

Escuela de Negocios
Tipo de documento: Tesis de maestría



Master in Management + Analytics

Predicción de la Demanda mediante Modelos Estadísticos y de Machine Learning: Un Estudio Aplicado en la Industria del Vidrio

Autoría: Harsányi, Pablo

Año: 2025

¿Cómo citar este trabajo?

Harsányi, P. (2025) "*Predicción de la Demanda mediante Modelos Estadísticos y de Machine Learning: Un Estudio Aplicado en la Industria del Vidrio*". [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella
<https://repositorio.utdt.edu/handle/20.500.13098/13677>

El presente documento se encuentra alojado en el **Repositorio Digital de la Universidad Torcuato Di Tella** bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional
Dirección: <https://repositorio.utdt.edu>



UNIVERSIDAD
TORCUATO DI TELLA

Master in Management + Analytics

*Predicción de la Demanda mediante Modelos Estadísticos y
de Machine Learning: Un Estudio Aplicado en la Industria
del Vidrio*

Francisco Pablo Harsányi

Tutor: Luigi Laporte

Resumen

La predicción precisa de la demanda constituye un elemento estratégico para optimizar la planificación operativa y la asignación eficiente de recursos

en entornos industriales. Este trabajo desarrolla un análisis comparativo de distintos enfoques de pronóstico aplicados a datos reales de ventas de una empresa líder en la industria del vidrio. Se implementaron modelos estadísticos tradicionales (ARIMA) y técnicas avanzadas de machine learning (Random Forest, XGBoost, Prophet y redes neuronales LSTM), evaluados en dos niveles de agregación: SKU-mes y demanda total mensual.

La metodología de evaluación combinó métricas absolutas y relativas (RMSE, WAPE, R^2), respetando la estructura temporal de los datos para preservar la validez predictiva. Los resultados evidencian una marcada superioridad de los modelos basados en machine learning, destacándose XGBoost con errores ponderados inferiores al 3% y niveles de ajuste superiores al 95% en predicción agregada.

El estudio resalta además la importancia de contar con predicciones desagregadas por producto, no solo para maximizar el nivel de servicio y minimizar inventarios, sino también para mejorar la flexibilidad operativa ante variaciones en la demanda. Finalmente, se discuten las implicancias prácticas de los hallazgos y se proponen líneas de investigación futuras para continuar fortaleciendo las capacidades predictivas en contextos industriales dinámicos.

Índice

1. Introducción

1.1. Contexto	6
1.2. Planteamiento del problema	8
1.3. Objetivos de investigación	9

2. Revisión de la literatura 10

2.1. Importancia del pronóstico de demanda en la industria del vidrio	10
2.2. Métodos de pronóstico de demanda	11
2.2.1. Series de tiempo	11
2.2.2. Métodos informales y promedios simples	13
2.2.3. Modelos de regresión	19
2.2.4. Modelos ARIMA / SARIMA y la metodología Box-Jenkins	24
2.2.5. Modelos aditivos: Prophet	27
2.2.5. Modelos de machine learning para series temporales	28
2.2.5.1. Random Forest	28
2.2.5.2. XGBoost	29
2.2.5.3. Redes neuronales y modelos avanzados de aprendizaje profundo	30

3. Datos 31

3.1. Estructura del dataset	31
3.1.1. Fuentes de datos	31
3.1.2. Variables relevantes	34
3.2. Análisis exploratorio de datos	36
3.2.1. Análisis de la demanda por SKU y cliente	37
3.2.2. Análisis por rubro	39
3.2.3. Identificación de patrones estacionales y de tendencia	39
3.2.4. Dispersión y volatilidad	43
3.2.5. Importancia empírica de variables predictoras	44

4. Metodología	45
4.1. Nivel y horizonte del pronóstico	45
4.1.1. Nivel: SKU–mes y total mensual	45
4.1.2. Horizonte de pronóstico: octubre–diciembre 2024	46
4.2. Preparación de los datos	47
4.3. Selección y descripción de los modelos predictivos	49
4.3.1. Regresión lineal múltiple	49
4.3.2. Modelos de series de tiempo: ARIMA y SARIMA	51
4.3.2.1. Aplicación en la serie mensual agregada	51
4.3.2.2. Aplicación a series individuales por SKU	55
4.3.3. Random Forest	56
4.3.4. XGBoost	58
4.3.5. Prophet	59
4.3.6. Modelo LSTM	60
4.4. Entrenamiento y validación	60
4.4.1. División temporal de los datos	61
4.4.2. Validación específica por tipo de modelo	61
4.4.3. Criterios de selección y precauciones metodológicas	63
4.5. Evaluación de desempeño	64
4.5.1. Métricas de evaluación (RMSE, MAPE, WAPE, R ²)	65
4.5.2. Niveles de agregación utilizados	66
5. Resultados	67
5.1. Enfoque de evaluación	68
5.2. Resultados por modelo	68
5.2.1. Modelos ARIMA y SARIMA	64
5.2.2. Random Forest	70
5.2.3. XGBoost	72
5.2.4. Prophet	73
5.2.5. Redes neuronales LSTM	74
5.3. Comparación entre modelos	75

5.4. Análisis SKU–mes con XGBoost	77
5.5. Reflexiones sobre los resultados	80
6. Conclusiones	81
6.1. Principales hallazgos	81
6.2. Limitaciones del estudio.....	83
6.3. Líneas futuras de investigación	84
Anexo	86
Bibliografía	91

I. Introducción

1. Contexto

En los negocios, los pronósticos juegan un papel fundamental en la planificación operativa, financiera y logística de una empresa, siendo un aspecto clave para la optimización de los costos.

La industria del vidrio es un claro: el gran volumen de capital involucrado, los enormes costos fijos y las distintas restricciones operativas -como poca flexibilidad en los volúmenes de producción o la necesidad de estar constantemente produciendo- hacen que tener un buen plan sea clave para el éxito.

La presente tesis busca enfrentarse a los desafíos de esta industria en relación al pronóstico de la demanda, específicamente para Vidriotec¹, la principal fábrica de vidrio de Argentina. Con una participación de mercado casi monopólica, su producción se centra en abastecer a los principales jugadores de la industria alimentaria de la región y, en segundo plano, producen productos para el hogar.

Se trata de una empresa del grupo Cristanox -que también produce para la industria farmacéutica- con ubicación en Zárate y más de 800 empleados que producen diariamente miles de envases para la industria alimenticia de productos de consumo masivo; como bebidas, productos de almacén o artículos de bazar. Más del 30% de su demanda se centra en grandes empresas de la industria, tales como los clientes Q, N, A y M, entre otros.

El contexto en el que se desenvuelve Vidriotec, sumado a las características propias de la industria del vidrio, hacen que se enfrente periódicamente a un gran volumen de demanda y variabilidad debido a la fuerte influencia que tienen factores económicos, estacionales y de consumo. Estos desafíos pueden resumirse en los siguientes factores:

- Estacionalidad de la demanda: productos como las botellas de bebidas pueden enfrentar ciertos picos en meses próximos al verano y caídas en el invierno
- Factores macroeconómicos: influyen de manera directa en la demanda de sus clientes, que se ve afectada por el contexto económico, la inflación y cambios en las políticas de importación/ exportación"

¹ Por motivos de confidencialidad, el nombre de la empresa ha sido reemplazado por uno ficticio.

- Tendencias del mercado: cambios en las preferencias de los consumidores finales o regulaciones medioambientales pueden afectar los patrones históricos de compras.

Es importante resaltar que los diferentes productos no experimentan la misma dinámica de demanda. Por ejemplo, los envases de cerveza pueden presentar una mayor estacionalidad debido a su alto consumo en verano, mientras que los productos destinados a la industria alimentaria pueden estar más influenciados por condiciones macroeconómicas, como el poder adquisitivo de los consumidores. Esta diferenciación en los patrones de demanda es un aspecto clave a analizar, ya que implica que un único modelo de pronóstico podría no ser suficiente para capturar con precisión estas variaciones.

En este contexto con varias fuentes de variabilidad, contar con modelos de pronósticos precisos de la demanda es clave para planificar una producción acorde a la demanda que al mismo tiempo ayude a optimizar la gestión de inventarios, como también para ayudar a la gerencia a tomar mejores decisiones estratégicas sustentadas en datos.

En este trabajo, se define un horizonte de pronóstico mensual con una anticipación de hasta tres meses, es decir, busca estimar la demanda mensual con uno, dos o tres meses de antelación. Este horizonte fue seleccionado en función de las necesidades operativas de Vidriotec, ya que la planificación de la producción requiere previsibilidad para tomar decisiones con tiempo, debido a las restricciones técnicas del proceso y sus altos costos asociados.

Por otra parte, se evaluarán pronósticos a nivel desagregado, al nivel de SKU por cliente. Este nivel de detalle resulta necesario para capturar los diferentes comportamientos de consumo entre productos y segmentos, y así permitir una planificación más ajustada a la realidad operativa. En particular, el análisis de esta tesis se enfoca en los 32 productos más relevantes para el negocio, definidos por la propia compañía en función de su peso en el volumen total y su impacto estratégico en la operación.

2. Problema

El problema a abordar es desarrollar herramientas cuantitativas que permitan mejorar la planificación de la producción, la cual depende en gran medida de la precisión en los pronósticos de la demanda. Actualmente, la empresa proyecta su producción a un horizonte de 3 a 6 meses según la demanda estimada, basándose principalmente en los forecasts provistos por sus clientes principales. Sin embargo, este enfoque presenta limitaciones, ya que no siempre captura de manera precisa las fluctuaciones de la demanda real, especialmente en aquellos segmentos donde los factores de variabilidad tienen un impacto más significativo.

La predicción de la demanda es un pilar fundamental en la gestión eficiente de la producción y los inventarios. Según Hanke (2010), un pronóstico preciso permite reducir costos operativos, mejorar la toma de decisiones y optimizar la asignación de recursos en la cadena de suministro. No obstante, en este caso, la dependencia de los forecasts de los clientes y la falta de modelos predictivos avanzados generan incertidumbre en la planificación, obligando a la empresa a mantener altos niveles de inventario para compensar la variabilidad en la demanda.

Esta variabilidad, sin embargo, no es homogénea en todos los productos. La fuente mencionada previamente indica que la demanda de ciertos productos puede estar influenciada por patrones estacionales, tendencias de mercado o factores macroeconómicos, lo que hace que su predicción sea más compleja. En el caso analizado, los envases de cerveza pueden experimentar fluctuaciones más marcadas debido a la estacionalidad y la competencia con productos sustitutos, como las latas de aluminio. En contraste, los productos destinados a la industria alimentaria (posición casi monopólica de Vidriotec), pueden mostrar una demanda más estable, influenciada principalmente por el poder adquisitivo del consumidor y el contexto macroeconómico. Estas diferencias en los patrones de demanda hacen necesario un análisis más profundo para determinar si diferentes tipos de productos requieren enfoques de modelado diferenciados en lugar de una estrategia de pronóstico única.

La empresa enfrenta diariamente varios desafíos en cuanto a la gestión de la demanda y sus decisiones respecto a la planificación de la producción:

- Dificultades para anticipar cambios en la demanda: Aunque los clientes más representativos proporcionan sus forecasts de consumo, estos no siempre reflejan la realidad del mercado, provocando desajustes en la producción.
- Riesgo de sobre-stock y desabastecimiento: La acumulación innecesaria de inventario en períodos de baja demanda impacta los costos operativos, desde almacenaje hasta costos de oportunidad por no producir artículos de mayor demanda en un momento dado. Esto repercute en el nivel de servicio a los clientes, especialmente en momentos de alta demanda. Hanke (2010) sostiene que una planificación basada en pronósticos imprecisos puede derivar en costos elevados debido a la acumulación innecesaria de stock o a la incapacidad de satisfacer la demanda en el momento adecuado.
- Complejidad en la programación de la producción: Las restricciones propias del proceso, como la imposibilidad de detener los hornos y la necesidad de ajustes específicos en las líneas de producción, limitan la flexibilidad para reaccionar ante cambios inesperados en la demanda.

Dado este contexto, la falta de un modelo preciso genera ineficiencias en la planificación de la producción y la gestión de inventarios, aumentando los costos operativos y el riesgo de desabastecimiento o sobre-stock. Esta situación impacta la capacidad de respuesta de la empresa ante variaciones en la demanda y dificulta la optimización de los recursos productivos, especialmente en productos estratégicos con alta rotación y demanda fluctuante.

3. Objetivo de investigación

El objetivo es desarrollar modelos cuantitativos de pronóstico de la demanda para Vidriotec, con el propósito de proporcionar a la gerencia una herramienta efectiva que permita prever la demanda futura de los SKU con mayor impacto en el negocio. Específicamente, se buscará predecir la demanda de productos para un horizonte de tiempo de uno a tres meses, contribuyendo a una planificación más precisa de la producción, una gestión más eficiente del inventario y un mayor nivel de servicio de cara al cliente.

Este trabajo compara distintas metodologías de predicción, incluyendo modelos estadísticos tradicionales (ARIMA, suavización exponencial) y técnicas de aprendizaje automático (Random Forest, XGBoost, redes neuronales). Dado que diferentes técnicas de pronóstico

pueden arrojar resultados variables según el contexto y los datos disponibles, es necesario evaluar múltiples enfoques (Masini & Vázquez, 2014).

Además, se analiza si los patrones de demanda varían significativamente entre diferentes tipos de productos, como envases para cerveza y aquellos destinados a la industria alimentaria. Este análisis permite determinar si se requiere una estrategia diferenciada de modelado para cada categoría, en lugar de aplicar un enfoque único para toda la producción de Vidriotec. Identificar estas diferencias contribuye al desarrollo de modelos más precisos y ajustados a las particularidades de cada segmento, optimizando la planificación de la producción y la gestión de inventarios.

Se trata de una investigación de impacto binario: brinda a la gerencia una herramienta basada en modelos cuantitativos, para facilitar la toma de decisiones en la planificación de la producción; y, a través de un análisis diferenciado de la demanda por tipo de producto, establece la necesidad de aplicar estrategias de modelado específicas para cada categoría, lo que contribuirá a reducir costos asociados al almacenamiento, minimizar riesgos de desabastecimiento y garantizar un alto nivel de servicio sin necesidad de mantener inventarios excesivos.

2. Revisión de literatura

2.1 Importancia del Pronóstico de Demanda en la Industria del Vidrio

De acuerdo con Hanke y Wichern (2009), una predicción precisa de la demanda, permite optimizar los niveles de inventario, reducir costos operativos y mejorar la eficiencia en la asignación de recursos.

En el caso de Vidriotec, la ausencia de modelos de pronóstico avanzados ha llevado a fluctuaciones en los niveles de stock, generando tanto sobrecostos de almacenamiento como riesgos de desabastecimiento en productos de alta demanda. La implementación de técnicas cuantitativas permitiría reducir esta incertidumbre y fortalecer la toma de decisiones en la planificación de la producción.

A continuación, se presenta una revisión de los principales enfoques para el pronóstico de demanda, desde métodos cualitativos hasta modelos cuantitativos avanzados.

2.2 Métodos de Pronóstico de Demanda

Los métodos utilizados para pronosticar demanda han evolucionado significativamente en las últimas décadas. Se ha pasado de enfoques cualitativos, basados en la experiencia y el juicio, a modelos cuantitativos apoyados en estadística y, más recientemente, en algoritmos de aprendizaje automático. Esta evolución ha sido impulsada por la necesidad de mejorar la precisión, la escalabilidad y la capacidad de adaptación ante contextos cada vez más dinámicos.

2.2.1. Series de Tiempo

Uno de los enfoques cuantitativos más difundidos en la predicción de demanda es el análisis de series de tiempo, que se basan en el análisis de observaciones de una variable (Y) recogidas secuencialmente en el tiempo. A diferencia de una secuencia aleatoria de datos, las series temporales presentan estructuras internas de dependencia entre observaciones cercanas, lo que requiere métodos específicos de análisis y modelado.

Esta relación entre las observaciones sucesivas da lugar a patrones de variabilidad que pueden ser aprovechados para estimar valores futuros. En otras palabras, se busca identificar regularidades en el comportamiento pasado que resulten útiles para anticipar la evolución futura de la serie. Sin embargo, dado que estas relaciones pueden cambiar con el tiempo, es fundamental reevaluarlas continuamente para validar su vigencia.

Descomposición de Series Temporales

Uno de los métodos más utilizados para comprender la estructura interna de una serie de tiempo es la descomposición, que permite desagregar la serie en distintos componentes con el objetivo de analizar su influencia individual y conjunta, tanto en el corto como en el largo plazo. Es importante aclarar que la descomposición no constituye en sí un modelo de

pronóstico, sino una herramienta de análisis que facilita la comprensión y el tratamiento posterior de la serie.

A partir de este enfoque, se pueden identificar los siguientes cuatro componentes básicos:

1) **Tendencia:** Corresponde a los movimientos de largo plazo en la serie, que reflejan un crecimiento o disminución sostenida a lo largo del tiempo. La tendencia puede tener forma lineal o no lineal, y su estimación implica ajustar una curva que represente la dirección general de los datos.

Una forma común de modelar la tendencia es a través de regresión, donde se busca estimar una ecuación del tipo:

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_k t^k$$

El objetivo es encontrar los coeficientes β_i que minimicen una función de error (por ejemplo, mínimos cuadrados), de modo que la estimación de la tendencia se acerque lo más posible a los valores reales observados Y_t .

2) **Estacionalidad:** Se refiere a patrones sistemáticos que se repiten a intervalos regulares, como pueden ser ciclos anuales, trimestrales o mensuales. Para aislar el componente estacional, es necesario seguir una secuencia de pasos:

1. Estimar y eliminar la tendencia de la serie original.
2. Suavizar el componente irregular o aleatorio.

Recolectar y resumir los valores estacionales para construir un índice estacional para cada período del ciclo.

Este análisis permite no solo identificar el patrón estacional, sino también eliminarlo (en caso de que se desee trabajar con series ajustadas), pronosticar valores futuros considerando la estacionalidad y tomar decisiones operativas como la programación de producción.

3) **Ciclo:** El componente cíclico refleja fluctuaciones que no tienen un patrón regular ni un período fijo, como expansiones y contracciones relacionadas con el contexto económico general. A diferencia de la estacionalidad, los ciclos pueden durar varios años y su identificación es más compleja, ya que no responden a un calendario. Por esta razón, en la práctica, el componente cíclico suele considerarse como parte de la tendencia.

Para poder analizarlo, es necesario previamente haber eliminado la tendencia y la estacionalidad de la serie.

4) **Componente Irregular:** Representa variaciones impredecibles que no se explican por ninguno de los componentes anteriores. Está asociado a eventos fortuitos, errores de medición u otros factores no sistemáticos.

Según Hanke y Wichern (2009), la descomposición de una serie temporal permite separar y analizar el comportamiento individual de sus componentes, así como generar combinaciones de los mismos con fines descriptivos o predictivos. Esto resulta útil para mejorar la interpretación de la serie, preparar los datos para modelado posterior o construir pronósticos específicos a partir de componentes seleccionados.

2.2.2 Métodos Informales y Promedios Simples

Antes de introducir modelos estadísticos más complejos, resulta útil repasar los enfoques informales y los métodos simples que muchas organizaciones—incluida la empresa objeto de análisis— aún utilizan debido a su simplicidad y bajo requerimiento técnico. Si bien estos enfoques pueden ofrecer cierto valor inicial, al depender del juicio humano presentan limitaciones importantes en términos de replicabilidad y objetividad.

Entre los métodos informales más comunes se destacan:

- **Juicio de expertos:** basado en la opinión de personas con conocimiento del mercado, como gerentes de ventas o responsables de operaciones. Aunque valioso, este enfoque es susceptible a sesgos cognitivos y errores sistemáticos.

- Forecasts proporcionado por clientes: muchas empresas toman como insumo los pronósticos de sus clientes clave. Si bien esta información puede reflejar intenciones reales de compra, también puede ser inestable y poco confiable.
- Analogía histórica: se basa en identificar situaciones previas similares y asumir que las tendencias se repetirán. La efectividad de este método depende en gran medida del grado de comparabilidad entre contextos pasados y actuales.

Aunque estos métodos pueden ofrecer valor inicial, su falta de estructura y rigurosidad motivó la adopción de técnicas más objetivas, comenzando por los modelos de promedio.

Los promedios simples son una técnica básica ampliamente utilizada por su facilidad de implementación. Como mencionan Hanke y Wichern (2009), este enfoque es adecuado “cuando los factores que producen la serie se han estabilizado y el entorno permanece sin cambios”. Es decir, cuando la demanda no está sujeta a grandes variaciones ni se espera una tendencia clara.

Este tipo de técnica resulta útil cuando las variables que afectan la serie temporal permanecen relativamente estables, y su comportamiento no sugiere cambios estructurales importantes en el corto plazo.

Sin embargo, en escenarios donde existen fluctuaciones, estacionalidad o tendencias, estos modelos pierden precisión, dando lugar a alternativas más refinadas como los promedios móviles y la suavización exponencial.

- Promedios Móviles: Este método calcula el promedio de los n valores más recientes de la serie temporal y lo utiliza como pronóstico para el siguiente período:

$$\hat{Y}_{t+1} = \frac{1}{n} \sum_{i=t-n+1}^t Y_i$$

donde:

- \hat{Y}_{t+1} = pronóstico para el período siguiente
- n = cantidad de períodos incluidos en el promedio

- Y_i = valor real observado en el período i

El promedio móvil asigna igual peso a todas las observaciones recientes, y su sensibilidad ante cambios en la demanda está determinada por el parámetro n : un valor bajo hace que el modelo sea más reactivo pero más volátil, mientras que un n alto lo estabiliza, aunque a costa de mayor rezago ante cambios abruptos.

Lo que hace esta técnica es calcular la media de las observaciones más recientes dentro de los n períodos definidos. Al asignar pesos iguales, su capacidad para adaptarse a tendencias o estacionalidades es limitada. Aunque mejora respecto al promedio simple, aún no permite capturar adecuadamente patrones estructurales si estos son marcados o variables.

- Suavización Exponencial: La suavización exponencial simple es una evolución del promedio móvil que otorga mayor peso a los datos más recientes, utilizando un coeficiente de suavización α (donde $0 < \alpha < 1$). Su fórmula principal es:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t$$

o, de forma equivalente:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t)$$

donde:

- \hat{Y}_t = pronóstico para el período t
- Y = valor real observado en el período anterior
- \hat{Y}_{t-1} = pronóstico del período anterior
- α = coeficiente de suavización (entre 0 y 1)

Además $(Y - \hat{Y}_t)$ representa el error de pronóstico del período anterior. Esta segunda forma permite visualizar que el modelo ajusta su estimación anterior en función del error observado, ponderado por α .

Este método asigna un peso decreciente en el tiempo a las observaciones pasadas, lo cual lo hace más sensible a los cambios recientes. El valor de α define qué tan reactivo será el modelo: valores cercanos a 1 hacen que el modelo responda con rapidez, mientras que valores bajos lo suavizan y estabilizan.

La selección óptima del parámetro α se realiza de forma iterativa, buscando minimizar el error cuadrático medio (MSE). Un concepto clave en este proceso es el uso de la señal de control, una técnica que permite monitorear la media de los errores a lo largo del tiempo. Esta señal sirve para detectar si el modelo necesita ser ajustado o si debe reconsiderarse el valor de α . Idealmente, un modelo balanceado debería sobrestimar y subestimar la demanda con la misma frecuencia, manteniéndose dentro de márgenes de desviación aceptables.

