

Forecasting de Demanda Multivariado con XGBoost

Arnau Sastre

[linkedin.com/in/arnausastre](https://www.linkedin.com/in/arnausastre)

September 9, 2025

Abstract

Este informe presenta un sistema de forecasting de demanda para múltiples productos mediante técnicas de aprendizaje automático, con un enfoque multivariado y utilizando el algoritmo **XGBoost**. Se simula un entorno estilo retail con múltiples series de ventas, se aplican técnicas de ingeniería de variables temporales y se entrena un modelo global capaz de predecir la demanda individual de cada producto. Se incluyen métricas de evaluación, análisis de errores y visualizaciones clave para la interpretación de resultados.

1 Introducción

La previsión de demanda es un componente esencial en la gestión de inventario, logística y producción. En escenarios con múltiples productos, se requiere un enfoque que permita capturar las dinámicas específicas de cada artículo sin entrenar un modelo por separado para cada uno.

Este trabajo propone una solución basada en **XGBoost**, un algoritmo eficiente de *boosting* de árboles de decisión, que permite abordar forecasting multivariado de forma tabular, escalable e interpretable.

2 Simulación del entorno multiproducto

Se genera una serie temporal de ventas para múltiples productos (**Product_A**, **Product_B**, **Product_C**) con características realistas:

- Tendencias individuales
- Estacionalidad semanal
- Variación aleatoria controlada

Cada observación contiene la cantidad vendida por producto y una serie de variables exógenas calculadas por ingeniería de features.

3 Ingeniería de variables predictoras

El modelo recibe múltiples variables por producto y por fecha, con el objetivo de capturar patrones temporales y calendáricos. Se incluyen:

- **Lags:** valores anteriores de demanda ($t - 1$, $t - 7$)

- **Rolling windows:** medias y desviaciones estándar móviles
- **Variables calendario:** día de la semana, mes, festivo (booleano)

Estas variables transforman la serie temporal en una estructura tabular apta para modelos tipo XGBoost.

4 Modelo utilizado: XGBoost

XGBoost (Extreme Gradient Boosting) es un algoritmo de boosting basado en árboles de decisión. Para forecasting, permite:

- Trabajar con series no regulares o incompletas
- Incluir múltiples señales de entrada
- Entrenar un único modelo para todos los productos
- Interpretar la importancia de las variables

Entrenamiento

El conjunto de datos es dividido en entrenamiento y test. Se entrena un modelo global con variables codificadas por producto, fecha y comportamiento pasado.

5 Métricas de evaluación

Se utilizaron las siguientes métricas para evaluar la calidad de las predicciones:

MAE (Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = 10.21$$

RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} = 12.95$$

Los valores fueron calculados de forma global y por producto. A modo de ejemplo, para el `Product_C`:

$$MAE_{\text{Product_C}} = 10.21, \quad RMSE_{\text{Product_C}} = 12.95$$

6 Visualizaciones clave

Para facilitar la interpretación de resultados se desarrollaron las siguientes gráficas:

- Comparación **predicción vs. valor real** por producto
- **Error absoluto** por producto
- **Importancia de variables** (feature importance)

Estas visualizaciones permiten entender tanto la calidad de las predicciones como el comportamiento del modelo.

7 Aplicaciones prácticas

El sistema es aplicable directamente a entornos reales, con pequeños ajustes de datos. Algunos casos de uso:

- **Retail:** previsión de demanda por SKU
- **Logística:** planificación de envíos y almacenaje
- **Producción:** dimensionamiento de fabricación según demanda esperada
- **E-commerce:** compras inteligentes y reducción de stock-outs

8 Ventajas del enfoque con XGBoost

- No requiere series continuas o perfectamente alineadas
- Acepta múltiples variables explicativas
- Entrenamiento rápido, apto para producción
- Posibilidad de interpretar las decisiones del modelo

Este tipo de modelo es ideal cuando hay múltiples productos, señales externas, y necesidad de actualización frecuente.

9 Conclusiones

Se ha desarrollado un sistema eficiente de forecasting multivariado basado en **XGBoost**, capaz de predecir la demanda futura de múltiples productos simultáneamente, integrando variables temporales, calendario y comportamiento histórico.

Los resultados muestran buena capacidad predictiva, interpretabilidad y escalabilidad, siendo una solución apta para implementarse en contextos empresariales reales.

Contacto

Si te interesa implementar una solución similar en tu empresa o necesitas ayuda en proyectos de forecasting, optimización o análisis de datos, puedes escribirme por **LinkedIn** o **Malt**. También puedes consultar otros proyectos técnicos en mi **GitHub**.