los residuales del modelo si se comportan como ruido blanco pero sin distribución normal, entonces los intervalos de confianza del modelo se calcularán por medio de bootstrapping.

7.4. Resultados

Se puede observar en las siguientes gráficas (ver *Figura 17. Predicción de 12 semanas del modelo ARIMA para el material 13* y *Figura 18. Predicción de 12 semanas del modelo Promedio Móvil para el material 900*) las 12 semanas pronosticadas de los modelos entrenados para el material 13 y 900. Ambas gráficas contienen datos de entrada (datos originales - azul claro), los datos limpios (con imputación de atípicos - azul oscuro punteado), los resultados de prueba (naranja), los resultados de predicción (verde) y sus intervalos de confianza (gris punteado).

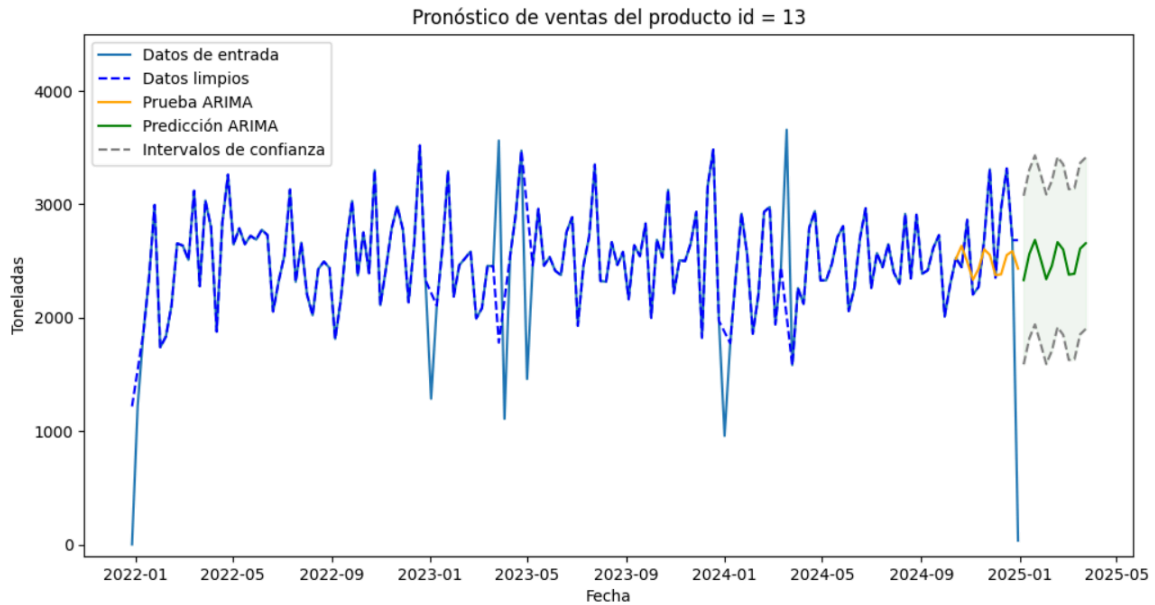


Figura 17. Predicción de 12 semanas del modelo ARIMA para el material 13.

Tabla XVI. Predicción de 12 semanas del modelo ARIMA para el material 13.

| Fecha | Ventas Kg | Intervalo confianza | |
|------------|------------|---------------------|------------|
| | | Inferior | Superior |
| 2025-01-06 | 2331667.67 | 1588852.8 | 3074482.54 |
| 2025-01-13 | 2560627.64 | 1817487.68 | 3303767.6 |
| 2025-01-20 | 2685325.03 | 1940813.02 | 3429837.03 |
| 2025-01-27 | 2518596.76 | 1771521.66 | 3265671.85 |
| 2025-02-03 | 2339092.99 | 1591999.28 | 3086186.7 |
| 2025-02-10 | 2458184.45 | 1708811.99 | 3207556.91 |
| 2025-02-17 | 2665568.55 | 1915825.31 | 3415311.79 |
| 2025-02-24 | 2600568.86 | 1849080 | 3352057.73 |
| 2025-03-03 | 2379372.29 | 1626973.62 | 3131770.95 |
| 2025-03-10 | 2387473.69 | 1633973.03 | 3140974.35 |
| 2025-03-17 | 2607884.16 | 1852901.25 | 3362867.06 |
| 2025-03-24 | 2655751.38 | 1900262.33 | 3411240.44 |