La suavización exponencial funciona bien cuando la serie temporal no presenta grandes saltos ni tendencia sostenida. En caso contrario, el modelo tenderá a "quedarse atrás" respecto a los valores reales, reaccionando lentamente ante cambios estructurales.

Si bien los métodos simples de suavización exponencial funcionan razonablemente bien en escenarios con demanda relativamente estable, su capacidad se ve limitada cuando los datos presentan tendencias marcadas o estacionalidad. Para abordar estas limitaciones, se han desarrollado extensiones dentro de la misma familia de modelos, como el método de Holt y el método de Winters. Estas variantes permiten una mayor adaptabilidad en contextos donde la demanda no es constante, sino que muestra patrones crecientes y/o cíclicos.

Método de Holt

El método de Holt, también conocido como suavización exponencial doble, fue diseñado para series temporales que exhiben una tendencia lineal sostenida. A diferencia de la suavización simple, este modelo suaviza tanto el nivel como la pendiente (tendencia) de la serie a lo largo del tiempo. Para ello, utiliza dos constantes de suavización: α para el nivel y β para la tendencia.

Las ecuaciones que definen este modelo son:

- Estimación de nivel:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

- Estimación de la tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

- Pronóstico p períodos delante:

$$\hat{Y}_{t+p} = L_t + pT_t$$

donde:

- L_t = nivel estimado de la serie en t
- T_t = tendencia estimada
- \hat{Y}_{t+h} = pronóstico h períodos hacia adelante
- α, β = coeficientes de suavización para nivel y tendencia respectivamente
- h = horizonte de pronóstico

Este modelo permite anticipar movimientos futuros cuando los datos presentan una tendencia clara. Como señalan Hanke y Wichern (2009), una posible desventaja es que requiere que la tendencia sea aproximadamente lineal, lo cual no siempre es habitual en entornos empresariales. Al igual que en la suavización simple, las constantes α y β se pueden optimizar mediante la minimización del error cuadrático medio (MSE), y su magnitud define cuánto sigue el modelo a los datos recientes.

Método de Winters

El método de Winters, también llamado suavización exponencial triple, extiende el modelo de Holt al incorporar un componente adicional: la estacionalidad. Esta variante es

especialmente útil cuando la serie presenta tanto una tendencia como un patrón estacional sistemático.

En su formulación multiplicativa —recomendada cuando la estacionalidad es proporcional al nivel—, se utilizan cuatro ecuaciones:

- Estimación del nivel:

$$L_t = \alpha \left(\frac{Y_t}{S_{t-s}} \right) + (1 - \alpha)(L_{t-1} + T_{t-1})$$

- Estimación de la tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

- Estimación del componente estacional:

$$S_t = \gamma \left(\frac{Y_t}{L_t} \right) + (1 - \gamma)S_{t-s}$$

- Pronóstico p períodos adelante:

$$\hat{Y}_{t+p} = (L_t + pT_t) \cdot S_{t-s+p}$$

donde:

- L_t : nivel suavizado en el tiempo t
- T_t : tendencia estimada
- S_t : componente estacional
- s : longitud del ciclo estacional
- p : horizonte de pronóstico

Tal como en los métodos anteriores, los parámetros α , β y γ pueden ajustarse para minimizar el MSE. La elección entre versión aditiva o multiplicativa dependerá del tipo de estacionalidad observada en los datos. Según Hanke y Wichern (2009), este modelo permite

capturar con eficacia la dinámica de series con patrones repetitivos, manteniendo una estructura relativamente simple y transparente para la interpretación.

Si bien estos métodos pueden resultar útiles para pronósticos de corto plazo y en contextos con baja variabilidad, su aplicabilidad se ve limitada cuando la serie temporal presenta estacionalidad, tendencia o cambios estructurales. Además, suelen requerir cierta estabilidad en las condiciones del entorno para generar estimaciones confiables.

En su implementación, los primeros valores de la serie suelen utilizarse para inicializar el modelo, mientras que los datos posteriores se reservan para evaluar su desempeño. Esta lógica anticipa lo que ocurre en enfoques más avanzados, donde la base de datos se divide explícitamente en conjuntos de entrenamiento, validación y prueba, con el objetivo de optimizar parámetros y contrastar resultados en distintos escenarios.

La necesidad de capturar patrones más complejos en los datos, así como de incorporar variables explicativas y adaptarse a entornos dinámicos, impulsó el desarrollo de modelos más sofisticados. Entre ellos se encuentran los modelos de regresión, las técnicas clásicas de series de tiempo y los enfoques de machine learning, que ofrecen mayor flexibilidad y precisión en la predicción de demanda.

2.2.3. Modelos de Regresión

Los modelos de regresión constituyen una herramienta estadística fundamental en el análisis de relaciones entre variables. En el contexto de la predicción de demanda, permiten estimar el valor de una variable dependiente a partir de una o más variables independientes, explicando parte de la variabilidad observada en los datos mediante una estructura lineal. Esta sección aborda la regresión lineal simple, la regresión múltiple y los supuestos generales que comparten ambos enfoques, con énfasis en su interpretación, limitaciones y condiciones de validez.

Regresión Lineal Simple

El modelo de regresión lineal simple busca establecer una relación lineal entre una variable dependiente Y y una única variable independiente X . Esta relación se representa mediante

una recta cuya pendiente indica el efecto estimado del cambio en X sobre Y , bajo el supuesto de una relación lineal constante.

El objetivo del modelo es encontrar la línea que mejor se ajusta a los datos observados, minimizando la suma de los cuadrados de las diferencias entre los valores reales de Y y los valores estimados \hat{Y} . Este criterio se conoce como mínimos cuadrados ordinarios (MCO) y su función objetivo se expresa como:

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Donde $\hat{Y} = b_0 + b_1X_i$ es la estimación del valor medio de Y para un valor dado de X , b_0 es la ordenada al origen y b_1 la pendiente de la recta.

La diferencia entre el valor observado Y y su estimación \hat{Y} se conoce como residuo, denotado como $e = Y - \hat{Y}$, esta descomposición también puede expresarse como $Y = \hat{Y} + e$.

En este contexto, \hat{Y} representa el patrón estimado de los datos y e la desviación aleatoria, lo que permite separar el comportamiento sistemático del ruido o error.

Desde una perspectiva probabilística, se asume que para cada valor de X , los valores de Y están distribuidos normalmente alrededor de una media $\mu_Y = \beta_0 + \beta_1X$. Esta línea teórica se denomina línea de regresión de la población, y el modelo estadístico se expresa como:

$$Y = \beta_0 + \beta_1X + e$$

donde e representa el término de error aleatorio.

Regresión Lineal Múltiple

En la práctica, la predicción de una variable dependiente a menudo requiere incorporar más de una variable explicativa. En estos casos, se utiliza la regresión lineal múltiple, que extiende la lógica de la regresión simple a un modelo con K variables independientes:

$$Y = \beta_0 + \beta_1X + \beta_2X_2 + \dots + \beta_KX_K + \varepsilon$$

La estimación de los parámetros del modelo también se realiza mediante el criterio de mínimos cuadrados, obteniendo así la función de regresión estimada:

$$\hat{Y} = b_0 + b_1X + b_2X_2 + \dots + b_KX_K + \varepsilon$$

Al igual que en el modelo simple, se conserva la identidad fundamental:

$$Y = \hat{Y} + e$$

donde \hat{Y} representa la parte explicada por el modelo y e el residuo.

El modelo múltiple permite evaluar simultáneamente el efecto individual de cada variable independiente sobre Y , manteniendo constante el efecto de las demás. Este enfoque mejora el poder explicativo y la precisión de los pronósticos, aunque también introduce desafíos adicionales, como la posibilidad de multicolinealidad entre variables.

Conceptos Comunes: Supuestos y Evaluación del Ajuste

Tanto en la regresión simple como en la múltiple, los modelos se basan en una serie de supuestos clave, cuyo cumplimiento es necesario para garantizar la validez de las inferencias estadísticas:

1. Linealidad: la relación entre las variables independientes y la dependiente es lineal.
2. Independencia de los errores: los residuos deben ser independientes entre sí.
3. Homocedasticidad: los errores tienen varianza constante a lo largo del rango de valores de las variables independientes.
4. Normalidad de los errores: los errores se distribuyen normalmente (particularmente importante para la inferencia).
5. Ausencia de multicolinealidad (solo en regresión múltiple): las variables independientes no deben estar altamente correlacionadas entre sí.

De todos ellos, la independencia de los errores resulta especialmente crítica en contextos donde los datos provienen de series de tiempo, ya que la autocorrelación puede sesgar los resultados. Este punto será retomado en el siguiente apartado, dedicado específicamente a modelos de regresión aplicados a datos temporales.

Otro concepto esencial compartido por ambos modelos es el coeficiente de determinación R^2 , que mide el porcentaje de variación total de Y explicado por el modelo. Se obtiene a partir de la descomposición de la varianza:

$$SST = SSR + SSE$$

donde:

- SST: suma total de cuadrados ($\sum (Y - \bar{Y})^2$)
- SSR: suma total de cuadrados explicada por la regresión ($\sum (\hat{Y} - \bar{Y})^2$)
- SSE: suma de cuadrados del error ($\sum (Y - \hat{Y})^2$)

El coeficiente R^2 se define como:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Cuanto más cercano a 1, mayor es la proporción de la variabilidad de Y explicada por el modelo. Si se extrae la raíz cuadrada de R^2 en el caso de regresión simple, se obtiene el coeficiente de correlación lineal r , que mide la fuerza y dirección de la relación lineal entre las variables.

Limitaciones de la regresión en datos temporales y transición a modelos dinámicos

Uno de los supuestos fundamentales de los modelos de regresión lineal, tanto simple como múltiple, es la independencia entre los errores. Este supuesto garantiza que los residuos del modelo no estén correlacionados entre sí, lo cual es esencial para asegurar la validez de las inferencias estadísticas. Sin embargo, en el análisis de datos que varían en el tiempo, este supuesto rara vez se cumple. Las observaciones recogidas en secuencia suelen estar influenciadas por fenómenos subyacentes que generan dependencia temporal entre los datos.

Esta dependencia se conoce como autocorrelación, e implica que los valores sucesivos de la variable dependiente y/o de sus errores tienden a estar relacionados entre sí. Una forma común de autocorrelación es la de primer orden, donde el error en un período determinado

depende linealmente del error del período anterior. Este comportamiento puede expresarse como:

$$e_t = \rho e_{t-1} + v_t$$

donde e_t representa el error en el período t , ρ es el coeficiente de autocorrelación de primer orden, y v_t un término de error aleatorio e independiente. Un valor de ρ distinto de cero implica que los residuos del modelo no son independientes, lo que invalida uno de los supuestos esenciales del enfoque clásico de regresión.

Para detectar la presencia de autocorrelación existen diversos procedimientos estadísticos. Entre los más utilizados se encuentra el estadístico de Durbin-Watson, que permite evaluar si existe autocorrelación de primer orden en los residuos. Este se calcula como:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

Un valor cercano a 2 indica ausencia de autocorrelación, mientras que valores significativamente inferiores sugieren autocorrelación positiva. Otra herramienta complementaria es el estadístico Q de Ljung-Box, que evalúa la autocorrelación conjunta de los residuos en múltiples rezagos:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{r}_k^2}{n-k}$$

Donde r^k es el coeficiente de autocorrelación muestral en el rezago k , n es el tamaño de la muestra y h el número de rezagos considerados.

Aunque la autocorrelación puede generar estimaciones ineficientes y errores estándar incorrectos, también puede utilizarse a favor si se modela adecuadamente. En lugar de aplicar una regresión tradicional, se puede considerar un enfoque en el que se utilicen como variables explicativas los valores pasados de la propia variable dependiente. Esto da lugar a los denominados modelos autorregresivos (AR), cuya formulación general de orden k es:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_k Y_{t-k} + \varepsilon_t$$

donde:

- Y_t : valor de la serie en el tiempo t .
- Φ_i : coeficientes autorregresivos
- ε_t : error aleatorio en el período t .

En este tipo de modelos, el valor actual de la serie se estima a partir de sus propios valores rezagados, aprovechando explícitamente la autocorrelación observada. Estos modelos representan una transición natural hacia estructuras dinámicas más robustas, como los modelos ARIMA, que se explorarán en la sección siguiente.

En definitiva, la presencia de autocorrelación en los residuos de una regresión aplicada a datos temporales puede llevar a una sobreestimación del coeficiente de determinación R^2 y a una subestimación de los errores estándar de los coeficientes. Una de las causas frecuentes de este problema es la omisión de variables explicativas relevantes, lo cual deja una porción significativa de la variabilidad sin modelar. Frente a esto, una estrategia adecuada es revisar qué variables podrían estar faltando en el modelo o, alternativamente, optar por modelos específicamente diseñados para captar la dinámica temporal de los datos, como los autorregresivos.

2.2.4 Modelos ARIMA / SARIMA y la metodología Box-Jenkins

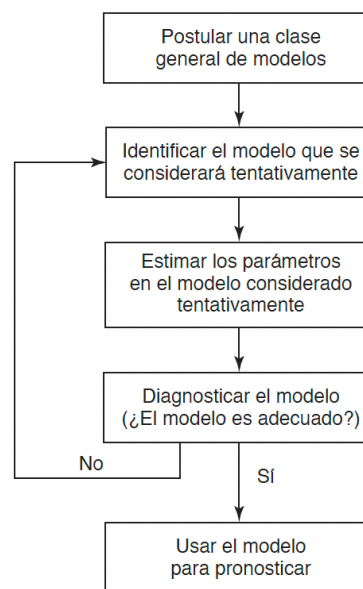
Los modelos ARIMA (Autoregressive Integrated Moving Average) constituyen una de las herramientas más potentes y flexibles para la predicción de series de tiempo. A diferencia de los modelos de regresión tradicionales, estos modelos no requieren de variables independientes para realizar pronósticos, sino que utilizan exclusivamente la información histórica de la propia serie para estimar su comportamiento futuro. Esta característica los vuelve particularmente útiles en contextos donde no se dispone de buenos predictores externos, o donde la dinámica de la variable en sí contiene suficiente información. Como afirman Hyndman y Athanasopoulos (2018), “los modelos ARIMA se basan en la hipótesis de que los valores pasados y los errores previos de la serie contienen suficiente información para pronosticar su comportamiento futuro”.

Además, estos modelos son especialmente valiosos cuando no se dispone de variables explicativas relevantes, ya que permiten capturar patrones internos de la serie, incluso cuando esta no es estacionaria (Makridakis, Wheelwright & Hyndman, 1998). A diferencia de los modelos de regresión, ARIMA no requiere suposiciones sobre la distribución de los errores ni relaciones lineales entre variables independientes y la dependiente, lo que amplía su aplicabilidad en entornos donde los supuestos clásicos no se cumplen (Hanke & Wichern, 2009).

El procedimiento más utilizado para construir modelos de esta familia es el propuesto por Box y Jenkins, conocido como metodología Box-Jenkins, que contempla tres fases iterativas: identificación del modelo, estimación de parámetros y verificación del ajuste. Este enfoque tiene un carácter exploratorio y progresivo: se propone una clase general de modelos, se ajusta un candidato inicial, se evalúan los residuos, y si el modelo no resulta adecuado, el proceso se reinicia.

Esta lógica puede representarse mediante el siguiente esquema presente en la siguiente página.

Figura 1. Esquema Box-Jenkins



Fuente: Box, G. E. P., Jenkins, G. M. y G. C. Reinsel, *Time Series Analysis: Forecasting and Control* (3a. ed.), Upper Saddle River, Nueva Jersey: Prentice Hall, 1994, p. 17. Reimpreso con permiso.

Dentro de los modelos ARIMA existen tres estructuras fundamentales:

- La parte autorregresiva (AR), que relaciona el valor actual de la serie con sus propios valores pasados.

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \varepsilon_t$$

- La parte de promedio móvil (MA), que modela el valor actual como función de errores pasados.

$$Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Un modelo ARMA general de orden (p,q) se expresa como:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

donde:

- Y_t : valor de la serie en el tiempo t.
- Φ_i : coeficientes autorregresivos
- θ_j : coeficientes de promedio móvil
- ε_t : error aleatorio en el período t.

Cuando la serie no es estacionaria, es posible aplicar una operación de diferenciación para remover la tendencia y estabilizar la media. En estos casos, se trabaja con modelos ARIMA(p, d, q), donde d representa el orden de diferenciación.

Si además se observa estacionalidad, se recurre a la extensión SARIMA, que incorpora componentes autorregresivos, de promedio móvil y de diferenciación en la dimensión estacional. Según Hanke y Wichern (2009), estos modelos “combinan términos autorregresivos regulares y de promedios móviles que explican la correlación en retrasos cortos, y términos autorregresivos estacionales y de promedio móvil que explican la correlación en retrasos estacionales”.

Esta familia de modelos resulta particularmente útil cuando la serie presenta múltiples estructuras superpuestas (tendencia, estacionalidad, ciclos), y su versatilidad permite

ajustarlos de forma progresiva a diferentes escenarios. Su aplicación práctica se analizará en la sección metodológica, donde se evaluará su desempeño para predecir la demanda de productos de Vidriotec con diferentes patrones de consumo.

2.2.5. Modelo Prophet

Prophet es un modelo de series temporales desarrollado por Facebook con el objetivo de proporcionar una herramienta robusta y accesible para usuarios no especializados en estadística, permitiendo la generación de pronósticos confiables a partir de datos con estacionalidades múltiples y potenciales cambios estructurales. Su estructura se basa en un modelo aditivo que descompone la serie en componentes de tendencia, estacionalidad, efectos especiales y ruido aleatorio:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

donde $g(t)$ representa la tendencia (posiblemente con puntos de cambio), $s(t)$ la estacionalidad (anual, semanal o definida por el usuario), $h(t)$ los efectos de feriados o eventos especiales, y ε_t un término de error aleatorio. Esta descomposición es conceptualmente similar a la utilizada en los Modelos Aditivos Generalizados (GAM), presentados por Hastie y Tibshirani (1987), donde la relación entre la variable dependiente y los predictores se expresa como la suma de funciones suaves que permiten capturar relaciones no lineales sin imponer una estructura paramétrica rígida.

En Prophet, las estacionalidades se modelan mediante series de Fourier, la tendencia puede ser lineal o logística con puntos de cambio detectados automáticamente, y los efectos especiales se introducen como regresores binarios definidos por el usuario. Esta configuración modular facilita la interpretación del modelo y permite incorporar conocimiento experto directamente en el proceso de pronóstico.

La documentación oficial de Prophet destaca que el modelo está especialmente diseñado para manejar estacionalidades múltiples y pronunciadas, incluyendo patrones diarios, semanales y anuales, así como efectos de feriados y eventos irregulares. Como resumen, *“Prophet es un procedimiento para pronosticar datos de series temporales basado en un modelo aditivo donde las tendencias no lineales se ajustan con estacionalidades anuales, semanales y*

diarias, además de efectos de feriados” (Facebook Prophet, 2024). A su vez, Moto DEI (2020) señala que el modelo puede integrar no solo estacionalidades y eventos, sino también variables externas relevantes, manteniendo control sobre los niveles de incertidumbre del pronóstico.

De esta forma, Prophet se presenta como una alternativa moderna dentro del ecosistema de modelos para series temporales, particularmente útil cuando se prioriza interpretabilidad, adaptabilidad a estacionalidades múltiples y participación del analista en el proceso de modelado. Su aplicación práctica se detallará en la sección metodológica, donde se evaluará su desempeño frente a otros enfoques tradicionales y de machine learning en el contexto de la demanda de Vidriotec.

2.2.5. Modelos de Machine Learning para series temporales

En los últimos años, el uso de modelos de machine learning para el pronóstico de series temporales ha adquirido creciente relevancia en contextos empresariales, debido a su capacidad para modelar relaciones no lineales complejas, capturar interacciones entre variables y adaptarse a entornos con múltiples dimensiones. A diferencia de los modelos estadísticos clásicos, que se apoyan en una estructura definida del proceso generador de datos, los enfoques de aprendizaje automático priorizan la precisión predictiva, construyendo modelos directamente a partir de los patrones presentes en los datos históricos. Esta flexibilidad permite incorporar información adicional, transformar variables, generar nuevos predictores y optimizar modelos sin necesidad de establecer supuestos tan estrictos. Como señalan James et al. (2021), este tipo de modelos puede ser especialmente útil en escenarios donde la relación entre variables es compleja y difícil de parametrizar analíticamente.

Entre los modelos más utilizados en la literatura aplicada se encuentran Random Forest y XGBoost, ambos basados en árboles de decisión, pero con estructuras y formas de entrenamiento diferenciadas.

2.2.5.1. Random Forest

Random Forest es un algoritmo de ensamble basado en la combinación de múltiples árboles de decisión, cada uno entrenado sobre una muestra aleatoria del conjunto de datos. Su

funcionamiento se apoya en la técnica conocida como *bagging* (bootstrap aggregating), en la cual se construyen numerosos modelos independientes que luego se combinan para generar una predicción final más robusta. La aleatoriedad introducida en la selección de muestras y variables explicativas permite reducir la varianza del modelo y mejorar su capacidad de generalización, incluso en presencia de ruido o relaciones no lineales (Hastie, Tibshirani & Friedman, 2009).

Este modelo ha sido utilizado ampliamente como benchmark en tareas de predicción de demanda debido a su buen desempeño sin necesidad de ajustes finos, su tolerancia a datos faltantes y su capacidad para estimar la importancia relativa de las variables predictoras. En contextos operativos, Random Forest se destaca por su facilidad de implementación, bajo riesgo de sobreajuste y flexibilidad para trabajar con diferentes estructuras de datos. Como muestran Jaimes Campos y López Zúñiga (2021), su aplicación en problemas logísticos permitió mejorar la precisión de los pronósticos y reducir costos operativos de forma significativa.

2.2.5.2. XGBoost

XGBoost (Extreme Gradient Boosting) es un algoritmo de *boosting* secuencial que construye árboles de decisión de manera iterativa, donde cada nuevo árbol intenta corregir los errores cometidos por el conjunto de árboles anteriores. A diferencia de Random Forest, que combina modelos en paralelo, XGBoost lo hace en forma secuencial, ajustando los nuevos árboles a partir de los residuos del modelo anterior. Esta estrategia le permite optimizar una función objetivo penalizada que incluye un término de regularización, lo que reduce el riesgo de sobreajuste y mejora la estabilidad del modelo (Chen & Guestrin, 2016).

Una de las ventajas distintivas de XGBoost es su eficiencia computacional, ya que incorpora técnicas como poda de árboles durante el entrenamiento, manejo interno de valores faltantes y paralelización del proceso de búsqueda de divisiones. Además, ofrece un alto grado de control sobre el modelo mediante la optimización de hiperparámetros como la tasa de aprendizaje (*eta*), la profundidad de los árboles (*max_depth*), el número de iteraciones (*nrounds*) y los coeficientes de regularización (*lambda*, *gamma*), entre otros.

En estudios recientes, como el de Pinzón Villanueva (2023), se han desarrollado procedimientos de optimización de hiperparámetros aplicados a XGBoost, mostrando su capacidad de adaptación en contextos reales de predicción. Estas características han posicionado a este algoritmo como uno de los más utilizados en la industria para tareas de pronóstico, especialmente en situaciones donde se dispone de un conjunto amplio de predictores derivados de la propia serie o de fuentes externas.

