

Detección de fraude en tarjetas de crédito

Problema

Las transacciones fraudulentas representan un costo operativo y de reputación enorme. El principal desafío técnico es detectarlas en un mar de datos legítimos, ya que en nuestro dataset solo el 0.17% de las transacciones eran fraudulentas. Un modelo que no aborde este desbalance extremo es ineficaz.

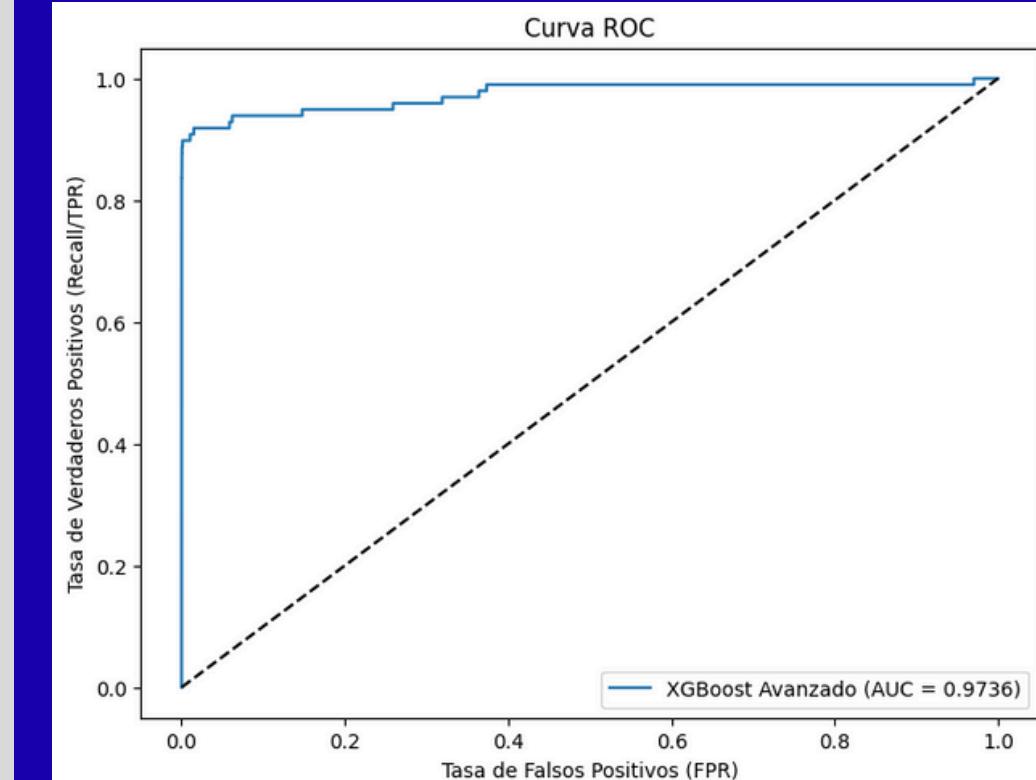
Solución propuesta

Prototipado Rápido en KNIME: Se usó para el Análisis Exploratorio de Datos (EDA), preprocesamiento, y la optimización de un modelo base con XGBoost, estableciendo un benchmark sólido de forma rápida y visual.

Refinamiento en Python: Se migró el flujo a un entorno profesional (VSCode, Jupyter) para aplicar Ingeniería de Características avanzada, creando nuevas "pistas" a partir de las variables de tiempo (transformación cíclica) y monto (indicador de valor cero).

Resultados clave

- # 84% de Fraudes Detectados (Recall)
Capaz de identificar a más de 8 de cada 10 transacciones fraudulentas reales.
- # 89% de Precisión en las Alertas
Cuando el sistema genera una alerta de fraude, es correcta casi 9 de cada 10 veces, minimizando el impacto en clientes inocentes.



Herramientas

- KNIME
- Python
- Pandas
- Scikit-learn & Imbalanced-learn
- XGBoost
- Git & GitHub

Explora el proyecto

