

Taller Final Proyección Productos 1 y 2 Empresa Comestibles

Juan Sebastian Orozco Hernandez - 1017173516

Introducción

En un entorno competitivo como lo es la industria de comestibles, se hace indispensable para las compañías del sector entender con precisión el comportamiento histórico de las ventas de sus productos.

El presente informe tiene como objetivo desarrollar y seleccionar los modelos de pronóstico más precisos y robustos para predecir las ventas mensuales de los dos productos estrella de una empresa del sector. Utilizando un histórico de datos que abarca desde junio de 2014 hasta noviembre de 2024, se han aplicado técnicas avanzadas de análisis de series de tiempo y aprendizaje automático para capturar las complejidades inherentes a cada producto, tales como tendencias, estacionalidad y posibles cambios estructurales en el comportamiento del consumidor.

Se han evaluado varios modelos cuyas técnicas van desde métodos de suavizado exponencial (Holt-Winters) hasta algoritmos estadísticos avanzados (SARIMA) y modelos de regresión aditiva modernos (Prophet). La selección del modelo ganador para cada producto se basó en sus resultados para minimizar el error de pronóstico que permita que las proyecciones entregadas sean confiables y accionables para la toma de decisiones del próximo mes.

A continuación, se detalla la metodología empleada, el análisis individual del comportamiento de cada producto y las proyecciones de ventas resultantes.

Metodología

Para garantizar la robustez de las proyecciones, se diseñó un flujo de trabajo analítico basado en tres pilares fundamentales: descomposición de series temporales, optimización bayesiana de hiperparámetros y validación cruzada temporal.

En primer lugar, se descartaron enfoques simplistas (como medias móviles) en favor de modelos capaces de capturar patrones complejos: **Suavizado Exponencial (Holt-Winters)** para capturar nivel y tendencia, **SARIMA** para modelar la estructura de autocorrelación y estacionalidad, y **Prophet** para manejar cambios de régimen y estacionalidades múltiples.

La optimización de cada modelo no se realizó mediante búsqueda exhaustiva (Grid Search), que suele ser computacionalmente costosa y limitada, sino mediante **Optimización Bayesiana (usando el framework Optuna)**. Este enfoque permite explorar el espacio de hiperparámetros de manera probabilística, convergiendo más eficientemente hacia la configuración óptima que minimiza el error.

Finalmente, para evitar el sobreajuste (*overfitting*), la validación se ejecutó mediante una estrategia de **Ventana Expandida (Expanding Window)**. En lugar de una única división de prueba, el modelo fue re-entrenado iterativamente en múltiples puntos del tiempo, simulando condiciones reales de pronóstico mes a mes.

Análisis Exploratorio de Datos (EDA) y Descomposición

Al analizar la descomposición del **Producto 1**, se evidencia un comportamiento no estacionario con una tendencia estructural bajista marcada a partir de 2021 (ver Figura 1). El componente estacional muestra un patrón aditivo claro y repetitivo cada 12 meses. La magnitud de la caída

sugiere un cambio de régimen en el mercado, lo cual presenta un desafío para modelos que dependen fuertemente de la media histórica, como los promedios móviles simples.