Cabe destacar que, si bien los modelos de machine learning presentados no modelan de manera explícita la estructura temporal de la serie —como lo hacen enfoques clásicos como ARIMA o Prophet—, pueden adaptarse eficazmente a tareas de forecasting mediante la generación de variables derivadas. La incorporación de rezagos, medias móviles u otras transformaciones basadas en la propia serie permite capturar dependencias temporales de manera indirecta, sin asumir una estructura funcional predeterminada. Esta flexibilidad metodológica resulta particularmente valiosa en contextos complejos o con múltiples fuentes de información, donde se prioriza la precisión del pronóstico por sobre la interpretabilidad del modelo.

2.2.5.3. Redes Neuronales y modelos avanzados de aprendizaje profundo

En los últimos años, el avance del aprendizaje profundo ha impulsado el desarrollo de modelos más complejos para el pronóstico de series temporales, entre ellos las redes neuronales artificiales. Estos modelos, inspirados en la estructura del cerebro humano, están formados por capas interconectadas de nodos (o “neuronas”) que permiten modelar relaciones no lineales complejas entre variables.

Las arquitecturas más simples, como los perceptrones multicapa (MLP), pueden utilizarse para tareas de regresión y clasificación en problemas estructurados. Su funcionamiento se basa en la propagación de señales a través de una red de capas ocultas y el ajuste de los pesos mediante el algoritmo de retro propagación (backpropagation). Este tipo de redes se explica con claridad en el texto de James et al. (2021), donde se presentan tanto su estructura como su aplicación a problemas prácticos.

En contextos de series temporales más complejas, se han desarrollado arquitecturas especializadas como las redes neuronales recurrentes (RNN) y, dentro de ellas, las LSTM (Long Short-Term Memory). Estas estructuras permiten capturar dependencias de largo plazo en secuencias de datos y han demostrado ser efectivas en series con alta variabilidad y patrones no lineales, aunque suelen requerir una mayor cantidad de datos, mayor capacidad computacional y una etapa de entrenamiento más delicada (Goodfellow, Bengio & Courville, 2016).

Si bien las redes neuronales no constituyen el eje principal de este trabajo, se decidió implementar de forma complementaria un modelo LSTM (Long Short-Term Memory) con el objetivo de explorar su aplicabilidad a la serie de demanda mensual. Como se detalla en la sección 5.2.5, los resultados no fueron satisfactorios, principalmente por el bajo volumen de datos disponible. Aun así, se considera relevante su inclusión como una línea exploratoria que complementa los enfoques estadísticos y basados en árboles, y que podría retomarse en futuras investigaciones con mayor granularidad de datos.

3. Datos

3.1.1. Fuentes de Datos

El presente trabajo se basa en datos históricos provistos por la empresa Vidriotec, cuya actividad principal es la producción y comercialización de envases de vidrio. La base de datos consolida información proveniente de distintas áreas operativas, entre ellas ventas, inventario y previsiones comerciales. A continuación, se detallan las principales fuentes utilizadas, su estructura y nivel de granularidad.

- Ventas: Constituye la base principal del análisis. Incluye información detallada de todas las operaciones de venta realizadas entre enero de 2021 y diciembre de 2024, con un volumen aproximado de 30.000 registros por año. Cada observación corresponde a una transacción diaria e incluye variables como:

Fecha	Tipo	Tipo en R	Descripción
Fecha	Fecha	date	Fecha de la venta (día/mes/año)
SKU	Catagórica	chr	Código único que identifica el producto vendido
SKU_Nombre	Catagórica	chr	Nombre comercial del producto
Modelo	Catagórica	dbl	Identificador de modelo del producto
Medida	Catagórica	chr	Unidad de medida (p. ej. 360, 730)
Nro_Sucursal	Catagórica	chr	Código de la sucursal que realizó la compra
Cliente_Nombre	Catagórica		Nombre del cliente
Cantidad	Numérica	dbl	Unidades físicas vendidas
Pesos	Numérica	dbl	Importe de la venta en moneda local (ARS)
Exportación	Binaria	integer	Flag de exportación (0 = nacional, 1 = exportado)
Dolar	Binaria	integer	Flag de facturación en dólares (0 = no, 1 = sí)
USD	Numérica	dbl	Importe de la venta en USD
Provincia	Catagórica	chr	Provincia donde se realizó la venta
País	Catagórica	chr	País donde se realizó la venta
Kg	Numérica	dbl	Peso total en kilogramos de las unidades vendidas

La unidad de análisis original es SKU–cliente–día, que fue posteriormente agregada a nivel SKU–cliente–mes para el desarrollo de modelos de predicción de demanda. Esta transformación permite capturar patrones estacionales, estabilizar la serie y facilitar la incorporación de variables rezagadas y agregadas.

- **Forecasts de clientes:** Estos datasets son proporcionados por los principales clientes de la empresa, en su mayoría grandes compañías del sector alimenticio y de bebidas. Incluyen el detalle de la cantidad de unidades que proyectan comprar en los próximos seis meses, desagregadas por producto. Su principal utilidad en esta investigación es servir como referencia externa (benchmark) para evaluar el desempeño de las técnicas de pronóstico desarrolladas. Si bien no fueron utilizados como insumo directo en todos los modelos, sí permiten contrastar los resultados obtenidos con las proyecciones realizadas por los propios clientes, aportando una perspectiva adicional sobre la precisión y utilidad de los enfoques propuestos.
- **Variables macroeconómicas (no utilizadas):** Se evaluó la inclusión de indicadores macroeconómicos como el Índice de Producción Industrial y el Índice de Confianza del Consumidor. Sin embargo, debido a su baja frecuencia, escasa variabilidad mensual y débil correlación con la demanda desagregada por producto y cliente, se optó por no incorporarlos en la versión final del modelo.
- **Calidad de los datos:** La calidad general de los datos fue satisfactoria. No se detectaron valores faltantes significativos ni outliers extremos que afectaran el análisis. Se realizaron tareas de validación, limpieza y estandarización de campos, así como ajustes en la codificación de variables categóricas y numéricas. Uno de los principales desafíos fue el tratamiento de variables con una gran cantidad de valores únicos, como SKU o Cliente_Sucursal, que presentan una amplia diversidad de categorías posibles. Esto implicó la necesidad de aplicar transformaciones específicas para su correcta incorporación en los modelos, tema que se aborda en la siguiente sección.

3.1.2 Variables Relevantes

A partir de las fuentes de datos descritas en el apartado anterior se construyeron las variables utilizadas para los modelos de predicción; tras una etapa de validación y preprocesamiento, y considerando su disponibilidad, calidad y relevancia para el objetivo del modelo, se seleccionaron las variables que se detallan a continuación, organizadas en tres categorías: variable objetivo, variables categóricas y variables numéricas.

Variable objetivo

- Cantidad_Unidades: variable numérica continua que representa la cantidad de unidades vendidas en cada observación diaria (posteriormente agregada a nivel mensual para el entrenamiento de los modelos). Constituye la variable objetivo del análisis.

Cabe destacar que esta variable refleja las ventas concretadas, es decir, las unidades efectivamente entregadas al cliente. No representa la demanda real o solicitada, sino el mínimo entre la demanda potencial y el inventario disponible en cada momento ($\text{Min}\{\text{Demanda}, \text{Inventario}\}$). Esta diferencia introduce una limitación relevante, ya que en situaciones de quiebre de stock la demanda observada puede subestimar la demanda efectiva.

Sin embargo, en el contexto de este trabajo, esta limitación es acotada: el nivel de servicio de la empresa, medido como OTIF (On Time In Full), se mantiene en torno al 85–90%, lo cual indica que en la mayoría de los casos la demanda es efectivamente cubierta. Las excepciones suelen deberse a cuestiones operativas como productos fuera de ciclo productivo o aumentos imprevistos de demanda por parte de ciertos clientes que no lograron anticiparse.

Desde un enfoque metodológico, la disponibilidad futura de información adicional –como pedidos no satisfechos o indicadores de stock insuficiente– permitiría modelar la demanda latente o ajustar por censura. En este estudio, no obstante, se opta por utilizar la variable de ventas concretadas como *proxy* de la demanda, con una justificación empírica razonable dada la alta cobertura observada.

Variables predictoras categóricas

Las siguientes variables categóricas fueron incluidas como insumos para los modelos predictivos, dado que explican parte relevante del comportamiento de la demanda:

- SKU (Artículo): identifica el producto vendido. Permite capturar patrones de consumo específicos por artículo, esenciales en un entorno con alta variedad de productos.
- Cliente_Sucursal: representa al cliente y su ubicación o punto de entrega. Es clave para segmentar comportamientos de compra y detectar efectos específicos por canal o relación comercial.
- Provincia: incorpora la dimensión geográfica, útil para capturar patrones logísticos, estacionales o de concentración de demanda por región.
- Rubro: cuando está disponible, permite agrupar artículos en categorías con dinámicas similares de venta.

Dado que modelos como XGBoost y Random Forest requieren variables numéricas como entrada, estas variables categóricas fueron transformadas mediante codificación one-hot (one-hot encoding). Esta técnica consiste en convertir cada categoría en una columna binaria independiente. Se implementó en R utilizando la función `model.matrix()`, y las matrices resultantes fueron integradas al conjunto de datos principal para el entrenamiento de los modelos.

Este enfoque permite que los algoritmos capten relaciones no lineales entre categorías y demanda. Sin embargo, incrementa considerablemente la dimensionalidad del dataset, en especial en variables con alta cardinalidad como Cliente_Sucursal o SKU. Para mitigar este efecto, se realizaron pruebas con subconjuntos representativos y se evaluó el impacto de la inclusión de cada variable sobre el desempeño del modelo.

En trabajos futuros o con otros tipos de modelos, como redes neuronales, podrían considerarse alternativas como embeddings o codificación ordinal, que permiten representar variables categóricas en espacios de menor dimensión manteniendo relaciones semánticas entre categorías.

Variables predictoras numéricas

Entre las variables predictoras numéricas se incluyen componentes temporales explícitos, así como variables construidas a partir de la serie histórica de ventas:

- Mes, Año: variables temporales explícitas, útiles para capturar efectos estacionales y tendencias a lo largo del tiempo.
- Lags de demanda: se incorporaron rezagos de la variable Cantidad_Unidades para capturar dependencia temporal:
 - lag_1, lag_2, lag_3, lag_6, lag_12: rezagos de 1, 2, 3, 6 y 12 meses respectivamente.
- Medias móviles: se calcularon promedios móviles para suavizar la serie y capturar patrones de nivel y tendencia:
 - ma_3, ma_6, ma_12: promedios móviles de 3, 6 y 12 meses.

Estas variables se utilizaron como insumos clave en los modelos basados en árboles de decisión, dado que permiten incorporar la dinámica histórica de cada serie sin necesidad de una especificación estructural del modelo temporal. Su inclusión se basa en la literatura sobre forecasting de series temporales con machine learning, y ha demostrado ser particularmente útil para capturar patrones repetitivos o de inercia en la demanda.

El conjunto final de variables constituye una base sólida para los modelos desarrollados, y será complementado con evidencia empírica en el análisis exploratorio presentado a continuación.

3.2 Análisis exploratorio de los datos

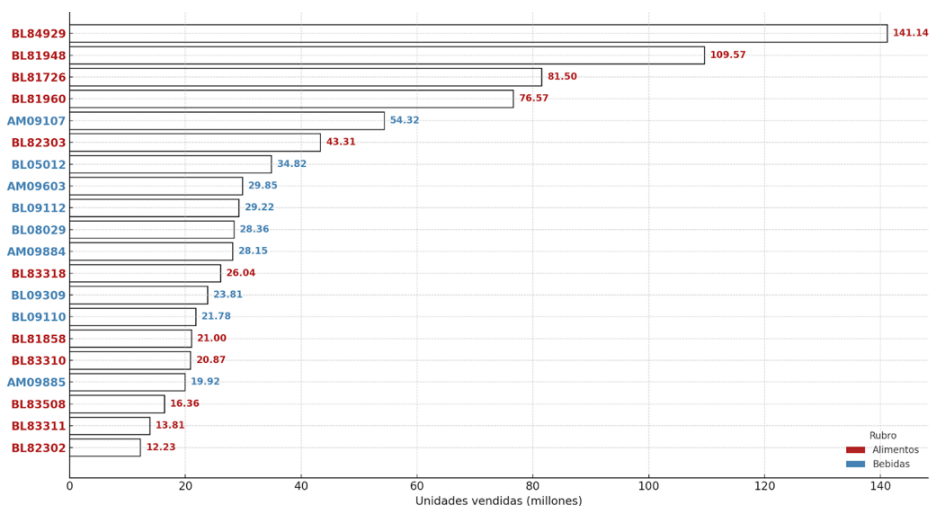
Esta sección tiene por objetivo explorar en profundidad la estructura y el comportamiento de la demanda observada, identificando patrones relevantes, posibles irregularidades y relaciones significativas entre las variables disponibles. Se abordan tanto aspectos cuantitativos propios del negocio —como el volumen de unidades vendidas por SKU, cliente o rubro— como también dimensiones estadísticas vinculadas a la variabilidad temporal, la estacionalidad y la correlación entre variables.

Además del análisis descriptivo, se incorporan herramientas estadísticas como el análisis de autocorrelación (ACF) para detectar dependencias temporales, y métodos de aprendizaje automático orientados a estimar la importancia relativa de los distintos predictores. Estos últimos permiten anticipar, de forma preliminar, qué variables podrían resultar más relevantes al momento de construir los modelos de pronóstico, tema que será desarrollado con mayor profundidad en la sección metodológica.

En conjunto, esta etapa exploratoria cumple una doble función: por un lado, facilita una comprensión detallada del fenómeno de demanda en Vidriotec y, por otro, permite fundamentar con evidencia empírica las decisiones de segmentación, selección de variables y diseño de modelos que se implementarán más adelante.

3.2.1 Análisis de la demanda por SKU y cliente

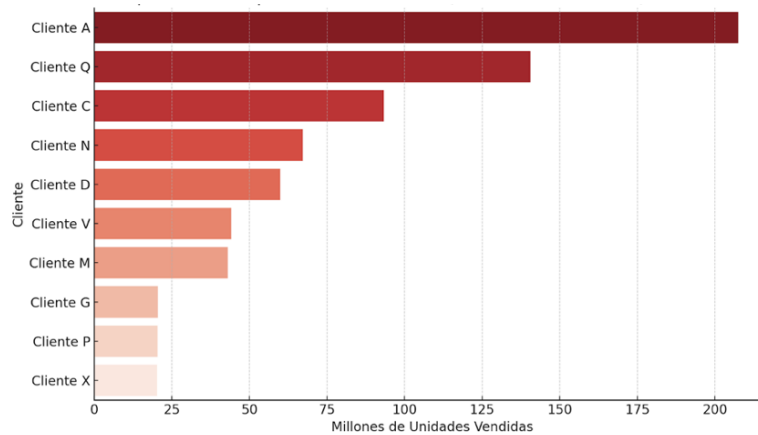
Figura 2. Top 20 SKUs por unidades vendidas (2021–2024)



El gráfico presenta los veinte SKUs con mayor volumen de unidades vendidas durante el periodo 2021–2024. Se aprecia una clara concentración de la demanda en un reducido conjunto de productos: los cinco primeros superan los 75 millones de unidades cada uno, y el SKU líder alcanza los 141 millones. Esta distribución pone de manifiesto una estructura de portafolio altamente concentrada, en la que dichos veinte SKUs representan aproximadamente el 60 % del volumen total comercializado.

La segmentación por rubro (alimentos y bebidas) permite observar que los productos alimenticios dominan en términos de volumen, aunque las bebidas también tienen representación destacada. Esta diferenciación es útil para entender dinámicas específicas de consumo y planificar estrategias diferenciadas de abastecimiento, producción y pronóstico.

Figura 3. Top 10 clientes por volumen vendido (2021–2024)

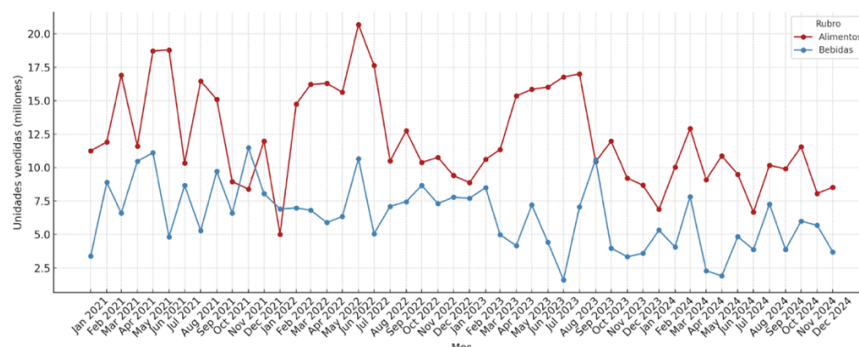


En cuanto a la demanda por cliente, se identifican los diez principales compradores en términos de unidades adquiridas durante el mismo período. El análisis revela una alta concentración en grandes cuentas: los tres primeros clientes —Arcor, Quilmes y Cía. Industrial— explican conjuntamente más de 440 millones de unidades, lo que representa una proporción significativa del total comercializado.

Este patrón sugiere una dependencia operativa relevante hacia un grupo reducido de clientes estratégicos, lo que implica desafíos y oportunidades. Por un lado, permite trabajar con mejores previsiones a partir de relaciones comerciales estables y flujos regulares de información (como los forecasts que algunos de ellos aportan). Por otro, resalta la necesidad de mitigar riesgos asociados a eventuales caídas de demanda o cambios contractuales por parte de estos actores clave.

3.2.2 Análisis por rubro

Figura 4. Evolución mensual de ventas totales por rubro – 2021 a 2024

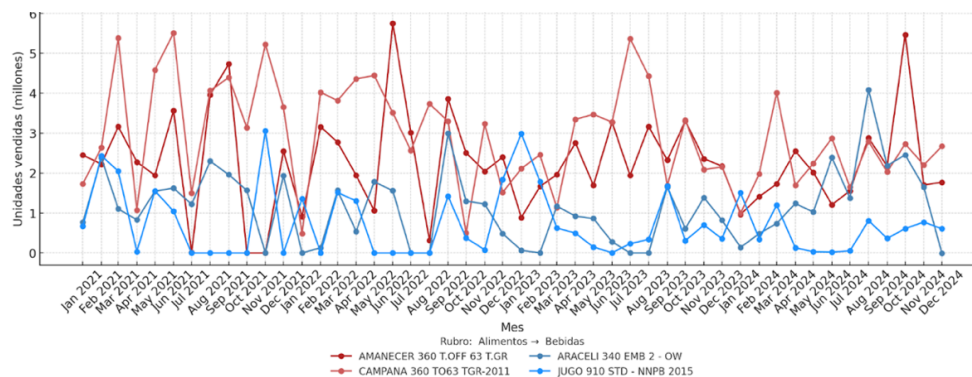


El gráfico presenta la evolución mensual del volumen de unidades vendidas, diferenciadas entre los rubros alimentos (en rojo) y bebidas (en azul) durante el período 2021–2024. Se observa que el rubro alimentos mantiene un nivel de ventas sistemáticamente superior, con varios picos que superan los 20 millones de unidades en determinados meses. Por su parte, las bebidas muestran una menor participación relativa, con un rango de ventas que oscila mayormente entre 5 y 12 millones de unidades.

A lo largo del período se identifican fluctuaciones mensuales recurrentes en ambos rubros, lo cual sugiere la presencia de comportamientos estacionales, aunque con distinta intensidad. Este análisis permite visualizar cómo varía la demanda en cada categoría a lo largo del tiempo y aporta elementos valiosos para entender la composición y dinámica general del volumen comercializado por la empresa.

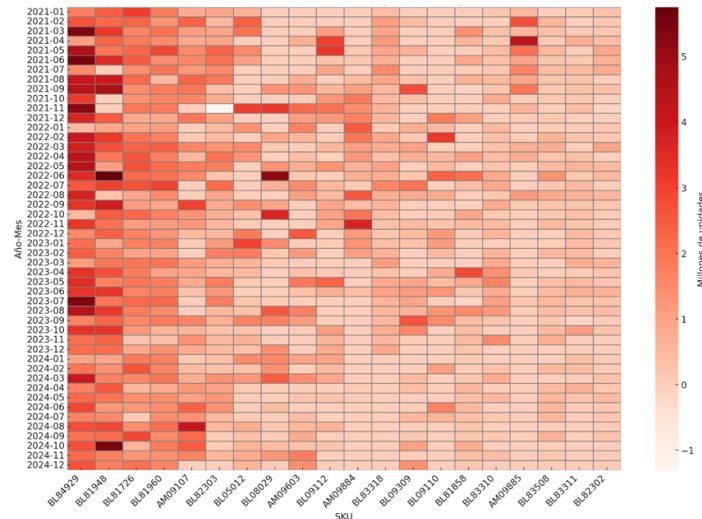
3.2.3 Identificación de patrones estacionales y de tendencia

Figura 5. Evolución mensual de unidades vendidas en SKUs seleccionados (2021–2024)



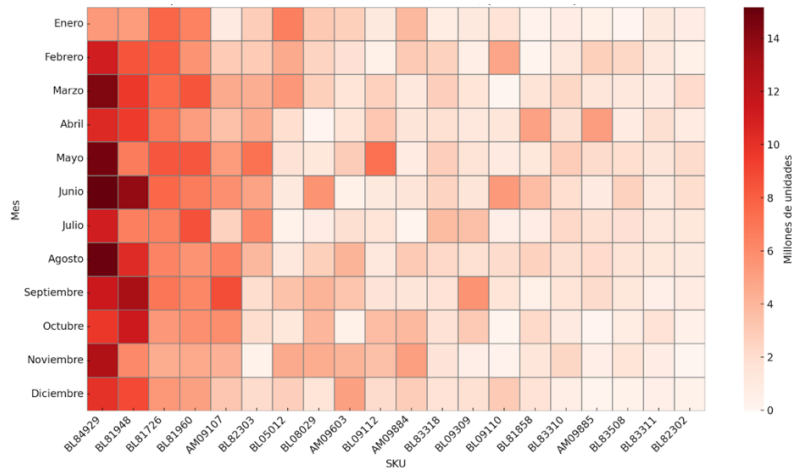
Este gráfico representa la evolución mensual de las ventas para un conjunto de SKUs seleccionados, pertenecientes a los rubros alimentos y bebidas. Se observan comportamientos heterogéneos entre productos, con algunos que presentan una demanda relativamente estable y otros con oscilaciones significativas a lo largo del tiempo. También se identifican diferencias claras entre rubros: en general, los SKUs de alimentos muestran mayor volumen y mayor variabilidad mensual, mientras que los SKUs de bebidas exhiben una estacionalidad más acentuada en determinados tramos del año. Esta visualización permite observar con mayor granularidad las dinámicas individuales que pueden quedar ocultas en los agregados por rubro.

