



Imagen: © 2018 Advantus Media, Inc. and QuoteInspector.com

INTRO

METODOLOGÍA

MODELOS

CONCLUSIONES

ALEJANDRA VAQUERO

DICIEMBRE 2024

SERIES TEMPORALES

ÍNDICE

03	INTRODUCCIÓN
04	METODOLOGÍA
10	MODELOS
13	CONCLUSIONES

INTRODUCCIÓN

TodoVentas S.A, una empresa de comercio retail online, está evaluando una ampliación de capital y necesita estimar sus ventas futuras para respaldar su valoración. Para ello, se ha desarrollado un modelo de predicción basado en series temporales que permite estimar el nivel de ventas totales para el próximo mes.

Utilizando datos históricos de ventas diarias por producto en los principales mercados, se han aplicado diferentes técnicas de modelado con el objetivo de identificar la mejor estrategia de predicción. Este informe presenta el análisis exploratorio de los datos, la selección y evaluación de modelos predictivos, y las conclusiones obtenidas a partir de las métricas de desempeño.

Los resultados de este estudio permitirán a la empresa tomar decisiones informadas sobre su estrategia de crecimiento y expansión.

OBJETIVO

Predecir cuál será el nivel de ventas totales durante el próximo mes.

META

Seleccionar y justificar el modelo predictivo que mejor se ajuste al objetivo.

METODOLOGÍA

El *dataset* contiene 394.330 registros de transacciones de ventas, entre las fechas 1 de diciembre de 2022 hasta el 9 de diciembre del 2023.

Incluye ventas diarias por producto en los mercados más importantes durante los últimos años. Se ha fijado un horizonte de predicción de un mes teniendo en cuenta la última fecha de transacciones disponible en los datos.

Los países representados son 10: United Kingdom, France, Netherlands, Germany, EIRE, Spain, Portugal, Italy, Belgium, Sweden.

METODOLOGÍA

El *dataset* contiene cinco variables:

- **InvoiceNo:** Número de factura (convertir a *string*).
- **StockCode:** Código del producto (convertir a *string*).
- **Date:** Fecha de la transacción (texto, requiere conversión a formato *datetime* *aaaa-mm-dd*).
- **Country:** País de la transacción (convertir a *category*).
- **TotalSales:** Ventas totales por transacción (numérico, decimales *float*).

No existen datos nulos en el dataset.

Existe una distribución de los datos sesgada hacia la derecha, ya que la media es mayor que la mediana en la columna de ventas o **TotalSales**, hay posibles outliers.

METODOLOGÍA

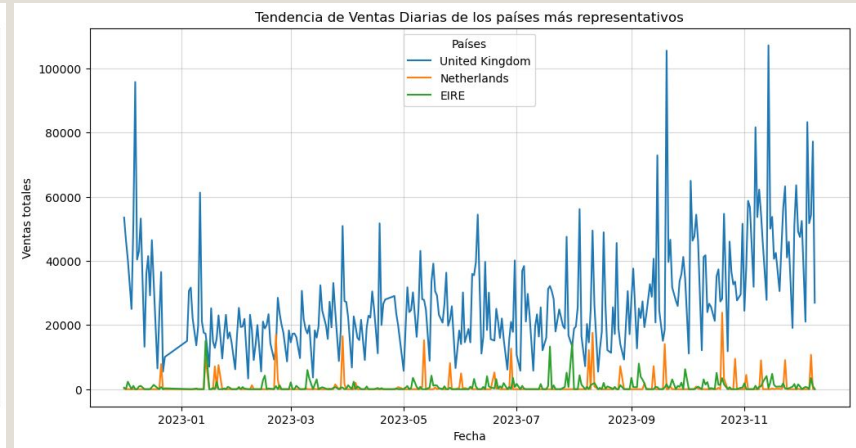
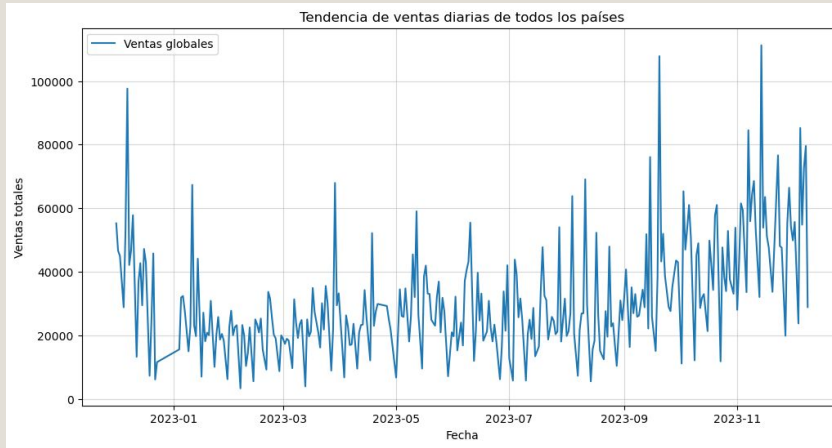
Se realizó el ejercicio con interpolación de datos lineales. Al completar la información diaria con una interpolación lineal se aumenta hasta casi en un 20% las ventas en algunos países, lo que altera significativamente los resultados de las ventas totales. Se decide no imputar los días y datos faltantes para mantener la fidelidad de los datos originales.