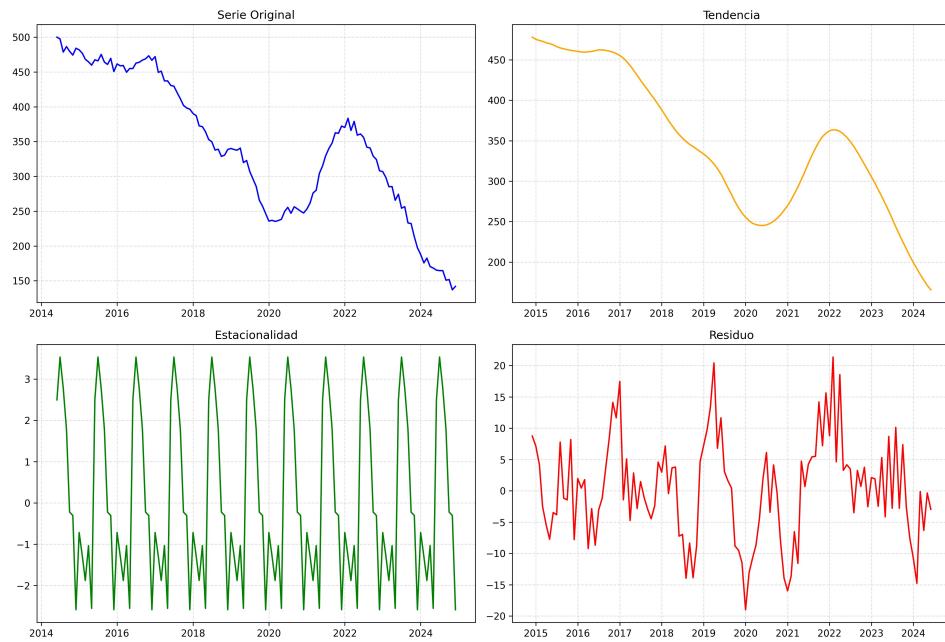


Figura 1. Descomposición de la Serie del Producto1.

El **Producto 2**, en contraste, muestra una tendencia positiva, con un crecimiento sostenido hasta 2023 seguido de una desaceleración reciente. La varianza del componente estacional se mantiene relativamente estable (homoscedasticidad), lo que justifica el uso de modelos aditivos. La detección de estos puntos de inflexión en la tendencia es crítica para evitar sobreestimar las ventas futuras.

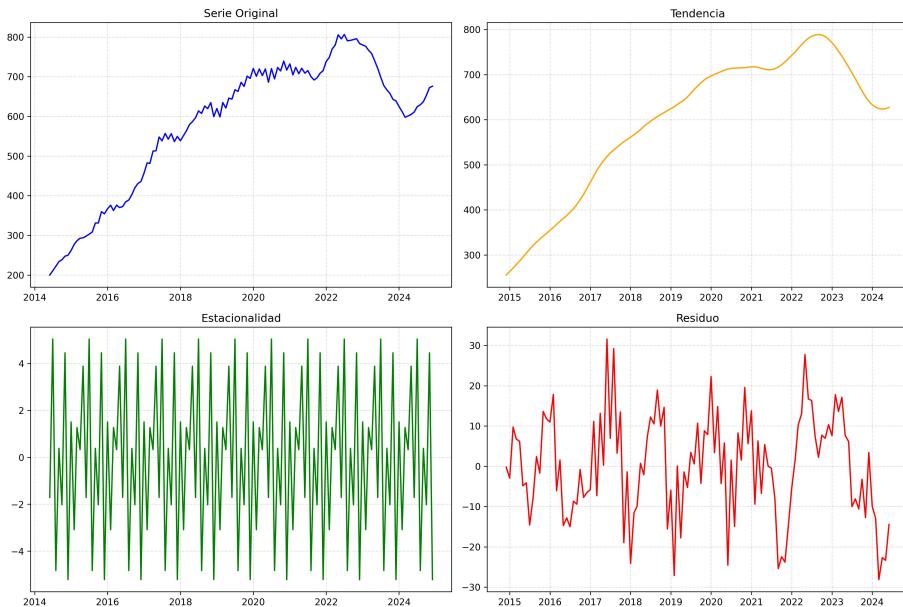


Figura 2. Descomposición de la Serie del Producto2.

Evaluación de Modelos y Selección

Tras ejecutar 20 iteraciones de optimización para cada tipo de modelo, se compararon los resultados bajo la métrica de la Raíz del Error Cuadrático Medio (RMSE), penalizando así los grandes desvíos que podrían causar rupturas de stock críticas.