Figura 6. Mapa de calor de ventas mensuales – Top 20 SKU (2021–2024)



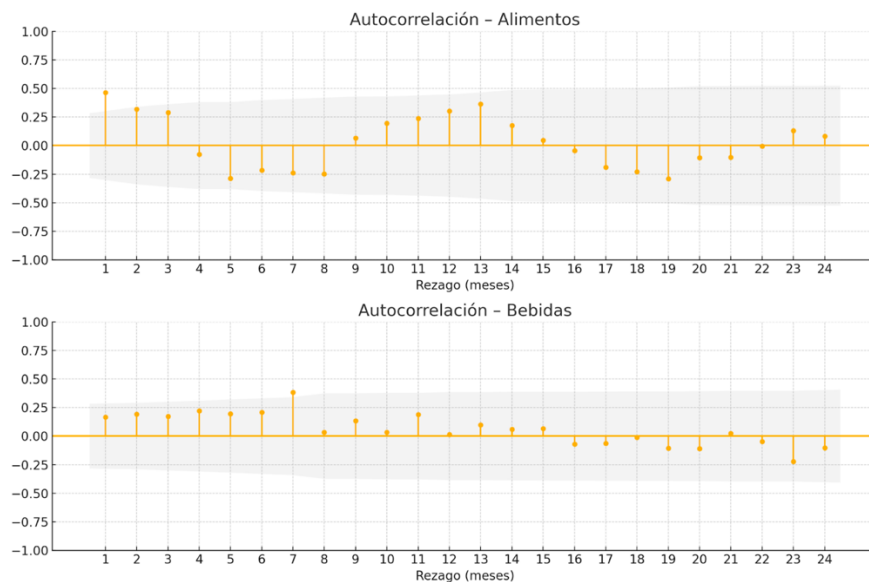
El mapa de calor permite visualizar la distribución mensual de ventas, en millones de unidades, para los 20 productos más vendidos a lo largo del período analizado. Cada fila representa un mes y cada columna un producto, con la intensidad del color indicando el volumen de ventas. Se observan ciclos de demanda marcados para determinados SKUs, con patrones de concentración de ventas en meses específicos. Esta herramienta facilita la detección de comportamientos repetitivos y variaciones interanuales, y refuerza la necesidad de considerar el componente estacional en el modelado de la demanda.

Figura 7. Mapa de calor de ventas por mes agregado – Top 20 SKUs (2021–2024)



A diferencia del gráfico anterior, este mapa de calor presenta la suma acumulada de unidades vendidas por mes calendario, agregada para todo el período 2021–2024. De esta manera, permite identificar estacionalidad intra-anual para los productos más relevantes. Se destaca, por ejemplo, una mayor concentración de ventas en los meses de abril, mayo y diciembre para varios SKUs, lo cual sugiere picos estacionales posiblemente asociados a eventos comerciales, patrones de consumo anual o planificación de abastecimiento de los clientes. Esta representación permite priorizar meses críticos para la planificación de producción y logística.

Figura 8. Gráfico de autocorrelación (ACF) mensual por rubro



Con el objetivo de identificar posibles patrones estacionales en la demanda, se realizó un análisis de autocorrelación para cada uno de los rubros disponibles en la base de datos: Alimentos y Bebidas.

En el caso de Alimentos, se observó una autocorrelación elevada en el rezago 1 (0.46), lo que sugiere una fuerte dependencia entre las ventas de un mes y las del mes inmediatamente anterior. También se identificaron valores significativos en el rezago 2 (0.32) y en el rezago 12 (0.30), este último indicando la presencia de estacionalidad anual moderada. El valor máximo de autocorrelación dentro de los primeros 24 rezagos fue de 0.46, lo que evidencia una alta estabilidad y regularidad en la serie temporal de este rubro.

En contraste, el rubro Bebidas mostró una autocorrelación más débil en los primeros rezagos (0.17 en el rezago 1 y 0.19 en el rezago 2), y prácticamente nula en el rezago 12 (0.02), lo que indica ausencia de estacionalidad anual significativa. El valor más alto de autocorrelación en los primeros 24 rezagos fue de 0.38, posiblemente atribuible a dinámicas internas o ciclos de reposición, pero sin evidencia clara de patrones estacionales sostenidos.

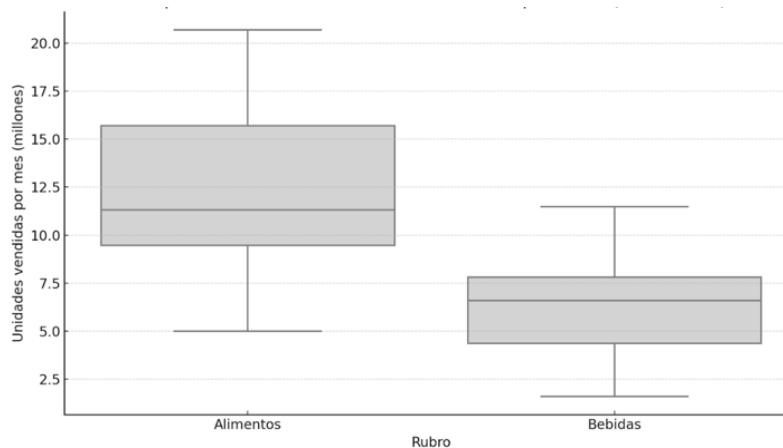
Según Hanke, “la autocorrelación de primer orden (rezago 1) es común en series de demanda con comportamiento persistente, mientras que la existencia de autocorrelación en rezagos múltiples, como el 12, es indicativa de patrones estacionales” (Hanke, capítulo 6). Esta información resulta fundamental al momento de seleccionar el tipo de modelo a utilizar, ya sea SARIMA, modelos de suavización estacional, o incluso enfoques híbridos que combinen componentes lineales y no lineales.

En función de estos resultados, se concluye que el rubro Alimentos presenta una demanda más estructurada y predecible, lo que respalda el uso de modelos estacionales en su modelización. Por el contrario, el comportamiento más errático del rubro Bebidas sugiere la necesidad de emplear modelos más flexibles o adaptativos, enfocados en capturar variaciones de corto plazo o sensibles a variables exógenas.

3.2.4 Dispersión y volatilidad

El gráfico tipo boxplot permite visualizar la dispersión mensual de las unidades vendidas para cada rubro en el período 2021–2024. El rubro Alimentos presenta una mediana de ventas más elevada, acompañada de un rango intercuartílico y una amplitud total considerablemente mayores. Esto sugiere no solo un mayor volumen de ventas mensual promedio, sino también una variabilidad significativa en la demanda, que podría estar asociada a picos estacionales marcados o a la influencia de productos de alta rotación con comportamiento irregular. Como se muestra en la Figura 9.

Figura 9. Dispersión mensual de unidades vendidas por rubro (2021–2024)

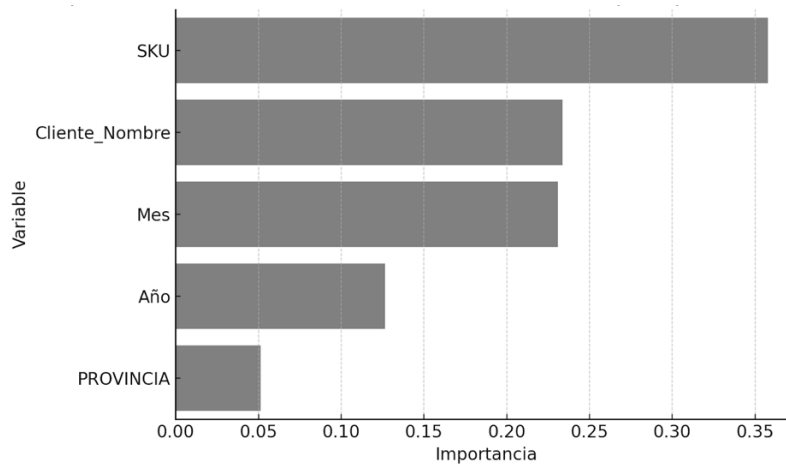


En contraste con lo explicado anteriormente para el rubro de Alimentos, el de Bebidas exhibe una mediana más baja y una dispersión más acotada, reflejando una dinámica de ventas más estable a lo largo del tiempo. Aunque su volumen absoluto es menor al de Alimentos, la menor variabilidad podría implicar una mayor previsibilidad en la demanda.

Este tipo de representación es particularmente útil para caracterizar la volatilidad de la serie por categoría, permitiendo distinguir entre rubros con comportamiento más errático y aquellos con patrones más estables. Esta información resulta relevante al momento de seleccionar modelos de pronóstico, ya que series con alta variación intermensual podrían beneficiarse del uso de modelos más robustos o adaptativos.

3.2.5 Importancia empírica de variables predictoras

Figura 10. Importancia de variables – Modelo Random Forest para predicción de cantidad



Con el objetivo de identificar los factores con mayor capacidad explicativa sobre la demanda, se entrenó un modelo de Random Forest utilizando como variables predictoras el identificador del producto (SKU), el nombre del cliente (Cliente_Nombre), la provincia de destino (PROVINCIA) y componentes temporales como el mes y el año de la operación.

El gráfico resultante muestra la importancia relativa de cada predictor en la construcción del modelo, medida a partir de su contribución a la reducción del error de predicción en los distintos árboles que componen el bosque. Esta métrica, conocida como *feature importance*, permite cuantificar empíricamente la relevancia de cada variable.

Los resultados indican que las variables SKU y Cliente_Nombre son las más influyentes en la predicción de la cantidad de unidades vendidas. Esto resulta coherente, ya que ambos atributos condensan los principales determinantes comerciales de la demanda: el tipo de producto y el cliente que lo adquiere. En tercer lugar aparece el Mes, reforzando la evidencia previa sobre la existencia de patrones estacionales. Las variables Año y PROVINCIA mostraron una menor contribución relativa, aunque aportan información complementaria sobre la evolución temporal y la distribución geográfica.

De acuerdo con James et al. (2021) en *An Introduction to Statistical Learning*, los modelos de ensamble como Random Forest son especialmente útiles no solo por su capacidad predictiva, sino también por permitir el análisis de importancia de variables, lo que aporta

valor interpretativo y puede guiar procesos de selección de atributos en futuras versiones del modelo.

4. Metodología

En este capítulo se presenta el enfoque metodológico empleado para generar los pronósticos de demanda de Vidriotec. Se inicia con la definición del nivel de desagregación (SKU–mes y total mensual) y el horizonte temporal de pronóstico (octubre–diciembre de 2024), delimitando así el alcance y la granularidad del análisis; a continuación, se describe el proceso de preparación de los datos, que incluye limpieza, transformación e ingeniería de variables; finalmente, se detallan los modelos predictivos seleccionados —desde técnicas clásicas de series de tiempo hasta algoritmos de aprendizaje automático— junto con la fundamentación teórica y empírica de su elección. Los desarrollos específicos de cada etapa se presentan en las Secciones 4.1 a 4.3.

4.1 Definición del nivel y horizonte de pronóstico

El objetivo de esta sección es establecer las unidades de análisis para la predicción y el horizonte temporal sobre el cual se realizaron los pronósticos de demanda. Se busca exponer y fundamentar las decisiones adoptadas durante el proceso de modelado, orientadas a obtener resultados que sean relevantes desde el punto de vista operativo y de mayor impacto para la planificación del negocio.

4.1.1 Nivel de pronóstico: SKU–mes y total mensual

El nivel principal de predicción seleccionado en este trabajo fue el de SKU–mes, es decir, la estimación de la cantidad de unidades demandadas para cada producto específico (SKU) en cada mes calendario. Este enfoque granular permite capturar variaciones en el comportamiento de demanda de productos individuales, proporcionando información clave para optimizar la planificación operativa, con el objetivo de maximizar el nivel de servicio de cara a los clientes, satisfacer de manera más eficiente la demanda y minimizar los niveles de inventario necesarios.

Adicionalmente, se realizó una predicción del volumen total mensual agregado, considerando la suma de las unidades vendidas de todos los SKUs. Esta predicción agregada resultó especialmente relevante para modelos como el ARIMA univariado y para el monitoreo del desempeño general de la planificación comercial.

La decisión de trabajar en ambos niveles responde a recomendaciones de la literatura especializada. Según Makridakis, Wheelwright y Hyndman (1998), “el nivel de agregación de los datos debe coincidir con el tipo de decisión que se busca tomar: las decisiones operativas requieren pronósticos desagregados, mientras que las decisiones estratégicas se benefician de pronósticos agregados”. De manera similar, Hanke y Wichern (2010) destacan que “los pronósticos a un nivel más desagregado permiten revelar tendencias específicas de productos y patrones estacionales que podrían quedar ocultos en datos agregados”.

La combinación de niveles de predicción permite abordar de manera integrada tanto necesidades operativas —como la asignación eficiente de recursos productivos— como estratégicas, facilitando una planificación a diferentes escalas de agregación del negocio. Esta práctica, además, se alinea con recomendaciones de Fildes, Goodwin, Lawrence y Nikolopoulos (2008), quienes sostienen que modelar distintos niveles de agregación en forma complementaria mejora la adaptabilidad de la planificación ante cambios operativos y de mercado.

4.1.2 Horizonte de pronóstico: octubre–diciembre 2024

En lo que refiere al horizonte de pronóstico desarrollado en esta tesis, es importante destacar que su definición se basó en el análisis de la operativa de la empresa y en su forma de planificar la producción. Asimismo, se consideró la periodicidad con la que reciben pronósticos de demanda de sus principales clientes. Tomando en cuenta este conocimiento del negocio, se definió realizar predicciones a nivel mensual, tanto de manera desagregada por SKU como en forma agregada, con un horizonte de uno a tres meses hacia adelante.

Esta decisión estuvo motivada por las características del proceso productivo de la compañía, que requiere tiempos operativos significativos desde el lanzamiento de una orden de

producción hasta la disponibilidad del producto terminado en inventario o en la sucursal del cliente. De este modo, la elección del horizonte buscó ajustarse a los tiempos típicos de planificación anticipada de la empresa.

Se evalúa el desempeño de los modelos en la predicción de los últimos tres meses de 2024, el período más reciente para el cual se contó con datos disponibles. Se seleccionó este trimestre como horizonte de pronóstico y evaluación por diversas razones: permite maximizar la cantidad de datos utilizados para el entrenamiento de los modelos, coincide con el período operativo típico de planificación y, además, representa una fase del año crítica para la demanda de la empresa. Durante el último trimestre, se combinan dos patrones relevantes: la demanda relativamente estable del rubro alimenticio y la demanda estacional del rubro de bebidas, que suele incrementarse de manera significativa.

La elección de un horizonte de tres meses permite evaluar la capacidad de los modelos en un contexto realista de planificación operativa, alineándose con recomendaciones de la literatura, que sugiere que los horizontes de corto a mediano plazo son los más relevantes para mejorar la precisión de la gestión operativa de la demanda (Makridakis, Wheelwright y Hyndman, 1998; Fildes et al., 2008).

4.2. Preparación de los datos

Las fuentes de datos utilizadas en el presente trabajo provienen principalmente de los sistemas internos de la empresa y de información adicional provista por especialistas del negocio. Se contó con una base de datos de ventas entre enero de 2021 y diciembre de 2024. Para el entrenamiento de los modelos se utilizaron observaciones hasta septiembre de 2024 inclusive, reservándose el último trimestre (octubre–diciembre de 2024) para la evaluación de los pronósticos. Además, se complementó la base de ventas con detalles relevantes sobre los productos más representativos para la empresa, los principales clientes, la planificación de la producción y los niveles promedio de inventario y servicio.

Estos conocimientos específicos fueron fundamentales para orientar la preparación del dataset de trabajo. En particular, la selección de los 31 SKUs más representativos se realizó en colaboración con expertos internos de la empresa, quienes aportaron un conocimiento

detallado sobre la importancia comercial y operativa de cada producto. Este proceso es fundamental en la construcción de modelos de pronóstico, ya que “un conocimiento profundo del negocio, de sus productos, clientes y procesos, es esencial para mejorar la precisión de los pronósticos” (Hanke & Wichern, 2010). La identificación de los artículos estratégicos permitió enfocar el análisis en productos de alta rotación y asegurar una base de observaciones adecuada para un entrenamiento robusto de los modelos.

Sobre esta base, se procedió a enriquecer el conjunto de datos mediante la construcción de variables adicionales. Se creó manualmente una variable que identifica el modelo del artículo, diferenciando productos que comparten características de diseño o tamaño. También se incorporó una variable de medida, para distinguir envases de diferentes capacidades (por ejemplo, 340, 710 o 1000 ml). Además, se clasificó cada SKU según el rubro comercial al que pertenece (Bebidas o Alimentos), representando casi la totalidad del volumen de ventas de la compañía. Estas variables fueron incorporadas para permitir un análisis más detallado de los patrones de demanda a nivel de producto y categoría.

En cuanto a la calidad de los datos, no se observaron cantidades significativas de valores faltantes. Las principales ausencias se registraron en la variable Provincia para registros de exportaciones, situación que fue corregida asignando el valor "Exportación" a dichos casos.

Posteriormente, los datos fueron importados en el entorno R para su procesamiento. En esta etapa, se crearon variables adicionales orientadas a capturar patrones temporales y estacionales de la demanda. Se generaron variables de rezagos (*lags*) y promedios móviles de unidades vendidas, buscando aprovechar la información histórica disponible. La incorporación de estas variables responde a la necesidad de capturar efectos de inercia y estacionalidad en la demanda, fenómenos comúnmente observados en series temporales de ventas. Asimismo, las variables categóricas principales (SKU, cliente, provincia, rubro) fueron codificadas numéricamente mediante técnicas de *one-hot encoding* para su inclusión en modelos de machine learning.

Se respetó estrictamente la estructura temporal de los datos, asegurando que todas las variables predictoras fueran construidas únicamente a partir de información disponible hasta

cada punto temporal, evitando así cualquier riesgo de *data leakage* en el entrenamiento de los modelos.

En resumen, el conjunto de datos final utilizado en el modelado predictivo integró variables numéricas (mes, año, rezagos, medias móviles) y categóricas (SKU, cliente, provincia, rubro), con el objetivo de predecir la cantidad de unidades demandadas a nivel SKU–mes y total mensual agregado.

4.3. Selección y descripción de los modelos predictivos

En este apartado se describen los modelos predictivos analizados y desarrollados para abordar el problema de pronóstico de demanda de Vidriotec. La selección de modelos combina enfoques clásicos de series temporales y técnicas modernas de aprendizaje automático, con el objetivo de capturar tanto patrones temporales como relaciones complejas entre múltiples variables explicativas. Se justifica su elección con base en criterios teóricos y empíricos, así como en su aplicabilidad concreta al contexto operativo de la empresa.

4.3.1. Regresión lineal múltiple

En una primera instancia se exploró la aplicación de un modelo de regresión lineal múltiple, dada su capacidad explicativa y su uso frecuente en análisis de demanda. El modelo incluía como predictores variables categóricas como el SKU, cliente, provincia y mes.

Sin embargo, al analizar los supuestos clásicos de este modelo, se detectó evidencia de heterocedasticidad en los residuos mediante el test de Breusch–Pagan (Breusch & Pagan, 1979), lo que comprometía la validez de las inferencias y su uso como modelo de predicción.

Las hipótesis del test fueron las siguientes:

H0: Los residuos tienen varianza constante (homocedasticidad)

H1: La varianza de los residuos depende de las variables explicativas (heterocedasticidad)

El resultado del test fue:

```
> print(bp_test)

studentized Breusch-Pagan test

data: modelo_lm
BP = 453.86, df = 361, p-value = 0.0006371
```

Dado que el *p-value* es considerablemente menor a 0.05, se rechaza la hipótesis nula, concluyendo que existe heterocedasticidad. Por lo tanto, se descartó el uso de la regresión múltiple como modelo final, priorizando modelos más robustos frente a este tipo de limitaciones.

Adicionalmente, se aplicó una versión extendida del test conocida como test de White (White, 1980), que incorpora términos cuadráticos para detectar formas más generales de heterocedasticidad. Como muestra la siguiente figura, los resultados fueron consistentes con el test de Breusch–Pagan tradicional:

```
> print(white_test)

studentized Breusch-Pagan test

data: modelo_lm
BP = 33.422, df = 8, p-value = 5.168e-05
```

Los resultados refuerzan la evidencia de heterocedasticidad, respaldando la decisión metodológica de no utilizar modelos lineales clásicos bajo estas condiciones.

Si bien en este trabajo se optó por descartar el uso de modelos lineales debido a la presencia de heterocedasticidad, otra alternativa metodológica que podría considerarse en futuras investigaciones es el uso de modelos ARCH o GARCH (Engle, 1982; Bollerslev, 1986), que permiten modelar directamente la varianza condicional de los errores en series temporales. Este enfoque podría ser especialmente útil si se detectara dependencia temporal en la volatilidad de los residuos.

4.3.2 Modelo de series temporales: ARIMA y SARIMA

En esta sección se presentan los modelos univariados de series temporales utilizados para el pronóstico de demanda mensual: ARIMA (AutoRegressive Integrated Moving Average) y su extensión estacional SARIMA (Seasonal ARIMA). Ambos modelos fueron seleccionados por su robustez teórica y amplia aplicación en contextos de forecasting, especialmente en dominios donde los datos presentan estructuras temporales con componentes de tendencia y estacionalidad (Box y Jenkins, 1976).

Los modelos fueron implementados utilizando la función `auto.arima()` del paquete `forecast` de R (Hyndman y Khandakar, 2008), que permite seleccionar automáticamente los parámetros óptimos del modelo en función del criterio de información AICc, y puede ajustarse tanto en su versión no estacional como estacional.

4.3.2.1. Aplicación a la serie mensual agregada

En primera instancia, se estimaron los modelos ARIMA y SARIMA sobre la serie temporal de demanda mensual total, es decir, considerando el volumen total de unidades vendidas por mes sin distinción de producto.

Para justificar la aplicación de modelos ARIMA, se realizó inicialmente el test de Dickey–Fuller aumentado (ADF), obteniendo un valor p de 0.04403, lo que indica que la serie puede considerarse estacionaria en su nivel original, aunque cercana al umbral crítico (Dickey y Fuller, 1979). A través de los gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF), se observó una autocorrelación significativa en los primeros rezagos, lo que respalda la elección de un modelo que capture explícitamente estas dependencias.

El modelo ARIMA seleccionado automáticamente fue $ARIMA(0,1,1)(1,0,0)[12]$. Aunque esta estructura incluye un componente estacional autorregresivo, la estacionalidad se validó más formalmente mediante la estimación de un modelo SARIMA con `seasonal = TRUE`, lo que habilita la búsqueda explícita de componentes estacionales adicionales. El modelo SARIMA resultante también adoptó la forma $(0,1,1)(1,0,0)[12]$, con transformaciones Box–Cox automáticas para estabilizar la varianza.

Ambos modelos fueron evaluados mediante análisis visual de residuos y el test de Ljung–Box. En el caso de ARIMA, el p-value fue de 0.04399, indicando una leve autocorrelación residual. Para el modelo SARIMA, el p-value fue de 0.6917, ampliamente superior al umbral de 0.05, lo cual indica una ausencia clara de autocorrelación residual significativa. Esto sugiere que el modelo SARIMA logró capturar de forma más completa las dependencias presentes en la serie, incluyendo aquellas de naturaleza estacional.