Se puede observar la evolución diaria de las ventas totales combinadas de los 10 países. Se observan algunos picos, denotando momentos de promociones o temporadas altas. Es posible que exista estacionalidad mensual o semanal pero lo evaluaremos en una descomposición de componentes de la serie temporal.

METODOLOGÍA

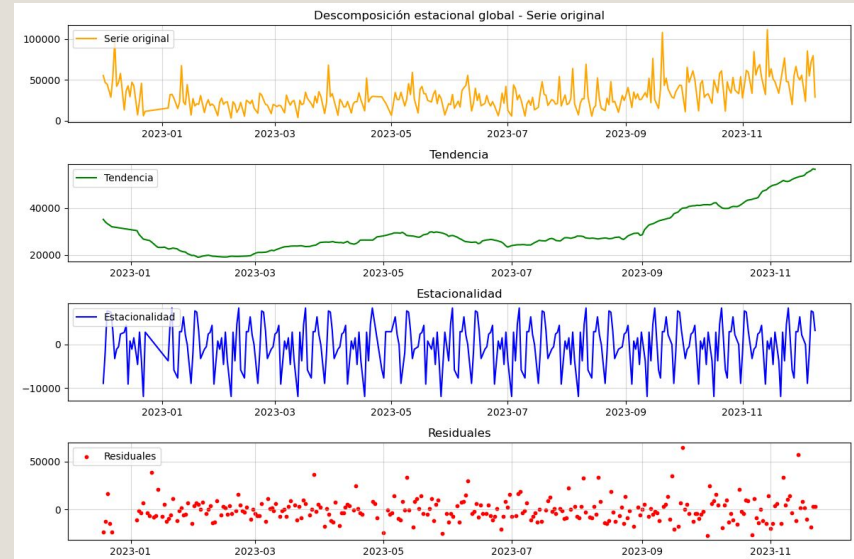
Existe una gran diferencia entre el volumen de venta de cada uno de los países.

Mientras unos países representan un comportamiento más estable otros presentan picos asociados probablemente a una estacionalidad.



METODOLOGÍA

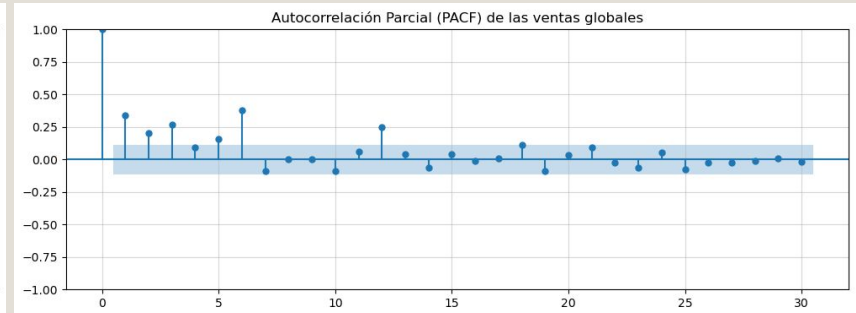
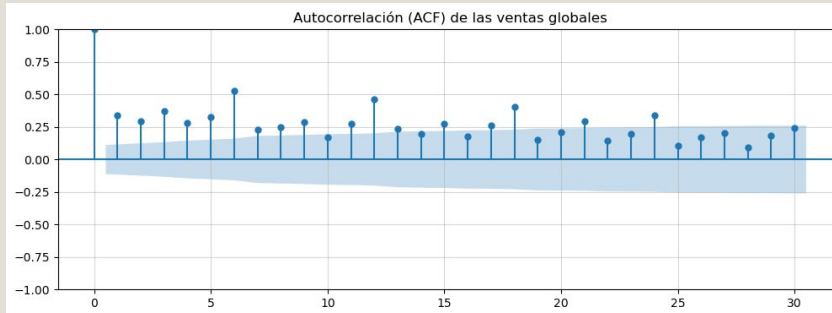
En la serie original se observan picos prominentes y algunas fluctuaciones. En la tendencia se confirma un aumento de las ventas a final del año. Existe una estacionalidad aparentemente recurrente de manera mensual. En la gráfica de residuales podemos observar algunos eventos impredecibles o ruido.



METODOLOGÍA

La autocorrelación muestra una caída gradual en la correlación conforme aumenta el número de lags, lo que sugiere posibles patrones recurrentes (mensual) y picos en lags específicos (semanal).