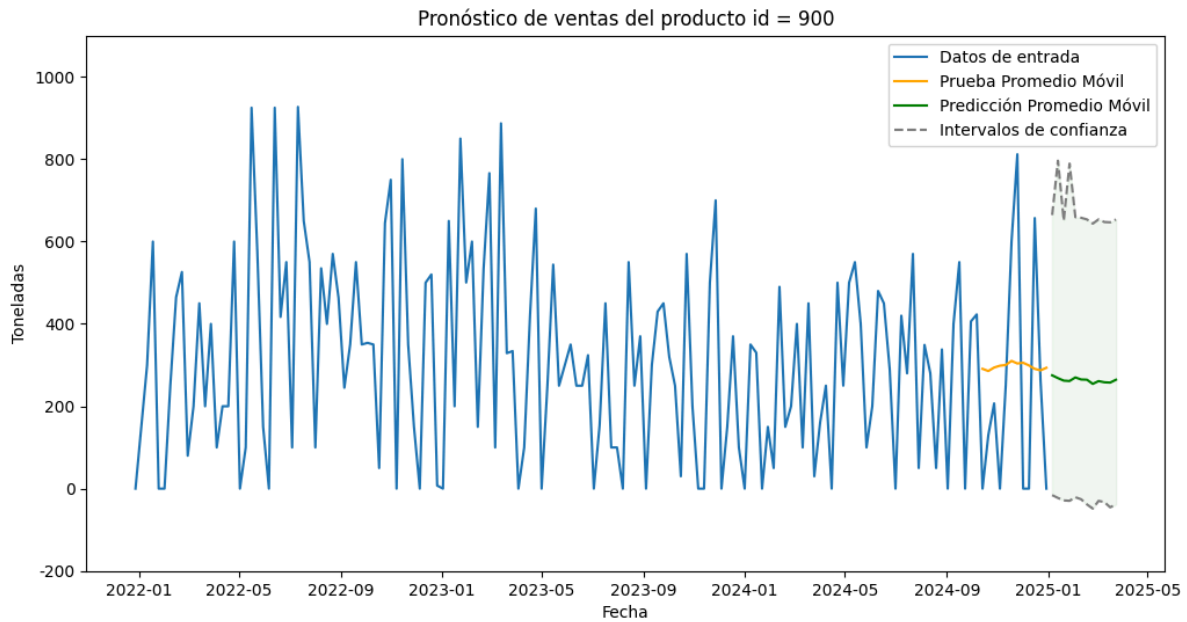


Figura 18. Predicción de 12 semanas del modelo Promedio Móvil para el material 900.

Tabla XVII. Predicción de 12 semanas del modelo Promedio Móvil para el material 900.

| Fecha | Ventas Kg | Intervalo confianza | |
|-------------------|-----------|---------------------|-----------|
| | | Inferior | Superior |
| 2025-01-06 | 275133.67 | -15741 | 664292.33 |
| 2025-01-13 | 268304.79 | -22569.88 | 796196.79 |
| 2025-01-20 | 262248.28 | -28626.38 | 651406.95 |
| 2025-01-27 | 261323.22 | -29551.44 | 789215.22 |
| 2025-02-03 | 270034 | -20840.67 | 659192.67 |
| 2025-02-10 | 265035.13 | -25839.53 | 657662.13 |
| 2025-02-17 | 264536.3 | -38305.03 | 653694.97 |
| 2025-02-24 | 254354.18 | -48487.15 | 643512.85 |
| 2025-03-03 | 261165.99 | -29708.68 | 653792.99 |
| 2025-03-10 | 258238.19 | -32636.48 | 647396.85 |
| 2025-03-17 | 257512.79 | -45328.54 | 646671.46 |
| 2025-03-24 | 264429.88 | -38411.45 | 653588.55 |

7.5. Comparación con valores reales

En esta sección se comparan los resultados obtenidos con el pronóstico colaborativo y de consenso realizado por las distintas áreas de la empresa. Se busca determinar que los modelos aquí evaluados son mucho más efectivos que dicha estimación.