A continuación, se presentan los gráficos correspondientes a ambos modelos:

Figura 11. Autocorrelación (ACF) - ARIMA

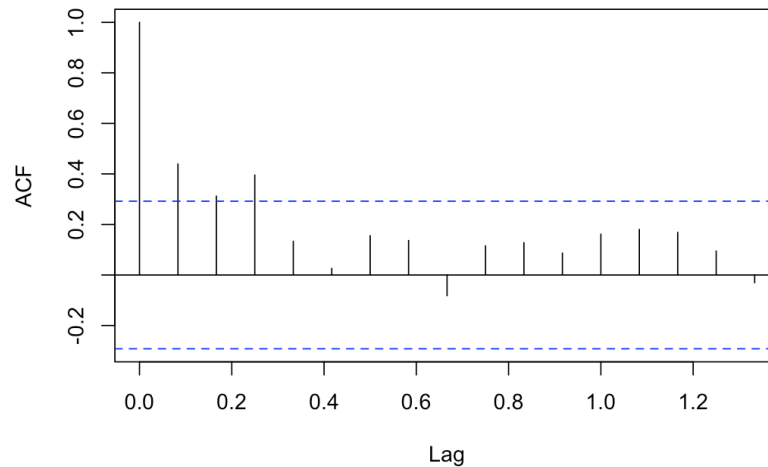


Figura 12. Autocorrelación Parcial (PACF) - ARIMA

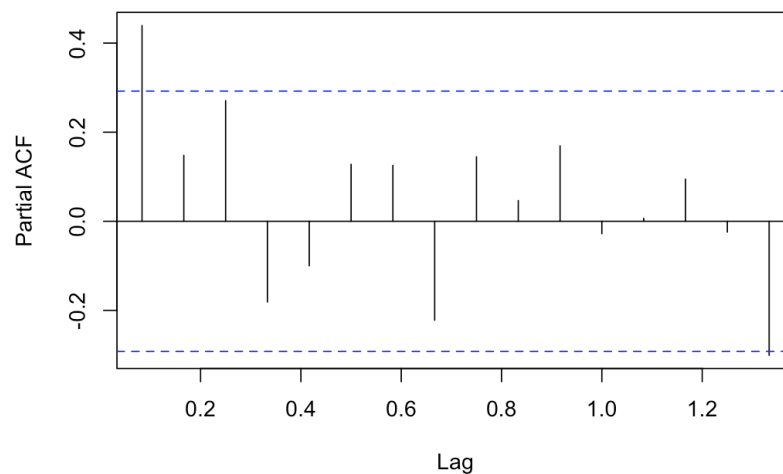


Figura 13. Diagnóstico de residuos del modelo ARIMA (0,1,1)(1,0,0)[12]

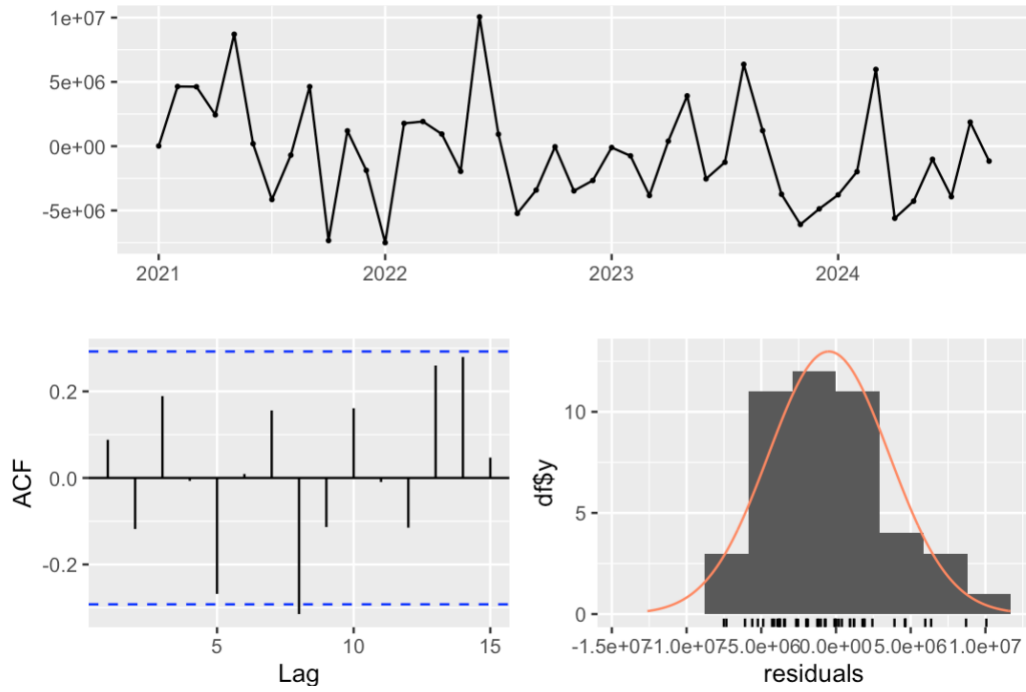


Figura 14. Autocorrelación (ACF) – SARIMA

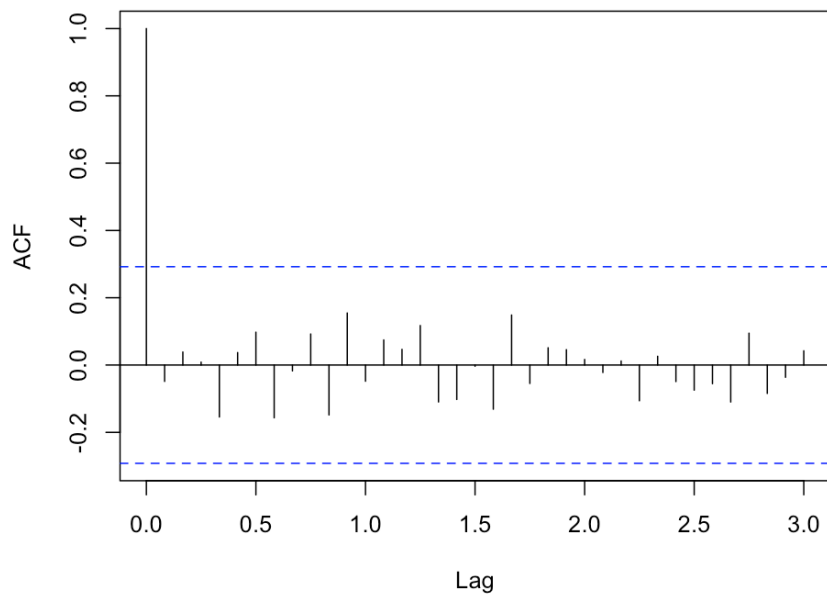
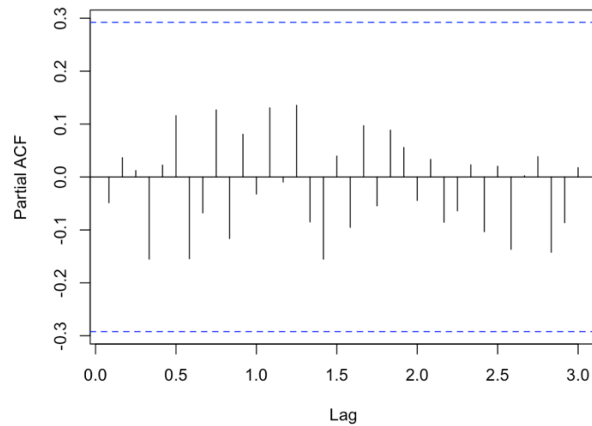
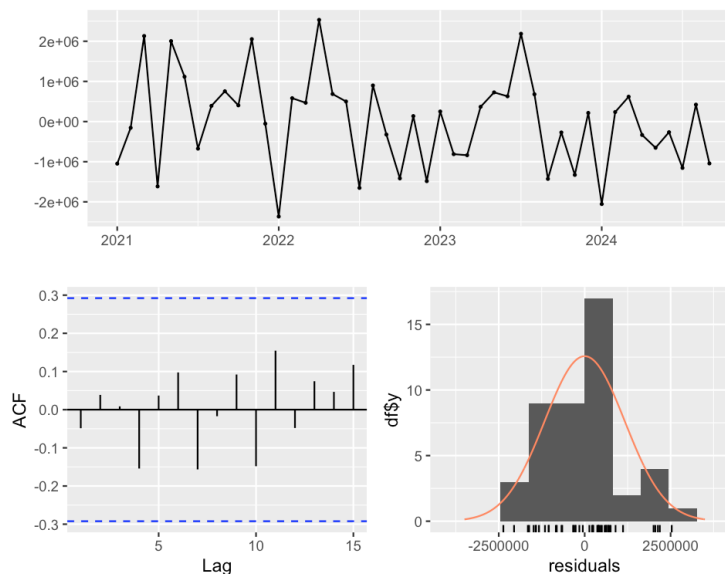


Figura 15. Autocorrelación Parcial (PACF) – SARIMA



Para determinar la significatividad estadística de los rezagos observados en los gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF), se utilizó el criterio estándar basado en intervalos de confianza del 95 %. Bajo el supuesto de ruido blanco, un rezago se considera significativo si su valor de autocorrelación supera en valor absoluto el umbral aproximado de $\pm \frac{2}{\sqrt{n}}$, donde n representa la cantidad de observaciones en la serie. Esta regla permite identificar rezagos cuya autocorrelación es estadísticamente distinta de cero, lo cual es útil para orientar la especificación de modelos de series temporales como ARIMA (Hyndman & Athanasopoulos, 2021).

Figura 16. Diagnóstico de residuos – SARIMA



Estos resultados permiten concluir que ambos modelos ofrecen un ajuste razonable para la serie mensual agregada, siendo el SARIMA una mejora natural del ARIMA al capturar adecuadamente la estacionalidad presente en la demanda. No obstante, ambos modelos funcionan como benchmarks robustos para comparar con métodos más complejos explorados posteriormente.

La elección de los modelos ARIMA y SARIMA responde a su capacidad comprobada para capturar dinámicas temporales con componentes de tendencia, ciclos y estacionalidad, comunes en contextos de demanda mensual. Ambos modelos ofrecen un marco interpretativo transparente, lo que resulta especialmente valioso para generar pronósticos comprensibles y auditables dentro de una organización industrial. En particular, SARIMA permite modelar explícitamente patrones estacionales como los que se observan en la demanda consolidada de Vidriotec, manteniendo una estructura flexible y sin requerimientos de grandes volúmenes de datos históricos. Esta versatilidad metodológica y su bajo costo computacional justifican su inclusión como modelos base de comparación frente a técnicas más complejas o menos interpretables, como los métodos de aprendizaje automático abordados posteriormente.

4.3.3. Aplicación a series individuales por SKU

Además del ajuste sobre la demanda mensual total, se desarrolló una versión desagregada de los modelos ARIMA y SARIMA, aplicada de forma individual a las series mensuales correspondientes a cada producto (SKU). Esta aproximación busca capturar patrones específicos en la dinámica de ventas de cada artículo, reconociendo que la heterogeneidad entre productos puede limitar significativamente la efectividad de modelos agregados.

El modelado se aplicó exclusivamente a aquellos productos que contaban con al menos 36 observaciones mensuales consecutivas. Este umbral mínimo de datos garantiza que cada serie tenga suficiente información para permitir un ajuste confiable del modelo, evitando sobreajustes o resultados poco robustos en productos con historial insuficiente.

Cada serie temporal fue estimada individualmente utilizando la función `auto.arima()` del paquete `forecast`, explorando dos variantes:

- ARIMA, sin componente estacional (seasonal = FALSE), adecuado para series sin estacionalidad marcada.
- SARIMA, con búsqueda automática de estructura estacional (seasonal = TRUE), capaz de capturar patrones cíclicos recurrentes, como los asociados a la estacionalidad anual.

En ambos casos, cada serie fue dividida en un conjunto de entrenamiento (enero 2021 a septiembre 2024) y un conjunto de prueba (octubre a diciembre 2024), sobre el cual se generaron pronósticos con horizonte de tres meses. El desempeño de los modelos fue evaluado a nivel SKU–mes mediante métricas de error como RMSE, MAPE, WAPE y R^2 . Asimismo, se consolidaron las predicciones por mes calendario para calcular métricas a nivel mensual agregado, lo que permitió comparar el desempeño de ambos enfoques también en escala global.

La decisión de modelar individualmente cada SKU se sustenta en la literatura especializada. Hyndman y Athanasopoulos (2021) sostienen que “cuando las series exhiben patrones sustancialmente diferentes, como efectos estacionales variables, tendencias o volatilidad, generalmente es preferible modelarlas de forma individual en lugar de intentar encontrar un modelo común que se ajuste a todas” (*Forecasting: Principles and Practice*). Esta recomendación resulta particularmente relevante en contextos de alta heterogeneidad como el de Vidriotec, donde distintos productos pueden presentar comportamientos de demanda claramente diferenciados en función de su segmento, uso, estacionalidad o canal de distribución.

El análisis detallado del desempeño de los modelos desagregados, así como su comparación con enfoques alternativos, se presenta en las secciones siguientes.

4.3.4. Random Forest

El marco teórico sobre Random Forest fue desarrollado previamente en la sección de Revisión de Literatura. Allí se describió su funcionamiento general como un método de agregación de múltiples árboles de decisión entrenados sobre subconjuntos aleatorios de los

datos (*bagging*), su robustez frente al sobreajuste y su capacidad para capturar relaciones no lineales (Breiman, 2001).

En el contexto del presente trabajo, se seleccionó Random Forest por su habilidad para modelar de manera flexible la demanda de Vidriotec, combinando variables categóricas (como SKU, cliente y provincia) y variables derivadas de la serie temporal (como rezagos y medias móviles). Su tolerancia al ruido y su capacidad para manejar datos de alta dimensionalidad lo convierten en un candidato natural para capturar la complejidad inherente a la predicción de demanda desagregada.

Metodológicamente, el modelo se implementó utilizando la librería `randomForest` en conjunto con el paquete `caret` en R. Se aplicó una estrategia de validación cruzada para series temporales (cross-validation mediante *time slicing*), garantizando la consistencia temporal entre conjuntos de entrenamiento y validación. Además, se llevó a cabo una búsqueda de hiperparámetros (grid search) optimizando el valor de `mtry` para maximizar el desempeño predictivo.

El proceso de entrenamiento, ajuste de hiperparámetros y validación del modelo se describe con mayor detalle en las secciones posteriores.

La elección de Random Forest como modelo de predicción responde a su reconocida capacidad para abordar problemas con estructuras complejas y no lineales, como los presentes en la demanda comercial de Vidriotec. Su flexibilidad para incorporar múltiples tipos de variables —categóricas, temporales y de interacción— permite capturar relaciones que modelos más estructurados no logran representar adecuadamente. Además, su robustez frente a la multicolinealidad y a valores atípicos lo convierte en una herramienta confiable para contextos de alta variabilidad e intermitencia en los datos. Estas características, sumadas a su buen desempeño empírico reportado en aplicaciones logísticas y de retail, justifican su incorporación como benchmark representativo de los enfoques de aprendizaje automático no paramétricos.

4.3.5. XGBoost (Extreme Gradient Boosting)

El modelo XGBoost fue abordado previamente en la sección de Revisión de Literatura, donde se explicó su funcionamiento basado en técnicas de *gradient boosting*, su capacidad para corregir errores iterativamente mediante el entrenamiento secuencial de árboles, y sus mecanismos de regularización para prevenir el sobreajuste (Chen & Guestrin, 2016; Friedman, 2001).

En el marco del presente trabajo, XGBoost se eligió por su reconocida eficacia en la modelización de datos tabulares, su capacidad para capturar relaciones no lineales complejas, y su robustez frente al sobreajuste. Estas características lo convierten en una herramienta adecuada para abordar la predicción de demanda de Vidriotec, combinando variables categóricas codificadas y variables derivadas de la serie temporal (como rezagos y medias móviles).

Metodológicamente, el modelo se implementó utilizando la librería *xgboost* de R. Se realizó una búsqueda de hiperparámetros (*grid search*) evaluando distintas combinaciones de tasas de aprendizaje (*eta*) y profundidades máximas de los árboles (*max_depth*). Cada combinación fue validada mediante validación cruzada de cinco particiones (5-fold cross-validation), empleando además un esquema de *early stopping* para prevenir el sobreentrenamiento.

El modelo final se entrenó utilizando los hiperparámetros óptimos encontrados y la cantidad de iteraciones correspondiente al mínimo error de validación. El proceso de entrenamiento, ajuste y evaluación de desempeño se desarrolla en detalle en las secciones siguientes.

La elección de XGBoost también se justifica por su solidez en contextos con estructuras de datos complejas, como el presente caso. Su mecanismo de boosting secuencial, junto con técnicas de control de complejidad como la profundidad máxima de los árboles, la tasa de aprendizaje, y el muestreo aleatorio de observaciones y variables, permite al modelo generalizar correctamente sin incurrir en sobreajuste (Chen & Guestrin, 2016; Nielsen, 2016). Estas características lo convierten en una alternativa especialmente adecuada para

capturar las relaciones no lineales y las interacciones presentes en la demanda de Vidriotec, manteniendo al mismo tiempo un equilibrio entre precisión y robustez.

4.3.5. Prophet

El modelo Prophet fue desarrollado originalmente por Taylor y Letham (2017) como una herramienta de predicción de series temporales pensada para escenarios donde se presentan patrones de estacionalidad fuerte, tendencias cambiantes y eventos especiales. Prophet se basa en un enfoque aditivo o multiplicativo, descomponiendo la serie temporal en componentes de tendencia, estacionalidad y feriados/eventos.

En el presente trabajo, Prophet se seleccionó como alternativa de modelado debido a su flexibilidad para capturar estructuras de series temporales relativamente complejas de forma automática, y por su facilidad de ajuste en contextos de negocios donde se dispone de datos agregados de ventas. Su capacidad para gestionar cambios de tendencia a través de changepoints automáticos resultaba atractiva dado el comportamiento estacional mixto observado en la demanda de Vidriotec.

Metodológicamente, Prophet fue implementado utilizando la librería prophet en R. Se entrenó el modelo utilizando datos de ventas mensuales agregados a nivel empresa para el período enero 2021 – septiembre 2024. Para optimizar su desempeño, se realizó una búsqueda sistemática sobre dos hiperparámetros principales: la escala de sensibilidad a cambios de tendencia (`changepoint_prior_scale`) y el modo de estacionalidad (`additive` o `multiplicative`). Se evaluaron múltiples combinaciones mediante un procedimiento de tuning interno, seleccionando aquellas que minimizaran los errores de pronóstico.

El entrenamiento se realizó exclusivamente sobre datos históricos disponibles hasta septiembre 2024, respetando la secuencia temporal para evitar filtraciones de información futura (`data leakage`). Posteriormente, se generaron predicciones sobre el horizonte octubre–diciembre 2024, cuyo desempeño se analiza detalladamente en el capítulo de resultados.

La elección de Prophet en este trabajo respondió a su propuesta metodológica innovadora para modelar series con estacionalidad y cambios estructurales. Si bien su aplicación se

justificaba inicialmente por la presencia de estacionalidad mixta en la demanda de Vidriotec, los resultados obtenidos indicaron un desempeño significativamente inferior al de otros modelos, con errores altos y bajo poder explicativo. No obstante, su inclusión permitió evaluar sus limitaciones en contextos de alta variabilidad e intermitencia de la demanda, aportando valor como benchmark adicional dentro del análisis comparativo.

4.3.6 Modelo LSTM

El modelo LSTM es una variante especializada de las redes neuronales recurrentes (RNN), desarrollada para captar dependencias temporales complejas y prolongadas en series de datos (Hochreiter y Schmidhuber, 1997). En el contexto de este trabajo, se implementó como modelo exploratorio complementario para evaluar su desempeño en la predicción mensual de la demanda en Vidriotec.

Metodológicamente, la implementación se realizó en R utilizando la librería keras, con una arquitectura simple formada por una capa LSTM de 50 unidades y una capa densa final con salida lineal. Se emplearon variables temporales derivadas de rezagos (lags) de 1, 2, 3, 6 y 12 meses como inputs. Los datos fueron normalizados, y el entrenamiento se realizó hasta septiembre de 2024, utilizando validación interna y early stopping para evitar sobreajuste.

Los resultados indicaron un desempeño predictivo significativamente inferior al de otros modelos evaluados (XGBoost, Random Forest), principalmente debido al volumen limitado de datos históricos disponibles y a la frecuencia mensual, insuficiente para aprovechar completamente las capacidades de aprendizaje profundo.

Su inclusión se justifica como benchmark adicional y exploración preliminar para futuras investigaciones que cuenten con series temporales más extensas o con mayor granularidad (e.g., datos diarios o semanales).

4.4. Entrenamiento y validación de modelos

Esta sección tiene como objetivo describir las técnicas y procedimientos utilizados para entrenar y validar los modelos desarrollados en este trabajo, asegurando la integridad temporal de los datos y la robustez de las métricas de desempeño.

4.4.1. División temporal de los datos

Dado que el objeto de estudio corresponde a series de tiempo de demanda, se adoptó una estrategia de división de los datos que respeta estrictamente la secuencia cronológica de las observaciones. Esto garantiza que el entrenamiento de los modelos se realice únicamente con información disponible hasta cada punto de predicción, evitando cualquier filtración de datos futuros (*data leakage*).

El conjunto de entrenamiento incluyó las observaciones registradas entre enero de 2021 y septiembre de 2024 inclusive, mientras que el conjunto de prueba se conformó con las observaciones correspondientes al trimestre octubre–diciembre de 2024. Esta división temporal fue definida considerando tanto la disponibilidad de datos como las prácticas de planificación operativa de la empresa, que suele trabajar con horizontes de proyección de uno a tres meses.

La adopción de esta estrategia permite simular un escenario realista de pronóstico, donde las decisiones deben tomarse basándose exclusivamente en información histórica. Como señalan Hyndman y Athanasopoulos (2018), en problemas de series de tiempo es esencial preservar el orden temporal para obtener evaluaciones de desempeño representativas y aplicables en entornos productivos reales.

4.4.2. Validación específica por tipo de modelo

El procedimiento de validación se adaptó a las características específicas de cada tipo de modelo desarrollado en este trabajo, respetando en todos los casos la estructura temporal de los datos para evitar cualquier tipo de filtración de información futura (*data leakage*).

En el caso de los modelos ARIMA, tanto en su versión univariada como por SKU, se realizó un ajuste tradicional sobre el conjunto de entrenamiento, seleccionando automáticamente los mejores parámetros (órdenes p , d , q) mediante criterios de información como el AIC (Akaike Information Criterion) y el BIC (Bayesian Information Criterion). Esta metodología, ampliamente aceptada en la literatura (Box, Jenkins & Reinsel, 1994), permite encontrar configuraciones parciales óptimas sin necesidad de validación cruzada explícita, respetando la lógica secuencial propia de las series de tiempo.

Para el modelo Random Forest, se utilizó un esquema de validación cruzada adaptado para series temporales mediante *time slicing*. En este enfoque, los datos se dividieron respetando su secuencia cronológica, creando ventanas de entrenamiento y validación móviles. La validación cruzada de este tipo permite evaluar el desempeño del modelo a lo largo del tiempo, asegurando que la información futura no influya en el entrenamiento. Adicionalmente, se realizó una búsqueda sistemática en grilla (*grid search*) sobre el hiperparámetro *mtry* (número de variables consideradas en cada división del árbol), seleccionando la configuración que optimizara el error de validación.