La autocorrelación parcial muestra una disminución lo que indica una relación importante en los lags más cercanos. También se pueden percibir evento recurrentes.



MODELOS



PROPHET

Prophet funciona mejor con series temporales que tienen efectos estacionales fuertes (ya sean semanales o mensuales) y una cantidad suficiente de datos históricos donde se representa esa estacionalidad. Es robusto ante datos faltantes y cambios de tendencia, y tiende a manejar bien los valores atípicos.

XGBOOST

Es flexible, tiene un rendimiento consistente en problemas de series temporales y soporta control de regularización para evitar sobreajuste. No tiene capacidades de manejo directo de series temporales como Prophet, entonces debemos transformarlas en un problema de regresión supervisada.

MODELOS

11

PROPHET

El MAPE en el conjunto de test (25,3%) es mejor que en el conjunto de train (36,8%). Las diferencias significativas entre las métricas de train y test (MAE y RMSE) sugieren que el modelo podría estar sobreajustando los datos de train.

XGBOOST

El modelo XGBoost muestra un buen rendimiento con un MAPE de 11,9% lo que indica que las predicciones están cerca de los valores reales.

COMPARACIÓN Y SELECCIÓN

Seleccionamos el modelo con el menor MAPE y MAE, ya que indican mayor precisión en las predicciones.

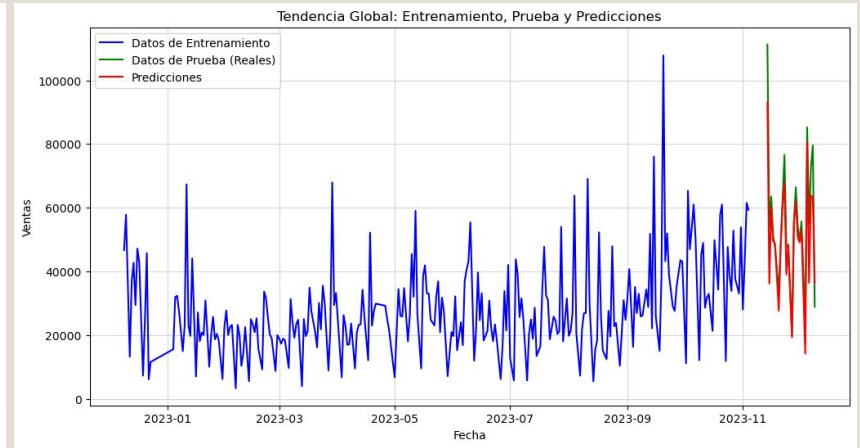
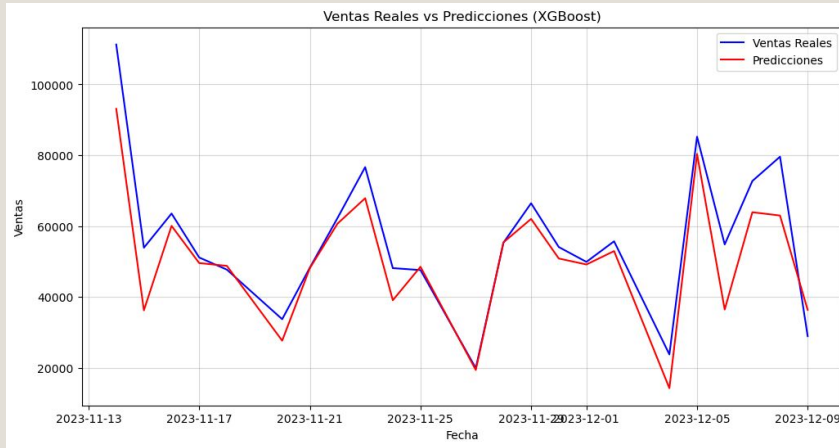
	Modelo	MAE	RMSE	MAPE
0	Prophet	15699.857452	21647.266444	0.253406
1	XGBoost	6344.963083	8742.378414	0.118703

MODELOS

12

XGBOOST

Mostró mejores resultados en términos de menor error de predicción. Esto significa que el modelo logra capturar mejor las variaciones en las ventas, reduciendo la incertidumbre en la estimación del próximo mes.



CONCLUSIONES

01

Tras evaluar distintos enfoques de modelado, se identificó a XGBoost como la mejor opción para la predicción de ventas, gracias a su capacidad para capturar patrones complejos, manejar datos con ruido e incorporar múltiples variables adicionales.

02

El desempeño del modelo podría optimizarse aún más mediante ajustes finos de hiperparámetros, la inclusión de datos históricos adicionales y la exploración de variables externas relevantes, como campañas publicitarias o factores económicos.

03

Las predicciones obtenidas proporcionan una base sólida para la toma de decisiones estratégicas, permitiendo a TodoVentas S.A. planificar mejor sus inventarios, ajustar sus estrategias de marketing y evaluar el impacto de una posible ampliación de capital.