En la *Tabla XVIII. Comparación semanal entre pronósticos del RMSE para el material 13* se puede ver la semana evaluada (Semana), el pronóstico obtenido en el presente trabajo (Pronóstico), el pronóstico de consenso (Colaborativo), el valor real de ventas Kg de esa semana (Valor Real) y el cálculo del RMSE al final de las columnas “(Pron. - Real)**2” y “(Colab. - Real)**2”.

Tabla XVIII. Comparación semanal entre pronósticos del RMSE para el material 13.

| Semana | Pronóstico | Colaborativo | Valor Real | (Pron. - Real)**2 | (Colab. - Real)**2 |
|------------|--------------|--------------|--------------|----------------------|-----------------------|
| 2025-01-06 | 2,331,667.67 | 4,016,710.00 | 751,750.00 | 2,496,139,843,978.23 | 10,659,963,801,600.00 |
| 2025-01-13 | 2,560,627.64 | 2,124,168.00 | 1,505,100.00 | 1,114,138,598,803.97 | 383,245,188,624.00 |
| 2025-01-20 | 2,685,325.03 | 2,124,168.00 | 2,059,750.00 | 391,344,118,159.50 | 4,149,678,724.00 |
| 2025-01-27 | 2,518,596.76 | 2,124,163.00 | 2,846,300.00 | 107,389,413,506.50 | 521,481,846,769.00 |
| 2025-02-03 | 2,339,092.99 | 4,270,085.00 | 2,660,150.00 | 103,077,603,670.14 | 2,591,890,704,225.00 |
| 2025-02-10 | 2,458,184.45 | 2,146,230.00 | 1,401,450.00 | 1,116,687,697,816.80 | 554,697,248,400.00 |
| 2025-02-17 | 2,665,568.55 | 2,146,230.00 | 1,908,800.00 | 572,698,638,269.10 | 56,373,004,900.00 |
| 2025-02-24 | 2,600,568.86 | 2,146,224.00 | 2,465,350.00 | 18,284,140,099.70 | 101,841,403,876.00 |
| 2025-03-03 | 2,379,372.29 | 3,953,063.00 | 3,155,650.00 | 602,607,083,042.84 | 635,867,492,569.00 |
| 2025-03-10 | 2,387,473.69 | 1,806,685.00 | 1,758,750.00 | 395,293,478,367.22 | 2,297,764,225.00 |
| 2025-03-17 | 2,607,884.16 | 1,806,683.00 | 2,190,500.00 | 174,209,537,018.91 | 147,315,489,489.00 |
| 2025-03-24 | 2,655,751.38 | 1,806,665.00 | 2,362,500.00 | 85,996,371,871.90 | 308,952,547,225.00 |
| | | | RMSE | 773,405.16 | 1,153,548.01 |

En la *Tabla XIX. Comparación mensual entre pronósticos del RMSE para el material 13* se puede ver el mes evaluado (Mes), el pronóstico obtenido en el presente trabajo (Pronóstico) que se calcula a partir de agregar por mes la suma de los pronósticos de cada semana (ej. , el mes de Enero 2025 es la suma de los pronósticos individuales de 2025-01-06, 2025-01-13, 2025-01-20 y 2025-01-27), pronóstico de consenso (Colaborativo), el valor real de ventas Kg de esa semana (Valor Real) y el cálculo del RMSE al final de las columnas “(Pron. - Real)**2” y “(Colab. - Real)**2”.

Tabla XIX. Comparación mensual entre pronósticos del RMSE para el material 13.

| Mes | Pronóstico | Colaborativo | Valor Real | (Pron. - Real)**2 | (Colab. - Real)**2 |
|------------|---------------|---------------|---------------|----------------------|----------------------|
| 2025-01-01 | 10,096,217.10 | 10,620,863.00 | 9,584,800.00 | 261,547,450,172.41 | 1,073,426,539,969.00 |
| 2025-02-01 | 10,063,414.85 | 10,730,688.00 | 8,903,600.00 | 1,345,170,486,280.52 | 3,338,250,559,744.00 |
| 2025-03-01 | 10,030,481.52 | 10,840,321.00 | 10,150,750.00 | 14,464,507,281.51 | 475,508,164,041.00 |
| | | | RMSE | 735,115.06 | 1,276,347.04 |

El pronóstico semanal mejora el RMSE en un 33% en comparación al pronóstico colaborativo semanal y el pronóstico mensual mejora el RMSE en un 42% en comparación al pronóstico colaborativo mensual. Tener un valor más preciso de predicción a corto y largo plazo ayuda a la planificación del material, reducir inventarios e incrementar la eficiencia de los sistemas de producción.