El modelo XGBoost se validó mediante un esquema de validación cruzada de cinco particiones (*5-fold cross-validation*), asegurando que los datos de entrenamiento se utilizaran de forma eficiente para evaluar la capacidad de generalización del modelo. La validación cruzada consiste en dividir el conjunto de entrenamiento en cinco subconjuntos de igual tamaño; en cada iteración, el modelo se entrena sobre cuatro particiones y se evalúa sobre la partición restante, rotando sucesivamente hasta cubrir todos los subconjuntos. Esta metodología permite obtener una estimación más robusta del error de predicción, mitigando la dependencia de una única partición de datos.

Para evitar el sobreajuste, se aplicó la técnica de *early stopping*, interrumpiendo el proceso de entrenamiento cuando el error de validación dejaba de mejorar tras un número determinado de iteraciones consecutivas.

Complementariamente, se llevó a cabo una búsqueda en grilla (*grid search*) de hiperparámetros, explorando combinaciones de valores de la tasa de aprendizaje (*eta*) y la profundidad máxima de los árboles (*max_depth*). El *grid search* consiste en entrenar múltiples modelos con diferentes combinaciones de hiperparámetros, evaluándolos bajo un mismo esquema de validación, con el objetivo de seleccionar la configuración que minimice el error de validación.

Esta estrategia combinada de validación cruzada, control temprano de sobreajuste y optimización de hiperparámetros permitió fortalecer la capacidad predictiva del modelo sobre nuevos datos, alineándose con las mejores prácticas de modelado supervisado.

Finalmente, el modelo Prophet se ajustó automáticamente sobre el conjunto de entrenamiento, descomponiendo la serie temporal en sus componentes de tendencia, estacionalidad y feriados. Dado que Prophet gestiona internamente la separación de información pasada y futura, no se requirió implementar esquemas de validación cruzada adicionales.

4.4.3. Criterios de selección y precauciones metodológicas

La selección de los mejores modelos y configuraciones se basó en la minimización del error cuadrático medio (RMSE) sobre los conjuntos de validación o testeo. Esta métrica fue priorizada debido a su capacidad para penalizar en mayor medida los errores de gran magnitud, característica especialmente relevante en el contexto de la planificación operativa, donde desviaciones significativas pueden derivar en sobrecostos logísticos, quiebres de inventario o pérdidas de oportunidad de venta.

Se tomaron medidas estrictas para preservar la integridad temporal de los datos a lo largo del proceso de modelado. Las variables predictoras generadas, tales como rezagos, medias móviles y codificaciones categóricas, fueron construidas exclusivamente a partir de información disponible hasta cada punto de predicción, evitando así cualquier riesgo de data leakage. Asimismo, la división entre conjuntos de entrenamiento y testeo respetó rigurosamente la secuencia cronológica de las observaciones, y los esquemas de validación cruzada implementados fueron diseñados para mantener la independencia temporal entre entrenamiento y validación. Como destacan Hyndman y Athanasopoulos (2018), en problemas de series de tiempo, preservar el orden temporal de los datos es esencial para obtener evaluaciones de desempeño representativas.

Adicionalmente, se priorizó la adopción de metodologías de entrenamiento y validación que fueran replicables en entornos empresariales. El uso de técnicas estandarizadas, como la validación cruzada y la búsqueda sistemática de hiperparámetros mediante grid search, facilita la futura implementación y monitoreo de los modelos por parte de equipos operativos sin requerir herramientas complejas o conocimientos avanzados de modelado estadístico.

Finalmente, aunque las metodologías adoptadas buscan maximizar la robustez de la evaluación, se reconoce que métricas como el RMSE pueden ser sensibles a valores atípicos, y que los esquemas de validación cruzada aplicados podrían no capturar completamente el impacto de cambios estructurales imprevistos en la demanda. Estas limitaciones fueron consideradas en el análisis de los resultados y serán discutidas en las secciones finales del trabajo.

4.5 Evaluación del desempeño de los modelos

La evaluación del desempeño de los modelos desarrollados en este trabajo se realizó utilizando un conjunto de métricas complementarias, seleccionadas con el objetivo de obtener una visión integral del ajuste y la capacidad predictiva de los algoritmos. Las métricas empleadas fueron el error cuadrático medio (RMSE), el error porcentual absoluto medio (MAPE), el error absoluto ponderado (WAPE) y el coeficiente de determinación ajustado (R^2 ajustado).

La elección de estas métricas responde a la necesidad de evaluar el desempeño de los modelos desde múltiples perspectivas. Mientras que el RMSE penaliza fuertemente los errores de gran magnitud, el MAPE proporciona una interpretación relativa intuitiva en porcentaje, el WAPE ofrece una medida más estable en presencia de productos de baja rotación, y el R^2 ajustado permite medir la proporción de variabilidad explicada por el modelo.

Asimismo, se evaluó el desempeño de los modelos en distintos niveles de agregación: a nivel SKU-mes, a nivel mensual agregado y a nivel total agregado (en el caso particular del modelo ARIMA univariado). Esto permitió analizar la precisión de las predicciones tanto en el detalle operativo como en una perspectiva más consolidada.

La combinación de múltiples métricas y niveles de análisis facilita un diagnóstico robusto del desempeño de los modelos y busca asegurar su aplicabilidad práctica en el entorno de planificación operativa de la empresa.

4.5.1 Métricas de evaluación (RMSE, MAPE, WAPE, R² ajustado)

A continuación, se describen las principales métricas utilizadas para evaluar el desempeño de los modelos:

- Error cuadrático medio (RMSE):

El RMSE se define como la raíz cuadrada del promedio de los errores cuadráticos entre los valores predichos y los valores observados. Su fórmula es:

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

donde y_i son los valores reales \hat{y}_i los valores predichos. El RMSE penaliza en mayor medida los errores de gran magnitud, por lo que es particularmente relevante en contextos donde grandes desvíos generan impactos operativos significativos.

- Error porcentual absoluto medio (MAPE):

El MAPE mide el error absoluto promedio expresado como porcentaje del valor real. Su fórmula es:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Esta métrica facilita una interpretación intuitiva de los errores, aunque puede ser sensible a valores reales cercanos a cero, donde los porcentajes se distorsionan.

Error absoluto ponderado (WAPE):

El WAPE calcula el error absoluto total en relación al volumen real total. Su fórmula es:

$$\text{WAPE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n y_i} \times 100$$

El WAPE es menos sensible a valores pequeños que el MAPE, y resulta particularmente útil en entornos con alta heterogeneidad de volúmenes entre productos.

Coefficiente de determinación (R^2):

El R^2 representa la proporción de la variabilidad total de los datos que es explicada por el modelo. Su fórmula es:

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

Donde \bar{y} es el valor medio de las observaciones reales.

En la evaluación de los modelos de este trabajo se utiliza el R^2 *ajustado* para penalizar la inclusión de variables irrelevantes. Es particularmente útil en modelos con múltiples predictores, ya que evita la sobreestimación del poder explicativo del modelo. Su fórmula es:

$$R_{ajustado}^2 = 1 - \left(\frac{(1 - R^2)(n - 1)}{n - k - 1} \right)$$

Donde n es el número total de observaciones y k es el número de predictores del modelo.

4.5.2 Niveles de agregación utilizados

El desempeño de los modelos fue evaluado en distintos niveles de agregación para capturar tanto la precisión en el detalle operativo como en una visión más estratégica:

- Nivel SKU–mes: Evaluación de la cantidad de unidades predichas para cada combinación específica de producto y mes. Este nivel es fundamental para analizar el comportamiento individual de cada SKU, permitiendo detectar patrones de sobreestimación o subestimación específicos.
- Nivel mensual agregado: Evaluación del total de unidades vendidas por mes, sumando todos los SKUs. Este nivel permite observar la capacidad de los modelos para capturar el comportamiento agregado de la demanda en el tiempo.

- Nivel total agregado (solo para ARIMA univariado): Evaluación del total de unidades vendidas acumuladas en el período de testeo, utilizado como benchmark para modelos que operan sobre series de demanda total sin distinción de SKU.

La combinación de estos niveles de evaluación permite analizar la robustez y consistencia de los modelos en distintos escenarios de uso práctico, alineándose con las necesidades reales de planificación de la empresa.

Con el objetivo de garantizar la robustez y aplicabilidad de los modelos desarrollados, este capítulo presentó en detalle la metodología implementada, desde la preparación de los datos hasta la definición de los criterios de evaluación. Se describieron los procedimientos de entrenamiento, las técnicas de validación adoptadas, las métricas seleccionadas para medir el desempeño de los modelos y los niveles de agregación utilizados para su análisis.

La metodología adoptada busca no solo maximizar la precisión predictiva, sino también asegurar su replicabilidad práctica en el contexto operativo de la empresa.

5. Resultados

5.1 Descripción general del enfoque de evaluación

Con el objetivo de evaluar el desempeño predictivo de los distintos modelos desarrollados, se implementó un esquema de evaluación basado en dos niveles de análisis: a nivel desagregado por combinación SKU–mes y a nivel agregado mensual consolidado. Para el caso particular del modelo ARIMA univariado, la evaluación se realizó a nivel de total agregado de unidades vendidas, dado que el modelo fue ajustado exclusivamente sobre la serie temporal total.

Las métricas de evaluación seleccionadas fueron el Error Cuadrático Medio (RMSE), el Error Absoluto Porcentual Medio Ponderado (WAPE) y el Coeficiente de Determinación (R^2). El RMSE se priorizó por su capacidad de penalizar errores de gran magnitud, atributo especialmente relevante en contextos de planificación operativa donde desvíos importantes pueden derivar en costos logísticos significativos. El WAPE se utilizó como métrica complementaria de interpretación relativa, más robusta que el MAPE en presencia de

demandas bajas, mientras que el R^2 permitió evaluar la proporción de variabilidad explicada por los modelos.

La evaluación se llevó a cabo respetando estrictamente la secuencia temporal de los datos: los modelos fueron entrenados sobre observaciones disponibles hasta septiembre de 2024 y testeados sobre el horizonte octubre–diciembre de 2024. Este enfoque garantizó que no existiera filtración de información futura (data leakage), asegurando así la validez y replicabilidad de los resultados obtenidos.

En las secciones siguientes se presentan los resultados de cada modelo en función de estas métricas y niveles de agregación definidos.

5.2 Resultados por modelos

5.2.1 ARIMA y SARIMA

En primer lugar, se desarrollaron modelos clásicos de series temporales para abordar el problema de predicción de la demanda. Se aplicaron dos enfoques principales: (i) un modelo ARIMA univariado sobre la serie agregada total de unidades vendidas, y (ii) múltiples modelos ARIMA ajustados individualmente por SKU.

En el caso del modelo aplicado a la serie mensual agregada, la especificación seleccionada por `auto.arima()` fue un **SARIMA(0,1,1)(1,0,0)[12]**, ajustado mediante búsqueda exhaustiva (sin `stepwise` ni `approximation`) sobre los datos de enero 2021 a septiembre 2024. Esta configuración incluye un término de media móvil no estacional (MA(1)) y un componente autorregresivo estacional de orden 1, con periodicidad mensual. Los coeficientes estimados fueron **ma1 = -0.803** y **sar1 = 0.329**, ambos con errores estándar bajos, y el modelo obtuvo un **AICc de 1473.25**, siendo el menor entre las alternativas evaluadas.

A su vez, se extendieron ambos enfoques incorporando estacionalidad a través de modelos SARIMA, con el objetivo de evaluar si la incorporación explícita de componentes estacionales mejoraba el desempeño predictivo.

La siguiente tabla presenta las métricas de desempeño obtenidas por cada modelo:

Figura 17. Demanda Real y predicción ARIMA - Año 2024

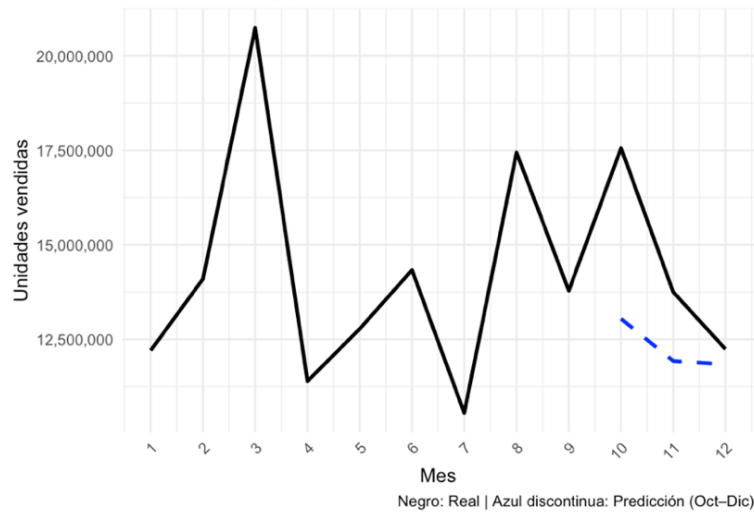


Tabla 1. Métricas de desempeño - Modelos ARIMA y SARIMA

Nivel de Evaluación	Métrica	ARIMA	SARIMA
SKU-mes (individual)	RMSE	856531.93	839547.62
SKU-mes (individual)	MAPE (%)	143.50	696.44
SKU-mes (individual)	WAPE (%)	41.66	41.33
SKU-mes (individual)	R ² ajustado	0.37	0.40
Mensual agregado	RMSE	1965137.19	1619392.07
Mensual agregado	MAPE (%)	9.77	8.87
Mensual agregado	WAPE (%)	11.03	9.61
Mensual agregado	R ² ajustado	0.09	0.37

Se observa que, en el caso de los modelos ajustados por SKU, tanto ARIMA como SARIMA presentaron limitaciones al enfrentar la alta dispersión e intermitencia propias de la demanda desagregada. En este nivel, el modelo SARIMA logró una ligera mejora en el coeficiente de determinación ajustado (R^2 ajustado = 0,40 vs. 0,37 en ARIMA) y una leve reducción del WAPE (41,33 % frente a 41,66 %), aunque presentó un MAPE considerablemente más alto (696,44 %), lo cual puede atribuirse al bajo volumen de demanda en muchos productos, donde pequeños errores absolutos generan grandes desvíos porcentuales. Al consolidar las predicciones por mes calendario (mensual agregado por SKU), el modelo SARIMA mostró mejoras más significativas frente a ARIMA. Se observó una reducción del RMSE (1.619.392 frente a 1.965.137), del MAPE (8,87 % frente a 9,77 %) y del WAPE (9,61 % frente a 11,03 %), así como un aumento considerable del coeficiente de

determinación ajustado (R^2 ajustado = 0,37 frente a 0,09). Estos resultados reflejan una mayor capacidad del modelo SARIMA para capturar patrones estacionales y ofrecer pronósticos más precisos en términos operativos, sin requerir transformaciones adicionales sobre los datos.

El test de Ljung–Box sobre los residuos del modelo SARIMA a nivel agregado arrojó un p-value de 0,6917, indicando una ausencia clara de autocorrelación residual y confirmando que el modelo fue capaz de capturar adecuadamente la estructura temporal de la serie.

En términos generales, aunque los modelos ARIMA presentaron una base razonable como línea de base metodológica, la incorporación del componente estacional en SARIMA permitió una mejora consistente en las métricas más relevantes para la toma de decisiones, especialmente en la planificación mensual consolidada.

Estos resultados, sin embargo, también evidencian que incluso con ajustes específicos por SKU y el tratamiento de estacionalidad, la complejidad de la demanda en Vidriotec requiere enfoques más robustos y flexibles. Esto justifica la exploración de modelos de aprendizaje automático, desarrollada en las secciones siguientes del trabajo.

5.2.2 Random Forest

Figura 18. Demanda Real y predicción Random Forest - Año 2024

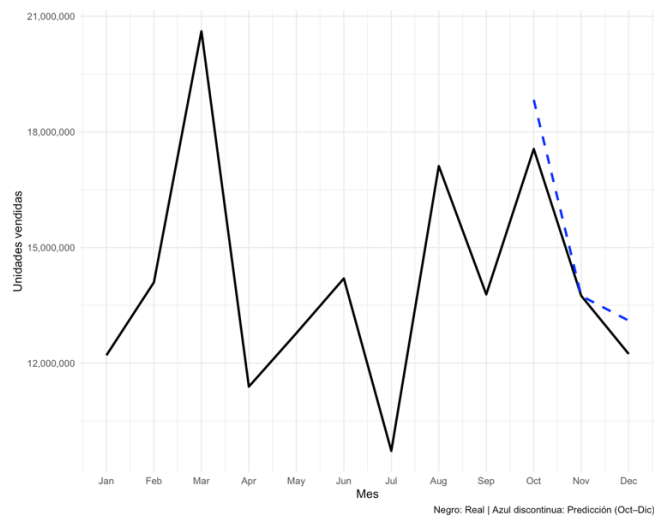


Tabla 2. Métricas de desempeño – Modelo Random Forest

Nivel de Evaluación	Métrica	Valor
SKU-mes (individual)	RMSE	237015.55
SKU-mes (individual)	MAPE (%)	702.67
SKU-mes (individual)	WAPE (%)	78.74
SKU-mes (individual)	R ² ajustado	0.53
Mensual agregado	RMSE	889192.11
Mensual agregado	MAPE (%)	4.78
Mensual agregado	WAPE (%)	4.92
Mensual agregado	R ² ajustado	0.82

El desempeño del modelo Random Forest se evaluó utilizando las métricas RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error) y WAPE (Weighted Absolute Percentage Error), tanto a nivel desagregado por SKU-mes como de forma agregada por mes calendario. Los resultados obtenidos evidencian un marcado contraste entre ambas escalas de análisis.

A nivel individual (SKU-mes), el modelo presentó un RMSE de 237.016 unidades, un MAPE de 702,7 % y un WAPE de 78,7 %. Estos valores reflejan un desempeño inadecuado para el pronóstico de demanda a nivel SKU-mes —con errores porcentuales extraordinariamente elevados—, por lo que el uso de Random Forest en este nivel de desagregación no es recomendable.

En contraste, al evaluar la capacidad predictiva de forma agregada por mes calendario, el RMSE registrado fue de 889.192 unidades, con un MAPE de 4,8 % y un WAPE de 4,9 %. Estos indicadores muestran que, aunque el modelo falla a nivel granular, logra capturar adecuadamente las fluctuaciones de la demanda total consolidada, resultando útil para aplicaciones de planificación operativa y proyección de volúmenes globales.

En síntesis, Random Forest no es válido para pronósticos desagregados por SKU-mes debido a sus altos niveles de error, pero sí puede considerarse una herramienta adecuada para estimar la demanda mensual agregada en entornos industriales como el de Vidriotec.

5.2.3 XGBoost

Figura 19. XGBoost con Cross Validation y Grid Search

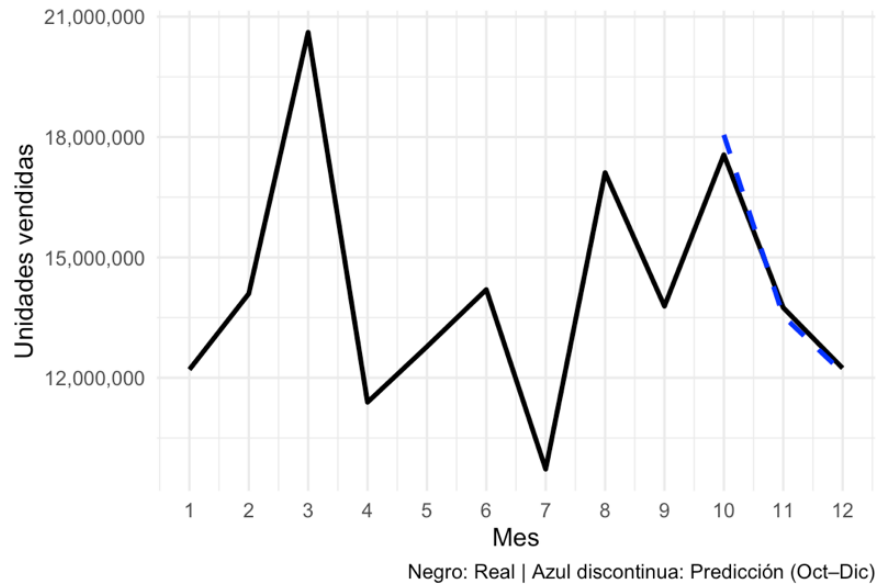


Tabla 3. Métricas de desempeño – Modelo XGBoost

Tipo de evaluación	Métrica	Valor
SKU-mes (individual)	RMSE	76411.93
SKU-mes (individual)	MAPE (%)	145.91
SKU-mes (individual)	WAPE (%)	18.02
SKU-mes (individual)	R ² ajustado	0.94
Mensual agregado	RMSE	491812.28
Mensual agregado	MAPE (%)	2.77
Mensual agregado	WAPE (%)	2.94
Mensual agregado	R ² ajustado	0.93

El modelo XGBoost fue evaluado a partir de métricas aplicadas tanto a nivel micro (SKU-mes) como macro (demanda total mensual), con el objetivo de capturar su capacidad predictiva en distintas escalas de análisis. Esta doble perspectiva permite identificar tanto los desafíos en la estimación individualizada por producto como el desempeño general en la predicción de volúmenes agregados. A nivel desagregado (SKU-mes), el modelo presentó un RMSE de 76.412 unidades, un

MAPE de 145,9 % y un WAPE de 18,0 %, alcanzando además un coeficiente de determinación (R^2 ajustado) de 0,94. Si bien el MAPE elevado refleja el impacto de combinaciones con volúmenes bajos de demanda, el valor de WAPE obtenido puede considerarse muy satisfactorio en términos operativos, dado que se sitúa dentro de los rangos considerados de buen desempeño en entornos de alta variabilidad de demanda. De acuerdo con Hyndman y Athanasopoulos (2021), en contextos de alta dispersión de la demanda, errores ponderados (como el WAPE) inferiores al 20 % son indicativos de un desempeño predictivo robusto, lo cual refuerza la solidez de los resultados obtenidos en este trabajo. El alto valor de R^2 ajustado sugiere, además, que el modelo logra capturar adecuadamente la estructura general del comportamiento de la demanda a nivel micro. Por otro lado, el desempeño del modelo mejora significativamente al analizar la predicción de la demanda mensual total. En este nivel agregado, XGBoost alcanzó un RMSE de 491.812 unidades, un MAPE de 2,8 % y un WAPE de 2,9 %, con un R^2 ajustado de 0,93, lo que indica una notable capacidad para capturar la dinámica global de la demanda y reproducir con precisión sus fluctuaciones a lo largo del tiempo. En síntesis, aunque XGBoost presenta ciertas limitaciones al modelar comportamientos puntuales de productos de baja rotación o alta volatilidad, se destaca como una herramienta robusta y confiable para el pronóstico agregado de demanda, con alto potencial para asistir en la planificación comercial, operativa y estratégica de la empresa.

5.2.4 Prophet

Se implementó el modelo Prophet, desarrollado por Facebook, por su capacidad para capturar de forma flexible componentes de tendencia y estacionalidad en series temporales. El modelo fue entrenado individualmente para cada SKU, considerando una frecuencia mensual. Asimismo, se llevó a cabo una optimización de hiperparámetros mediante una búsqueda en grilla, explorando combinaciones de `changepoint_prior_scale` y `seasonality_mode`, con el objetivo de minimizar el error de predicción. La evaluación del desempeño se realizó tanto a nivel SKU–mes como en el agregado mensual. A continuación, se presenta la tabla con las métricas obtenidas para el período octubre–diciembre de 2024.