Producto	Modelo	Error	RMSE
Producto 1	Holt-Winters	Trend: Multiplicativo, Seasonal: Aditivo	3.72
	SARIMA	Order (0,1,1) x Seasonal (2,0,1,12)	3.23
	Prophet	Seasonality: Aditivo, Prior Scale: 0.44	25.51
Producto 2	Holt-Winters	Trend: Multiplicativo, Seasonal: Aditivo	8.98
	SARIMA	Order (0,1,2) x Seasonal (2,1,0,12)	4.16
	Prophet	Seasonality: Multiplicativo, Prior Scale: 0.09	21.81

Tabla 1. Resultados de la Competencia de Modelos y Parámetros Óptimos

Como se evidencia en la **Tabla 1**, el algoritmo **SARIMA** superó en gran medida a las otras aproximaciones.

- En el **Producto 1**, SARIMA logró reducir el error (RMSE) en un **13%** frente a su competidor más cercano (Holt-Winters), demostrando una mayor capacidad para interpretar la tendencia no lineal reciente.
- En el **Producto 2**, la diferencia fue aún más notable: SARIMA redujo el error a menos de la mitad en comparación con Holt-Winters (4.16 vs 8.98), validando la importancia de los términos de media móvil (MA) y la diferenciación estacional detectados por la optimización bayesiana.

Diagnóstico de Residuos (Validación de Supuestos)

La selección del modelo no se basó únicamente en la minimización del error. Se realizó un importante diagnóstico de los residuos del modelo ganador para asegurar la validez estadística de las predicciones (ver Figuras 3 y 4 - Tableros de Diagnóstico).

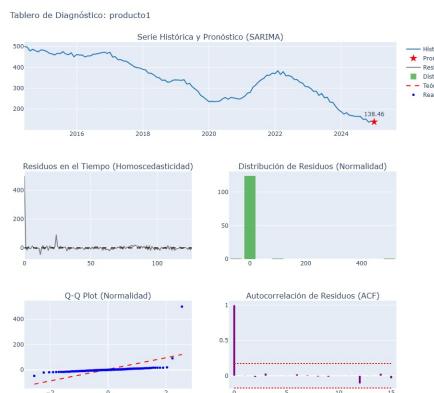


Figura 3 - Tablero de Diagnóstico Producto 1



Figura 4 - Tablero de Diagnóstico Producto 2

- Test de Ljung-Box (Autocorrelación):** Se evaluó la hipótesis nula de independencia en los residuos. Con un *p-valor* superior a 0.05, confirmamos que los residuos se comportan como **Ruido Blanco**, lo que implica que el modelo ha capturado toda la información sistemática disponible en la serie.
- Test de Shapiro-Wilk / Jarque-Bera (Normalidad):** Se analizó la distribución de los errores. Aunque en datos reales es común encontrar desviaciones en las colas (leptocurtosis), la distribución se aproxima lo suficiente a la normalidad para validar los intervalos de confianza generados.

Pronóstico Final y Recomendaciones de Negocio

Producto	Modelo	Pronóstico Mes Siguiente	RMSE_CV
Producto 1	SARIMA	138.46	3.23
Producto 2	SARIMA	681.67	4.16

Tabla 2. Modelo ganador con el valor pronosticado para el mes de Enero de 2025

Con base en el modelo re-entrenado con la totalidad de la historia disponible, se proyectan las siguientes cifras para el mes de Enero del año 2025:

- **Producto 1:** 138.46 unidades.
- **Producto 2:** 681.67 unidades.

Dada la tendencia bajista del Producto 1, se sugiere revisar otras variables o características para evitar pérdidas por un sobre-stock, ajustando los niveles de seguridad a la baja. Para el Producto 2, a pesar de la corrección reciente, la volatilidad estacional sugiere mantener un inventario de seguridad flexible para cubrir posibles picos de demanda no lineales detectados por el modelo.