

Modelo de Propensión de Compra con XGBoost y SHAP

Arnau Sastre

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

August 9, 2025

Abstract

Este informe presenta un sistema avanzado para predecir la probabilidad de compra de clientes mediante un modelo de clasificación basado en **XGBoost** y análisis interpretativo con **SHAP**. Se describe la metodología utilizada, las métricas obtenidas, la optimización del umbral de conversión y ejemplos de aplicación empresarial. El sistema es adaptable a diferentes segmentos de clientes y está orientado a maximizar el retorno de inversión en campañas de marketing.

1 Introducción

En entornos competitivos, conocer la probabilidad de compra de cada cliente permite optimizar recursos de marketing y maximizar el impacto de las acciones comerciales. Este trabajo desarrolla un modelo de propensión de compra con **XGBoost**, complementado con interpretabilidad avanzada mediante **SHAP**, para priorizar clientes y diseñar estrategias segmentadas.

2 Definición del problema

El objetivo es estimar, para cada cliente, la probabilidad de realizar una compra en un periodo determinado. Esto permite:

- Priorización de clientes con mayor probabilidad de conversión.
- Optimización de umbrales de decisión para campañas.
- Simulación de retorno de inversión (ROI) de acciones específicas.
- Estrategias diferenciadas por segmento de cliente.

3 Modelo utilizado

El modelo principal es **XGBoost**, un algoritmo de boosting que combina múltiples árboles de decisión para mejorar la precisión. Se entrenó un modelo de clasificación binaria (*compró* vs. *no compró*) sobre datos simulados de comportamiento de clientes.

Variables simuladas

- Visitas web, clics en emails, productos vistos.
- Descuento aplicado, frecuencia de compra.
- Segmento de marketing.

Interpretabilidad con SHAP

Se aplicó **SHAP** para:

- Identificar las variables más influyentes (importancia global).
- Explicar predicciones a nivel individual (contribución de cada variable).
- Exportar reportes interpretativos por cliente.

4 Optimización del umbral

El umbral de decisión se ajustó para maximizar el F1-score, obteniendo un valor óptimo de 0.43 con $F1 = 0.7912$. Se evaluaron métricas adicionales como precisión, exhaustividad (*recall*), exactitud (*accuracy*) y AUC.

5 Resultados

- Umbral óptimo (F1): $0.43 \rightarrow F1 = 0.7912$
- ROI estimado en campaña sobre top 10% más propensos: $5.20\times$
- Tasa de conversión real del top 10%: 42.00%
- Reporte SHAP exportado: `reporte_shap_clientes.csv`

Entrenamiento por segmento

- Segmento 0: AUC = 0.91, Clientes = 2500
- Segmento 1: AUC = 0.93, Clientes = 1500
- Segmento 2: AUC = 0.95, Clientes = 1000

6 Aplicación empresarial

Este sistema puede integrarse en flujos de trabajo de marketing y ventas para:

- Campañas de marketing predictivo.
- Segmentación y *scoring* de clientes.
- Automatización de *retargeting*.
- CRM inteligente con acciones en tiempo real.
- Fidelización y predicción de abandono (*churn*).

7 Conclusiones

El modelo de propensión de compra con **XGBoost** y **SHAP** permite una priorización precisa y explicable de clientes, optimiza la inversión en marketing y ofrece una interpretación clara de las variables que influyen en la decisión de compra. Su flexibilidad lo hace aplicable en diferentes sectores y bases de clientes.

Contacto

Si te interesa implementar una solución similar en tu empresa, no dudes en escribirme a través de **LinkedIn** o **Malt**. También puedes ver otros proyectos técnicos en mi **GitHub**.