Tabla 4. Métricas de desempeño – Modelo Prophet

nivel	changepoint	seasonality	RMSE	MAPE	WAPE	R ² ajustado
Mensual agregado	0.05	additive	7221450	37.9	32.4	-9.5
Mensual agregado	0.10	additive	7239019	38.0	32.5	-9.6
Mensual agregado	0.10	multiplicative	5473279	36.2	32.3	-5.1
Mensual agregado	0.05	multiplicative	6083285	40.5	37.1	-6.7
Mensual agregado	0.30	additive	14728273	77.2	65.9	-42.5
Mensual agregado	0.30	multiplicative	92804597	571.0	544.0	-1715.0
SKU-mes (individual)	0.05	multiplicative	1781226	34.5	105.0	-1.9
SKU-mes (individual)	0.10	multiplicative	1789440	-278.0	106.0	-2.0
SKU-mes (individual)	0.05	additive	2852088	602.0	161.0	-6.3
SKU-mes (individual)	0.10	additive	2856336	602.0	161.0	-6.3
SKU-mes (individual)	0.30	additive	4304392	1040.0	199.0	-15.8
SKU-mes (individual)	0.30	multiplicative	25466555	8219.0	647.0	-577.0

Estos resultados evidencian las limitaciones del modelo Prophet en contextos con alta volatilidad, dispersión y falta de estacionalidad marcada. Tal como advierten sus creadores (Taylor & Letham, 2017), Prophet tiende a fallar en escenarios con estructuras de datos poco definidas, lo cual ha sido reafirmado por estudios recientes (Kwarteng & Andreevich, 2024) en aplicaciones con series cortas o altamente irregulares.

En conclusión, si bien Prophet resulta una herramienta versátil y de rápida implementación, en este caso particular no logró capturar adecuadamente la dinámica de la demanda de Vidriotec. Su bajo desempeño tanto a nivel agregado como por SKU motiva la exploración de enfoques alternativos, particularmente modelos basados en aprendizaje automático, que puedan incorporar múltiples variables explicativas y mayor capacidad para adaptarse a entornos de alta complejidad.

5.2.5. Redes Neuronales LSTM

En línea con la exploración de enfoques avanzados de modelado predictivo, se implementó un modelo de redes neuronales recurrentes del tipo Long Short-Term Memory (LSTM) con el objetivo de evaluar su desempeño en el pronóstico de la demanda de Vidriotec. La elección de esta arquitectura se fundamentó en su reconocida capacidad para modelar secuencias temporales complejas y capturar dependencias de largo plazo en series de datos (Hochreiter & Schmidhuber, 1997).

El modelo fue entrenado inicialmente a partir de los datos agregados a nivel mensual. Posteriormente, se adaptó el enfoque a datos diarios para incrementar el volumen de observaciones y permitir un mayor aprovechamiento de las capacidades secuenciales del

LSTM. En ambos casos, se utilizaron variables de rezagos y medias móviles como entradas, respetando estrictamente la estructura temporal para evitar filtraciones de información futura.

Los resultados obtenidos, sin embargo, no fueron satisfactorios. A nivel agregado, el modelo presentó un error cuadrático medio (RMSE) elevado, altos valores de MAPE y WAPE, y un coeficiente de determinación R^2 negativo. Este desempeño refleja una limitada capacidad del LSTM para generalizar patrones de la demanda en el contexto específico analizado.

Se identifican varias posibles causas de este resultado: en primer lugar, el volumen de datos disponibles —aún considerando la granularidad diaria— resulta relativamente bajo para el entrenamiento efectivo de redes neuronales profundas, las cuales tienden a requerir grandes cantidades de datos para evitar el sobreajuste (Goodfellow, Bengio & Courville, 2016). En segundo lugar, la volatilidad inherente a la demanda industrial, sumada a la ausencia de información externa enriquecedora (como campañas comerciales, variables meteorológicas o datos económicos), dificulta la identificación de patrones robustos que el LSTM pueda aprender.

A pesar de estos resultados, la inclusión de este experimento en el trabajo resulta valiosa, dado que permite contrastar empíricamente la efectividad de técnicas tradicionales de machine learning frente a métodos de deep learning en contextos reales de negocio. Asimismo, deja abierta la puerta para futuras investigaciones que, disponiendo de mayores volúmenes de datos o de variables explicativas adicionales, podrían reconsiderar el uso de redes neuronales avanzadas para mejorar la precisión de los pronósticos.

5.3 Comparación de desempeño entre modelos

Tabla 5. Comparación entre modelos - Pronóstico Mensual Agregado

Modelo	RMSE	WAPE (%)	R^2 ajustado
ARIMA Mensual Agregado	1965137	11.03	0.09
SARIMA Mensual Agregado	1619392	9.61	0.37
Random Forest Mensual Agregado	889192	4.92	0.82
XGBoost Mensual Agregado	491812	2.94	0.93
Prophet Mensual Agregado	7221450	32.40	-9.50

Tabla 6. Comparación entre modelos - Pronóstico SKU-mes

Modelo	RMSE	WAPE (%)	R² ajustado
ARIMA SKU-mes	856532	41.66	0.37
SARIMA SKU-mes	839548	41.33	0.40
Random Forest SKU-mes	237016	78.74	0.53
XGBoost SKU-mes	76412	18.02	0.94
Prophet SKU-mes	1781226	105.00	-1.90

Las Tablas 5 y 6 presentan los resultados comparativos de los modelos en los niveles de demanda mensual agregada y SKU-mes, respectivamente. En ambos casos, XGBoost consigue el mejor desempeño: alcanza un RMSE de 491.812 y un WAPE de 2,94 % con un R² ajustado de 0,93 en la predicción agregada, y un RMSE de 76.412, WAPE de 18,02 % y R² ajustado de 0,94 en SKU-mes, superando ampliamente al resto de los modelos evaluados. Random Forest se ubica en un punto intermedio. A nivel SKU-mes presenta un RMSE bajo (237.017) y un R² ajustado moderado (0,53), aunque con un WAPE elevado (78,74 %) que limita su aplicabilidad en análisis desagregados. En cambio, en la predicción mensual consolidada logra un buen desempeño (WAPE 4,92 %; R² ajustado 0,82), convirtiéndose en una opción sólida para estimaciones globales. Entre los modelos clásicos de series temporales, SARIMA mostró mejoras consistentes respecto a ARIMA, tanto en el modelado por SKU como en la demanda mensual. A nivel SKU-mes, ambos modelos presentaron limitaciones ante la alta intermitencia, pero SARIMA redujo el RMSE (839.547 frente a 856.532) y mejoró el R² ajustado (0,40 frente a 0,37). En la agregación mensual por SKU, la mejora fue más marcada: SARIMA alcanzó un WAPE de 9,61 % frente al 11,03 % de ARIMA, y elevó el R² ajustado de 0,09 a 0,37, reflejando una mejor captura de los patrones estacionales presentes en la demanda. ARIMA, por su parte, mostró un rendimiento moderado como benchmark inicial, con un desempeño inferior al de los modelos estacionales y de aprendizaje automático, especialmente en el nivel agregado. Prophet, en cambio, fue el modelo con menor precisión relativa en ambos niveles, con WAPE superiores al 100 % en SKU-mes y del 32,40 % en la

serie mensual, y coeficientes de determinación ajustados negativos, lo cual indica un ajuste deficiente ante las características específicas del caso de estudio. Dado que XGBoost ofrece el mejor equilibrio entre precisión granular y agregada, las siguientes secciones profundizan en su análisis de errores, interpretabilidad y desempeño a nivel de producto. Además, en el anexo se incluyen el detalle de las predicciones por SKU, así como los gráficos en panel que ilustran visualmente las predicciones y su evolución a lo largo del tiempo.

5.4 Análisis de predicciones y errores a nivel SKU – Modelo XGBoost

Dado que el modelo XGBoost ha mostrado ser el que mejor desempeño tiene en términos de precisión, se opta por desglosar sus resultados a nivel de SKU en la siguiente tabla. Este desglose proporciona una visión detallada de cómo el modelo ajusta sus predicciones para cada producto, permitiendo así una evaluación más precisa de su capacidad para prever la demanda futura y optimizar la planificación de la producción.

Tabla 7. Resultados Pronóstico XGBoost - Nivel SKU

SKU	Rubro	Unidades reales	Unidades predichas	RMSE	WAPE (%)	R ² ajustado
BL08029	Bebidas	1706800	1721865	137691.0	0.85	0.90
BL81858	Alimentos	1425007	1451865	7062.8	1.88	0.95
BL81948	Alimentos	8941600	9147899	49607.7	2.31	0.94
BL83310	Alimentos	399168	388766	10402.5	2.61	NA
AM09603	Bebidas	3494976	3601406	91402.0	3.05	0.82
BL84929	Alimentos	7606271	7297183	70775.8	4.06	0.96
BL81960	Alimentos	4833280	5130253	115591.0	6.14	0.63
BL05011	Bebidas	1123979	1054439	14316.7	6.19	0.90
BL82303	Alimentos	630000	688263	30556.4	9.25	0.96
BL09112	Bebidas	248436	223362	25073.8	10.09	NA
BL05012	Bebidas	1984319	2184756	17398.4	10.10	0.96
BL05033	Bebidas	206980	178341	12993.5	13.84	0.83
BL05045	Bebidas	70560	60388	6413.3	14.42	-0.62
BL09309	Bebidas	2438868	2034909	137611.0	16.58	0.85
BL81726	Alimentos	3923720	4593595	393964.0	17.06	0.71
AM09107	Bebidas	4096456	5032418	390074.0	22.85	0.83
BL82302	Alimentos	120960	163759	22553.4	35.38	-7.90
BL05034	Bebidas	11520	18882	3883.8	63.90	NA
BL83508	Alimentos	278784	494993	112909.0	77.55	-6.22
AM09109	Bebidas	0	91416	91415.8	NA	NA

Esta tabla presenta el desempeño del modelo XGBoost a nivel de predicción por SKU. Se observa que, para los productos de mayor volumen de ventas, el modelo logró niveles de error porcentual (WAPE) considerablemente bajos, en algunos casos inferiores al 1 %, y coeficientes de determinación ajustados (R^2 ajustado) superiores al 0,90, indicando una elevada capacidad de ajuste a los patrones de demanda individuales. La disponibilidad de predicciones desagregadas por SKU representa una ventaja estratégica sustancial para la empresa. Permite anticipar de manera precisa la demanda futura de cada producto, habilitando una planificación de la producción más eficiente, reduciendo tiempos de respuesta y mejorando la asignación de recursos. Esto impacta directamente en el objetivo de maximizar el nivel de servicio hacia los clientes, garantizando disponibilidad de stock cuando sea necesario, y minimizando a su vez la acumulación innecesaria de inventarios. La optimización resulta particularmente crítica en SKUs que son exclusivos para determinados clientes, ya que su rotación depende directamente de la demanda de un único comprador. En estos casos, un sobrestock no solo implica costos financieros asociados al almacenamiento y a la inmovilización de capital, sino también riesgos comerciales en caso de cambios en el comportamiento de compra del cliente. Como caso ilustrativo, se analizó en detalle el desempeño del modelo para el SKU BL08029, asociado exclusivamente al Cliente C. Durante el último trimestre de 2024, el forecast oficial proporcionado por el cliente sobreestimó significativamente la demanda futura, registrando un error acumulado superior a 1,5 millones de unidades. En contraste, el modelo XGBoost desarrollado en este trabajo predijo con alta precisión, alcanzando un RMSE inferior a 140.000 unidades en el mismo período. Además, al analizar mes a mes, se observa que la predicción propia mantuvo consistentemente menores errores absolutos respecto al forecast externo.

Cabe destacar que, al tratarse de un forecast no vinculante, los errores de estimación impactan directamente sobre la empresa proveedora, quien asume los costos asociados a la sobreproducción, la acumulación de inventarios y las eventuales pérdidas comerciales. Como afirman Mentzer y Moon (2004), “las predicciones de ventas no vinculantes generan una transferencia de riesgo hacia el proveedor, quien debe optimizar su planificación para mitigar los efectos de errores de estimación”. En este sentido, disponer de herramientas internas de predicción más precisas resulta estratégico no solo para optimizar los procesos productivos

y logísticos, sino también para reducir riesgos financieros y fortalecer los niveles de servicio ofrecidos a los clientes.

Con el objetivo de evaluar la robustez del modelo XGBoost en distintos contextos temporales, se realizó además una prueba *out-of-sample* complementaria. Dado su destacado desempeño en el período octubre–diciembre 2024, se decidió testear nuevamente el modelo sobre un tramo anterior no considerado en la evaluación principal. Para ello, se entrenó el modelo con datos históricos hasta junio de 2024, utilizando como nuevo período de validación los meses de julio a septiembre de ese mismo año. Esta estrategia buscó verificar si el modelo mantenía un desempeño sólido frente a cambios temporales y con una menor cantidad de datos disponibles para el entrenamiento.

Tabla 8. Comparación del desempeño del modelo XGBoost en dos ventanas temporales (validación out-of-sample)

Tipo de evaluación	Métrica	Jul-Sep 2024	Oct-Dic 2024
SKU-mes (individual)	RMSE	84672.45	76411.93
SKU-mes (individual)	MAPE (%)	158.27	145.91
SKU-mes (individual)	WAPE (%)	20.37	18.02
SKU-mes (individual)	R ² ajustado	0.91	0.94
Mensual agregado	RMSE	561984.32	491812.28
Mensual agregado	MAPE (%)	3.14	2.77
Mensual agregado	WAPE (%)	3.28	2.94
Mensual agregado	R ² ajustado	0.89	0.93

Los resultados obtenidos en esta validación alternativa fueron altamente consistentes con los valores registrados originalmente. El error absoluto ponderado (WAPE) y el coeficiente de determinación ajustado (R² ajustado) mostraron variaciones menores, confirmando la estabilidad del modelo en su aplicación práctica. La leve disminución en las métricas de precisión se atribuye principalmente al menor volumen de datos de entrenamiento, lo cual es esperable en contextos de aprendizaje supervisado. En conjunto, este análisis refuerza la confiabilidad del modelo XGBoost para tareas de pronóstico operacional en distintos escenarios temporales.

5.5 Reflexiones finales sobre los resultados

En términos operativos, la posibilidad de anticipar la demanda total mensual con un nivel de precisión elevado representa un aporte significativo para la gestión de la empresa. Contar con predicciones confiables permite optimizar la planificación de la producción, anticipando con mayor certeza las necesidades de materias primas, la programación de cargas y mantenimientos en los hornos industriales, y la asignación eficiente de turnos de personal.

Asimismo, desde una perspectiva financiera, proyectar la demanda futura facilita la estimación más precisa de los ingresos por ventas, permitiendo mejorar la gestión del flujo de caja, la planificación de compras estratégicas y la toma de decisiones de inversión a corto y mediano plazo.

Por otro lado, a nivel desagregado, disponer de estimaciones por SKU permite a la empresa mejorar sustancialmente su flexibilidad operativa. Una predicción precisa de la demanda de cada producto facilita una planificación de la producción más ajustada a las necesidades reales, reduciendo los tiempos de anticipación requeridos para lanzar órdenes de fabricación. Esto no solo permite satisfacer de manera más eficiente la demanda de los clientes —mejorando el nivel de servicio ofrecido—, sino que también contribuye a evitar la acumulación innecesaria de inventarios, optimizando el uso de espacio, capital de trabajo y costos asociados al almacenamiento.

De este modo, los resultados obtenidos no solo tienen relevancia académica en términos de validación de modelos predictivos, sino que además presentan una aplicabilidad práctica concreta, alineándose con los objetivos de eficiencia operativa, servicio al cliente y rentabilidad estratégica de la empresa. Como sostienen Makridakis, Wheelwright y Hyndman (1998), “la capacidad de prever con precisión la demanda futura es uno de los elementos más críticos para optimizar las decisiones de producción, inventario, y asignación de recursos, impactando directamente en la eficiencia y rentabilidad de las organizaciones.”

6. Conclusiones

6.1 Principales hallazgos

A lo largo del trabajo se evaluaron distintos modelos de predicción de demanda con el objetivo de estimar con precisión las ventas futuras de productos de Vidriotec. Las predicciones se realizaron en dos niveles de agregación: por un lado, a nivel mensual agregado, fundamental para optimizar la eficiencia operativa global mediante una adecuada planificación de la producción, la asignación de capacidad, la gestión de hornos y turnos de trabajo. Por otro lado, se trabajó a nivel SKU–mes, que constituye el nivel de mayor relevancia práctica para la gestión del inventario y la maximización del nivel de servicio hacia los clientes, al permitir anticipar con mayor precisión las necesidades de abastecimiento por producto.

Ambos niveles ofrecen valor complementario para la empresa. Mientras que la predicción agregada facilita decisiones estratégicas de volumen y planificación general, la predicción desagregada permite una asignación más precisa de recursos, evitando sobrestock o faltantes, y garantizando un servicio eficiente a los clientes.

Entre los modelos evaluados, XGBoost se destacó ampliamente en ambos niveles. A nivel SKU–mes, alcanzó un RMSE de 76.412, un WAPE de 18,02% y un R^2 de 0,96, lo cual refleja una gran capacidad para capturar las variaciones individuales de la demanda. A nivel mensual agregado, su desempeño fue aún más notable: RMSE de 491.812, WAPE de 2,94% y R^2 de 0,95, posicionándolo como el modelo más robusto para la planificación general.

La métrica RMSE fue priorizada en esta tesis por su capacidad para penalizar con mayor severidad los errores grandes, lo que la convierte en una herramienta más adecuada que el desvío estándar para la estimación de stock de seguridad. Como señalan Hanke y Wichern (2010), “el error cuadrático medio debe ser preferido sobre otras métricas cuando el objetivo es minimizar el impacto de errores de gran magnitud en decisiones operativas, como la planificación de inventarios”.

En contraste, el modelo Random Forest mostró un buen rendimiento en el nivel mensual agregado (RMSE de 889.192, WAPE de 4,92%, R^2 de 0,84), pero presentó limitaciones

importantes a nivel SKU–mes, donde alcanzó un WAPE del 78,74% y un MAPE extremadamente alto, lo que restringe su aplicabilidad para planificación desagregada.

Los modelos ARIMA y SARIMA, aunque válidos como benchmarks metodológicos, quedaron rezagados frente a los modelos de machine learning, especialmente en el nivel SKU–mes. Sin embargo, SARIMA mostró mejoras consistentes respecto a ARIMA en el agregado mensual, particularmente al capturar componentes estacionales.

En cuanto a Prophet y las redes neuronales LSTM, sus desempeños fueron insatisfactorios en ambos niveles. Prophet presentó WAPE superiores al 100% en el nivel SKU–mes y errores significativos incluso en el nivel agregado. Las redes LSTM, a pesar de su potencial teórico, no lograron generalizar correctamente debido a las limitaciones en el volumen y la granularidad de los datos disponibles.

En términos de ingeniería de variables, la incorporación de rezagos, medias móviles y la codificación one-hot de variables categóricas como SKU, cliente y provincia resultó fundamental para mejorar el desempeño predictivo de los modelos de árboles, permitiéndoles capturar relaciones no lineales y efectos históricos relevantes.

En suma, los resultados obtenidos validan empíricamente la utilidad de aplicar enfoques de machine learning avanzados para la predicción de demanda en contextos industriales complejos. Además, resaltan la importancia de adaptar el nivel de agregación del pronóstico a las necesidades concretas del negocio, generando valor tanto en la planificación operativa general como en la toma de decisiones específicas por producto. No obstante, y en línea con lo señalado por Hanke y Wichern (2010), “ningún modelo de pronóstico, por sofisticado que sea, puede reemplazar el conocimiento experto del negocio; la clave está en integrarlo de manera complementaria al modelo para mejorar la calidad y aplicabilidad de los resultados”. En esta investigación, dicha integración se manifestó tanto en la selección de los SKUs más relevantes —definidos junto a especialistas internos— como en el diseño del horizonte de predicción y en la interpretación de los resultados, permitiendo desarrollar una herramienta que no solo tiene validez técnica, sino también utilidad práctica concreta para la gestión de Vidriotec.

6.2 Limitaciones del estudio

Si bien los resultados obtenidos fueron sólidos y muestran una clara ventaja de los modelos de machine learning frente a los enfoques tradicionales, el estudio presenta ciertas limitaciones que es importante considerar al interpretar sus conclusiones y proyectar su aplicabilidad.

Una de las principales restricciones está vinculada al volumen y la granularidad de los datos disponibles. La serie histórica, con frecuencia mensual y limitada a cuatro años, acota la capacidad de los modelos para identificar patrones complejos o capturar ciclos de largo plazo. Esta limitación no solo es técnica, sino también estructural del negocio, ya que una proporción importante de los productos analizados están destinados a clientes exclusivos. Esto genera una dinámica de demanda muy particular: algunos SKUs presentan ventas en determinados meses y ausencia total en otros, lo cual dificulta la generación de series temporales consistentes y reduce la densidad de datos útiles para el entrenamiento de modelos más exigentes.

Por otro lado, la variable objetivo utilizada —la cantidad de unidades vendidas— fue considerada un buen proxy de la demanda efectiva, dadas las condiciones de alto nivel de servicio observadas en la empresa. No obstante, en situaciones de quiebre de stock o productos fuera de ciclo productivo, esta variable podría subestimar la demanda real. La disponibilidad futura de información como pedidos no satisfechos o niveles de stock permitiría ajustar el modelo para captar mejor la demanda latente y mejorar su capacidad explicativa.

Finalmente, cabe mencionar que la evaluación del modelo se realizó sobre datos históricos y fuera de un entorno productivo real. Aunque la división temporal fue diseñada cuidadosamente para simular un escenario operativo, aún no se ha medido el impacto de estos modelos en la toma de decisiones concretas, tales como la planificación de producción, la gestión de inventarios o el nivel de servicio al cliente. Esta validación práctica será clave para consolidar su utilidad en el negocio.

Estas limitaciones no desmerecen los hallazgos del estudio, pero sí marcan oportunidades claras de mejora metodológica y aplicación práctica, las cuales se abordan en la sección siguiente.

6.3 Líneas futuras de investigación

A partir de los resultados obtenidos y de las limitaciones previamente discutidas, surgen diversas oportunidades para profundizar esta línea de investigación y fortalecer el desarrollo de modelos predictivos en contextos industriales como el de Vidriotec.

En primer lugar, se destaca la posibilidad de retomar la aplicación de redes neuronales recurrentes, como las arquitecturas LSTM, cuya capacidad para capturar dependencias secuenciales de largo plazo podría aprovecharse mejor en escenarios con series más extensas, mayor granularidad temporal o datasets enriquecidos con variables externas. Si bien en este estudio su desempeño no superó al de los modelos tradicionales, su potencial sigue vigente y podría explorarse con enfoques alternativos o bajo configuraciones más ajustadas a la naturaleza de los datos.

Asimismo, una línea de trabajo de alto valor sería la incorporación de variables macroeconómicas que puedan estar correlacionadas con la demanda del sector, tales como el Índice de Producción Industrial, el consumo masivo o el nivel de actividad manufacturera. Dado que los productos de Vidriotec funcionan como insumos en etapas tempranas de la cadena productiva de sus clientes, resulta razonable esperar que sus ventas reaccionen con cierto rezago ante cambios en estos indicadores. Modelar explícitamente estos desfases podría aportar información predictiva valiosa, especialmente para horizontes de mediano plazo.