Para el material 900, no se tiene pronóstico colaborativo pero de igual forma se calculó el RMSE en comparación al valor real (ver *Tabla XX. Cálculo del RMSE del pronóstico semanal para el material 900* y *Tabla XXI. Cálculo del RMSE del pronóstico mensual para el material 900*) el cual queda aquí consignado como referencia.

Tabla XX. Cálculo del RMSE del pronóstico semanal para el material 900.

| Semana | Pronóstico | IC Inferior | IC Superior | Real | (Pron. - Real)**2 |
|------------|------------|-------------|-------------|--------------|----------------------|
| 2025-01-06 | 275133.67 | -15741 | 664292.33 | 0.00 | 0.00 |
| 2025-01-13 | 268304.79 | -22569.88 | 796196.79 | 0.00 | 0.00 |
| 2025-01-20 | 262248.28 | -28626.38 | 651406.95 | 253,000.00 | 64,009,000,000.00 |
| 2025-01-27 | 261323.22 | -29551.44 | 789215.22 | 220,000.00 | 48,400,000,000.00 |
| 2025-02-03 | 270034 | -20840.67 | 659192.67 | 710,000.00 | 504,100,000,000.00 |
| 2025-02-10 | 265035.13 | -25839.53 | 657662.13 | 200,000.00 | 40,000,000,000.00 |
| 2025-02-17 | 264536.3 | -38305.03 | 653694.97 | 0.00 | 0.00 |
| 2025-02-24 | 254354.18 | -48487.15 | 643512.85 | 290,000.00 | 84,100,000,000.00 |
| 2025-03-03 | 261165.99 | -29708.68 | 653792.99 | 1,000,000.00 | 1,000,000,000,000.00 |
| 2025-03-10 | 258238.19 | -32636.48 | 647396.85 | 280,000.00 | 78,400,000,000.00 |

| | | | | | |
|------------|-----------|-----------|-----------|-------------|--------------------|
| 2025-03-17 | 257512.79 | -45328.54 | 646671.46 | 0.00 | 0.00 |
| 2025-03-24 | 264429.88 | -38411.45 | 653588.55 | 922,000.00 | 850,084,000,000.00 |
| | | | | RMSE | 471,618.93 |

Tabla XXI. Cálculo del RMSE del pronóstico mensual para el material 900.

| Mes | Pronóstico | Real | (Pron. - Real)**2 |
|------------|-------------------|--------------|--------------------------|
| 2025-01 | 1,067,009.96 | 1,183,000.00 | 13,453,689,379 |
| 2025-02 | 1,053,959.61 | 1,490,000.00 | 190,131,221,711 |
| 2025-03 | 1,041,346.85 | 1,848,410.00 | 651,350,928,088 |
| | | RMSE | 533,833.88 |

8. Conclusiones

En esta segunda fase del proyecto se logró identificar y priorizar los 2 productos de mayor facturación utilizando análisis exploratorio de la serie de tiempo basados en datos semanales. A diferencia de la primera fase, hubo mayor cantidad de datos para analizar pero con la desventaja de tener mucha dispersión a través del tiempo. Sin embargo, el análisis de correlación demostró la existencia de una relación entre Ventas_Kg y Ventas_Pesos (en otras palabras entre el peso vendido y el valor monetario), lo que proporcionó una base útil para el desarrollo de modelos multivariados.

Se evaluaron modelos de promedio móvil, suavización exponencial y ARIMA para una serie de tiempo con datos sin imputar e imputación por media aritmética. A diferencia de la primera fase, en esta etapa de entrenamiento todos los modelos utilizaron optimización bayesiana, lo que permitió automatizar el proceso de ajuste de hiper parámetros para los dos productos priorizados. Este enfoque no solo redujo significativamente el tiempo requerido para el entrenamiento, sino que también optimizó el uso de recursos computacionales, facilitando así la posibilidad de escalar la solución a un mayor número de productos en el futuro.

