Resumen Ejecutivo del Proyecto de Detección de Fraude en Tarjetas de Crédito

Resumen Ejecutivo

Este proyecto aborda la detección de transacciones fraudulentas en tarjetas de crédito utilizando un dataset altamente desbalanceado. Se aplicó la metodología CRISP-DM para guiar el análisis, desde la comprensión del negocio hasta el despliegue. Se entrenaron modelos de Machine Learning (RandomForest y XGBoost) para identificar patrones que diferencien transacciones legítimas de fraudulentas. Los resultados muestran que XGBoost ofrece un rendimiento superior en términos de ROC-AUC, lo que lo convierte en la opción recomendada para este problema.

Metodología

El proyecto siguió las fases del proceso CRISP-DM, resumidas en la siguiente tabla:

Fase	Nombre	Descripción
1	Comprensión del Negocio	Definir objetivos y criterios de éxito.
2	Comprensión de los Datos	Explorar y analizar la estructura y calidad del dataset.
3	Preparación de los Datos	Limpieza, eliminación de duplicados, escalado y balanceo.
4	Modelado	Entrenamiento de modelos RandomForest y XGBoost.
5	Evaluación	Validación con métricas como ROC-AUC, Precision, Recall.
6	Despliegue	Generación de recomendaciones y documentación.

Principales Resultados

Se compararon dos modelos: RandomForestClassifier y XGBoost. La siguiente tabla resume su rendimiento:

Modelo	ROC-AUC	Características más importantes
RandomForest	0.85	V17, V12, V14, V10, V11, V16
XGBoost	0.98	V17, V12, V14, V10, V11, V16

Recomendaciones

- Priorizar el uso de XGBoost para la detección de fraude por su alto rendimiento.
- Aplicar técnicas de balanceo (SMOTE) y escalado para mejorar la calidad del modelo.
- Monitorear métricas como Precision, Recall y F1-score además de ROC-AUC.
- Implementar un sistema de alertas basado en las variables más discriminantes (V17, V12, V14, V10, V11, V16).
- Actualizar el modelo periódicamente para adaptarse a nuevos patrones de fraude.