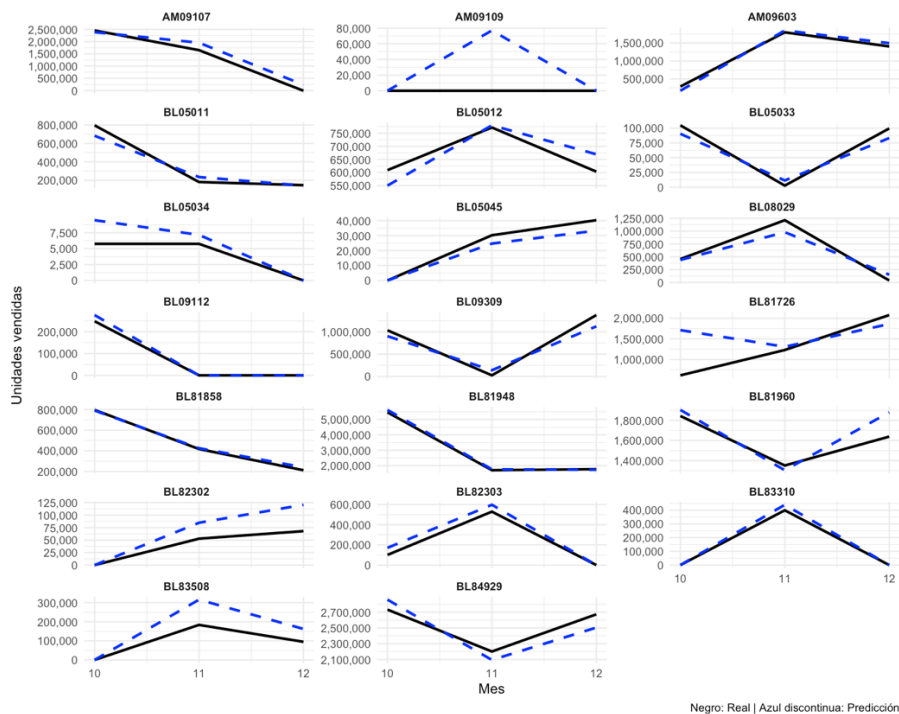
Desde una perspectiva endógena, también resultaría enriquecedor ampliar la ventana temporal de análisis incluyendo más años de información histórica, lo que permitiría estabilizar los patrones de demanda de productos con comportamiento irregular. Del mismo modo, la inclusión de variables internas como los niveles de inventario diario por SKU contribuiría a distinguir si las ventas observadas reflejan la demanda real o estuvieron condicionadas por restricciones de stock. Esta diferenciación no solo mejoraría la precisión

de los modelos, sino que también permitiría evaluar con mayor detalle el nivel de servicio alcanzado y detectar oportunidades de optimización logística y operativa.

Finalmente, futuras investigaciones podrían avanzar hacia la implementación de los modelos en contextos reales, ya sea mediante la integración de herramientas de visualización o dashboards operativos, como a través del diseño de sistemas de pronóstico colaborativos que combinen inputs estadísticos con ajustes realizados por expertos del negocio. Este enfoque permitiría capitalizar la precisión de los modelos predictivos y al mismo tiempo incorporar el conocimiento tácito de los responsables de la planificación, alineándose con las recomendaciones de la literatura para lograr pronósticos más robustos y accionables.

Anexo
Tabla 7. Resultados Pronóstico XGBoost - Nivel SKU

SKU	Rubro	Unidades reales	Unidades predichas	RMSE	WAPE (%)	R ² ajustado
BL08029	Bebidas	1706800	1721865	137691.0	0.85	0.90
BL81858	Alimentos	1425007	1451865	7062.8	1.88	0.95
BL81948	Alimentos	8941600	9147899	49607.7	2.31	0.94
BL83310	Alimentos	399168	388766	10402.5	2.61	NA
AM09603	Bebidas	3494976	3601406	91402.0	3.05	0.82
BL84929	Alimentos	7606271	7297183	70775.8	4.06	0.96
BL81960	Alimentos	4833280	5130253	115591.0	6.14	0.63
BL05011	Bebidas	1123979	1054439	14316.7	6.19	0.90
BL82303	Alimentos	630000	688263	30556.4	9.25	0.96
BL09112	Bebidas	248436	223362	25073.8	10.09	NA
BL05012	Bebidas	1984319	2184756	17398.4	10.10	0.96
BL05033	Bebidas	206980	178341	12993.5	13.84	0.83
BL05045	Bebidas	70560	60388	6413.3	14.42	-0.62
BL09309	Bebidas	2438868	2034909	137611.0	16.58	0.85
BL81726	Alimentos	3923720	4593595	393964.0	17.06	0.71
AM09107	Bebidas	4096456	5032418	390074.0	22.85	0.83
BL82302	Alimentos	120960	163759	22553.4	35.38	-7.90
BL05034	Bebidas	11520	18882	3883.8	63.90	NA
BL83508	Alimentos	278784	494993	112909.0	77.55	-6.22
AM09109	Bebidas	0	91416	91415.8	NA	NA

Figura 20. Demanda real y predicción por SKU - XGBoost (Oct - Dic 2024)


Negro: Real | Azul discontinua: Predicción

Tabla 9. Resultados Pronóstico Random Forest - Nivel SKU

SKU	Rubro	Unidades reales	Unidades predichas	RMSE	Error absoluto	WAPE (%)	R ²
BL08029	Bebidas	1.706.800	1.679.944	460383.52	26.856	1.57	0.11
BL84929	Alimentos	7.606.271	8.058.088	364757.24	451.817	5.94	0.44
BL81726	Alimentos	3.923.720	4.192.198	748890.33	268.478	6.84	0.08
BL81948	Alimentos	8.941.600	9.670.107	202151.74	728.507	8.15	0.27
BL83310	Alimentos	399.168	474.163	74995.29	74.995	18.79	NA
BL81960	Alimentos	4.833.280	3.621.260	451686.84	1.212.020	25.08	-4.02
AM09107	Bebidas	4.096.456	3.055.686	1064292.57	1.040.770	25.41	-0.08
BL09309	Bebidas	2.438.868	1.749.312	399623.92	689.556	28.27	0.00
BL05033	Bebidas	206.980	277.611	35017.45	70.631	34.12	-0.02
AM09603	Bebidas	3.494.976	2.271.363	321948.02	1.223.613	35.01	-0.84
BL05011	Bebidas	1.123.979	1.695.256	37943.11	571.277	50.83	0.47
BL05012	Bebidas	1.984.319	3.030.627	102496.34	1.046.308	52.73	0.43
BL82303	Alimentos	630.000	1.057.594	303002.76	427.594	67.87	-1.00
BL81858	Alimentos	1.425.007	2.401.850	37834.66	976.843	68.55	0.12
BL09112	Bebidas	248.436	453.332	204895.62	204.896	82.47	NA
BL05045	Bebidas	70.560	157.960	48597.63	87.400	123.87	-91.98
BL83508	Alimentos	278.784	904.392	312940.94	625.608	224.41	-48.78
BL82302	Alimentos	120.960	454.245	167055.38	333.285	275.53	-487.29
BL05034	Bebidas	11.520	96.707	42603.68	85.187	739.47	NA
AM09109	Bebidas	0	378.046	378046.29	378.046	NA	NA

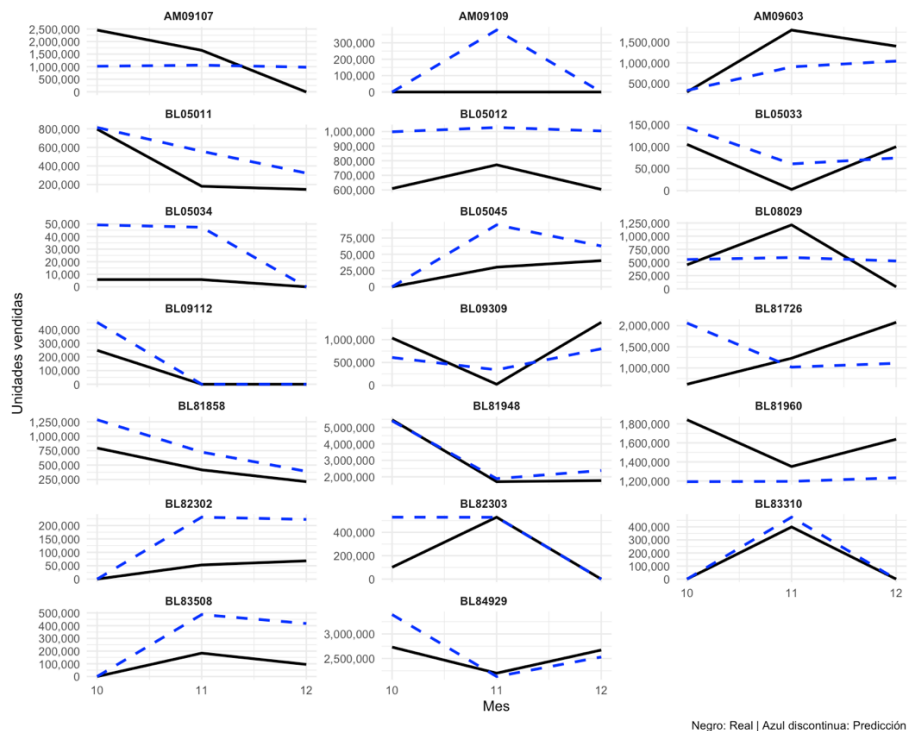
Figura 21. Demanda real y predicción por SKU - Random Forest (Oct - Dic 2024)


Tabla 10. Resultados Pronóstico ARIMA - Nivel SKU

SKU	Unidades reales	Unidades predichas	RMSE	Error absoluto	WAPE (%)	R ²
AM09107	4.096.456	3.983.456	888242.1	113.000	2.76	0.25
BL81858	1.425.007	1.631.261	251253.5	206.254	14.47	-0.08
BL84929	7.606.271	8.901.939	492783.6	1.295.668	17.03	-3.31
BL05012	1.984.319	1.582.972	154910.5	401.347	20.23	-2.93
BL81960	4.833.280	3.815.973	399270.5	1.017.307	21.05	-2.92
BL09309	2.438.868	1.885.941	603928.6	552.927	22.67	-0.10
BL81948	8.941.600	6.708.755	1910726.7	2.232.845	24.97	-0.18
AM09603	3.494.976	2.396.146	737512.2	1.098.830	31.44	-0.33
BL81726	3.923.720	5.440.208	874608.3	1.516.488	38.65	-1.13
BL05011	1.123.979	683.565	332732.9	440.414	39.18	-0.24
BL08029	1.706.800	4.277.200	1250981.3	2.570.400	150.60	-5.60

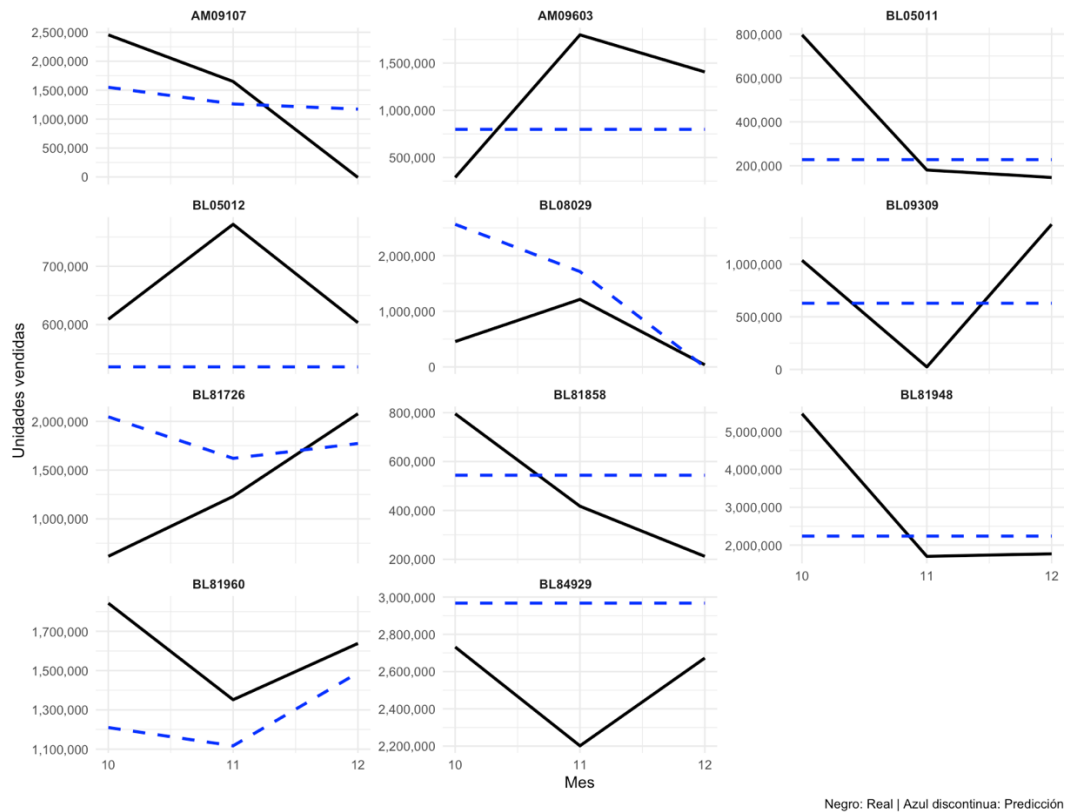
Figura 22. Demanda real y predicción por SKU - ARIMA (Oct - Dic 2024)


Tabla 11. Resultados Pronóstico SARIMA - Nivel SKU

SKU	Unidades reales	Unidades predichas	RMSE	Error absoluto	WAPE (%)	R ²
BL84929	7.606.271	7.769.649	761841.7	163.378	2.15	-9.31
BL81858	1.425.007	1.631.261	251253.5	206.254	14.47	-0.08
BL05012	1.984.319	1.582.972	154910.5	401.347	20.23	-2.93
BL81960	4.833.280	3.815.973	399270.5	1.017.307	21.05	-2.92
BL09309	2.438.868	1.885.941	603928.6	552.927	22.67	-0.10
BL81948	8.941.600	6.839.754	1757845.7	2.101.846	23.51	0.00
AM09603	3.494.976	2.396.146	737512.2	1.098.830	31.44	-0.33
BL81726	3.923.720	5.440.208	874608.3	1.516.488	38.65	-1.13
BL05011	1.123.979	683.565	332732.9	440.414	39.18	-0.24
AM09107	4.096.456	6.011.501	867756.6	1.915.045	46.75	0.29
BL08029	1.706.800	4.277.200	1250981.3	2.570.400	150.60	-5.60

Figura 23. Demanda real y predicción por SKU - SARIMA (Oct - Dic 2024)

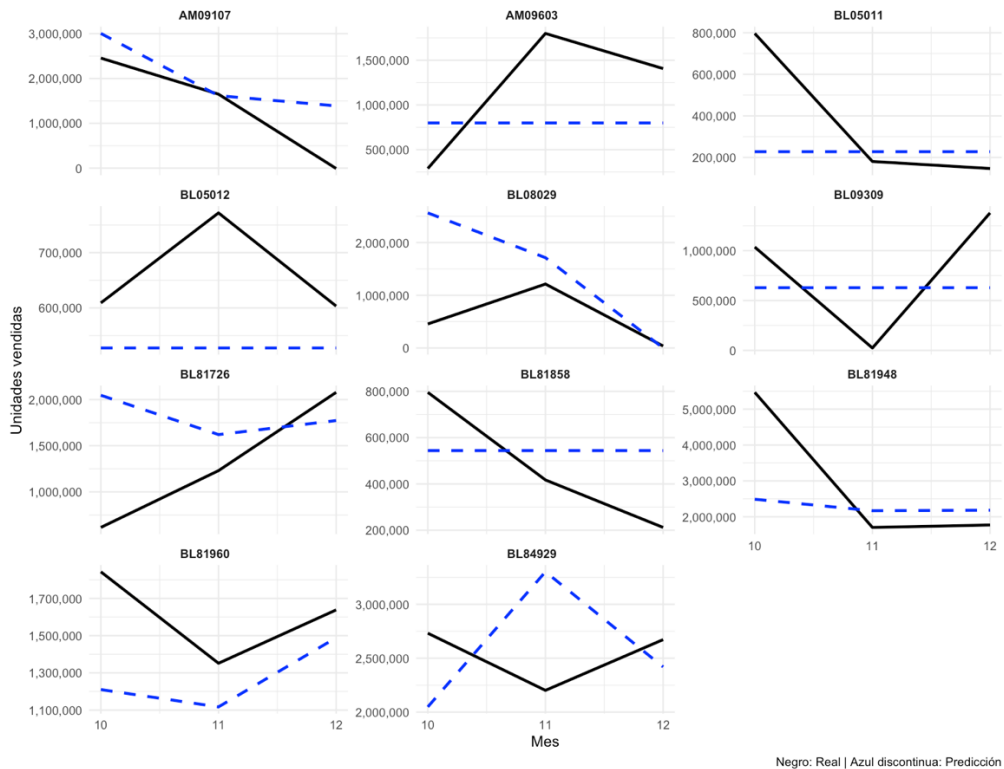
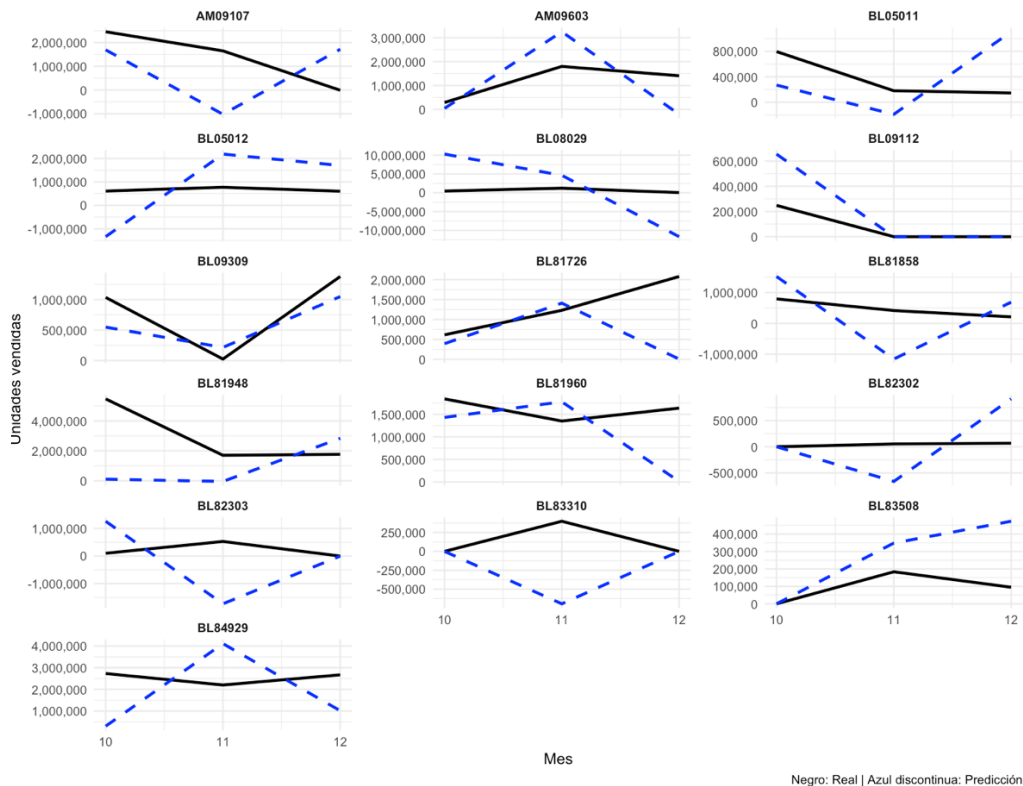


Tabla 12. Resultados Pronóstico PROPHET - Nivel SKU

SKU	Unidades reales	Unidades predichas	RMSE	Error absoluto	WAPE (%)	R ²
BL05011	1.123.979	1.189.023	667865.1	65.044	5.79	-4.00
AM09603	3.494.976	3.085.009	1260475.7	409.967	11.73	-2.88
BL09309	2.438.868	1.812.866	357078.3	626.002	25.67	0.61
BL81858	1.425.007	1.052.878	1038376.3	372.129	26.11	-17.46
BL05012	1.984.319	2.547.930	1526260.1	563.611	28.40	-380.91
BL84929	7.606.271	5.426.880	2025705.3	2.179.391	28.65	-71.88
BL81960	4.833.280	3.222.150	999761.3	1.611.130	33.33	-23.60
AM09107	4.096.456	2.386.903	1889393.2	1.709.553	41.73	-2.39
BL81726	3.923.720	1.809.301	1207400.2	2.114.419	53.89	-3.05
BL81948	8.941.600	2.903.265	3312074.7	6.038.335	67.53	-2.54
BL08029	1.706.800	3.200.097	9041923.4	1.493.297	87.49	-343.85
BL82302	120.960	252.444	785123.1	131.484	108.70	-10784.29
BL09112	248.436	655.304	406868.4	406.868	163.77	-Inf
BL82303	630.000	-469.757	1796935.5	1.099.757	174.56	-69.38
BL83508	278.784	821.965	291855.1	543.181	194.84	-42.30
BL83310	399.168	-695.162	1094329.9	1.094.330	274.15	-Inf

Figura 24. Demanda real y predicción por SKU - PROPHET (Oct - Dic 2024)


Bibliografía

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden-Day.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd ed.). Prentice Hall.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), 1287–1294. <https://doi.org/10.2307/1911963>

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007. <https://doi.org/10.2307/1912773>

Facebook Prophet. (2024). Prophet: Forecasting at scale. Meta Open Source. <https://facebook.github.io/prophet/>

Fildes, R., Ma, S., & Kolassa, S. (2019). Retail forecasting: Research and practice. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2019.06.004>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hanke, J. E., & Wichern, D. (2010). *Pronósticos en los negocios* (9.^a ed.). Pearson Educación.

Hastie, T. J., & Tibshirani, R. J. (1987). Generalized additive models. *Statistical Science*, 2(3), 297–318. <https://doi.org/10.1214/ss/1177013604>

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. <https://otexts.com/fpp2/>

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. <https://www.statlearning.com/>

Kwarteng, S. B., & Andreevich, P. A. (2024). Comparative analysis of ARIMA, SARIMA and Prophet model in forecasting. *Research & Development*, 5(4), 110–120. <https://doi.org/10.11648/j.rd.20240504.13>

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). Wiley.

Masini, F. J., & Vázquez Ledesma, F. (2014). *Modelos cuantitativos de pronósticos* (7^a ed.). AVG Consulting.

McCarthy, T. M., Davis, D. F., Golicic, S. L., & Mentzer, J. T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. *Journal of Forecasting*, 25(4), 303–324. <https://doi.org/10.1002/for.989>

Mentzer, J. T., & Moon, M. A. (2004). *Sales forecasting management: A demand management approach*. SAGE Publications.

Moto DEI. (2020). Facebook Prophet: (Almost) everything you should know to get started. *Medium*. <https://medium.com/swlh/facebook-prophet-426421f7e331>

Taylor, S., & Letham, B. (2017). Forecasting at scale. *PeerJ Preprints*. <https://doi.org/10.7287/peerj.preprints.3190v2>