Adicionalmente, se evaluó un modelo de vectores autorregresivos (VAR), incorporando las variables Ventas_Kg y Ventas_Pesos, con el objetivo de determinar si este enfoque multivariado podría ofrecer una mejora en la capacidad predictiva frente a los modelos univariados. Aunque el modelo VAR no superó en desempeño a las demás alternativas, su implementación resultó útil como ejercicio exploratorio y evaluar su efectividad en comparación a los demás modelos.

Finalmente, se realizó un análisis comparativo entre las proyecciones obtenidas por los mejores modelos (mensual y semanal) frente a la demanda colaborativa y las ventas reales. Este análisis reveló mejoras significativas en la precisión de las predicciones para el material 13, con incrementos de exactitud del 33% y 42%, respectivamente. Estos resultados indican que el mejor modelo desarrollado ofrece una estimación de la demanda más precisa y confiable, lo que representa un avance sustancial para la planificación y toma de decisiones comerciales.

9. Limitaciones

Disponibilidad: Los datos de variables exógenas (precio por material), dinámicas comerciales, No siempre se cuenta con suficientes datos históricos o relevantes.

Calidad de los datos: Datos incompletos, inconsistentes, duplicados o con errores afectan los resultados.

Tiempo de procesamiento: Algunos algoritmos pueden demorar mucho en ejecutarse si no están optimizados.

Herramientas limitadas: Restricción al uso de ciertas tecnologías por licencias, políticas corporativas o desconocimiento.

Integración con sistemas existentes: Dificultad para incorporar el modelo en los flujos operativos actuales de la empresa.

Acceso restringido: Por temas legales, de privacidad o confidencialidad, puede que no se tenga acceso a toda la información deseada.

10. Consideraciones Futuras

El análisis se enfoca en dos materiales específicos. Después, se podría ampliar el enfoque de modelado para incluir un mayor número de productos o grupos de productos que compartan características similares. Esto nos permitiría escalar la solución a nivel corporativo y mejorar la precisión de nuestras proyecciones.

Adicionalmente, se deberá fortalecer el modelo al incluir más variables externas que impactan la demanda, tales como:

- Promociones o campañas comerciales
- Calendario de producción
- Festivos o temporadas especiales
- Factores macroeconómicos o del mercado

Esto podría potenciar nuestra capacidad de predicción, especialmente en entornos inestables o para productos que tienen una fuerte estacionalidad.

Algo a tener en cuenta es que para asegurar que el modelo siga siendo preciso a lo largo del tiempo, es fundamental establecer un proceso regular de evaluación y reentrenamiento. Esto es especialmente importante si se observan diferencias significativas entre las predicciones y la demanda real,

Finalmente, como parte del desarrollo a futuro, se puede añadir un módulo de simulación de escenarios (optimista, pesimista, base) que permita evaluar cómo diferentes supuestos (como cambios en tendencias, aumento de la demanda o estacionalidades inusuales) impactan en las proyecciones y facilitan una planificación estratégica más efectiva.

11. Referencias

- [1] “©Copyright IBM Corporation 1994, Actualizado 2021-08-17, *Introducción al CRISP-DM*,
<https://www.ibm.com/docs/es/spss-modeler/saas?topic=guide-introduction-crisp-dm>”
- [2] *Figura 1. Esquema del ciclo CRISP-DM estándar*, BLOG: IIC / Innovación / La metodología CRISP-DM en ciencia de datos,
<https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/>
- [3] (Slimstock, Última actualización: February 22, 2024).
- [4] *Supply Chain Management*, Marzo 2020,
<https://cadenadesuministrosidat.blogspot.com/2020/03/supply-chain-management.html>
- [5] *Objetivos y Ventajas de un plan maestro de producción*,
<https://www.ekon.es/blog/plan-maestro-de-produccion/>, Equipo EKON, 30.05.2022
- [6] (¿Que es la gestión del inventario, IBM,
<https://www.ibm.com/es-es/topics/inventory-management>)
- [7] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. Wiley.
- [8] Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer.
- [9] Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*. Springer.
- [10] Gardner, E. S. (2006). *Exponential smoothing: The state of the art—Part II*. *International Journal of Forecasting*, 22(4), 637-666.
- [11] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
- [12] Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications*. Wiley.

[13] *Vectores autorregresivos. Utel Universidad.*

[14] *Alonso, J. Introducción a los pronósticos con modelos estadísticos de series de tiempo para científico de datos (en R) (2020).*