Detección de Fraude con Tarjetas de crédito

YENDRI FERREIRA - LEONARDO ARAQUE QUINTERO

May 2025

2 Descripción del problema

2.1 Contexto del problema

El fraude con tarjetas de crédito es un problema que afecta globalmente a instituciones financieras, comerciantes y usuarios. A medida que los métodos de pago se vuelven más digitales y rápidos, también lo hacen las técnicas utilizadas por los estafadores, quienes diseñan patrones de fraude cada vez más sofisticados y difíciles de detectar mediante reglas fijas o análisis manual. En este contexto, las instituciones enfrentan el reto de detectar comportamientos inusuales con precisión y rapidez, sin afectar negativamente la experiencia del cliente legítimo.

Desde la perspectiva de ingeniería y ciencia de datos, este problema representa un caso emblemático para el aprendizaje automático: es un fenómeno raro, costoso, etiquetado, desequilibrado y con un alto impacto real. Los algoritmos tradicionales, como los basados en umbrales o reglas, suelen ser insuficientes frente a la evolución constante de los esquemas de fraude. Por tanto, implementar soluciones de clasificación y detección de anomalías mediante Machine Learning (ML) permite anticiparse a comportamientos sospechosos, actualizarse dinámicamente y detectar casos incluso no observados previamente.

Este proyecto se enmarca dentro de ese propósito: desarrollar un modelo predictivo que permita identificar de forma automática las transacciones fraudulentas a partir de variables derivadas del comportamiento financiero de los usuarios, utilizando técnicas vistas en el curso como árboles de decisión, modelos ensemble, redes neuronales, validación cruzada, regularización y análisis de métricas.

2.2 Composición de la base de datos

Para abordar este problema, se utiliza el conjunto de datos público titulado Credit Card Fraud Detection, disponible en la plataforma Kaggle. Este dataset contiene 284,807 transacciones de clientes europeos recolectadas en el transcurso de dos días, y clasifica cada transacción como legítima o fraudulenta mediante una etiqueta binaria.

La base de datos está compuesta por 31 columnas:

- **Time**: representa el tiempo transcurrido en segundos desde la primera transacción registrada.
- **Amount**: el valor monetario de la transacción.
- **Class**: variable objetivo (0 para transacciones legítimas y 1 para fraudulentas).
- **V1 a V28**: variables transformadas mediante PCA (Análisis de Componentes Principales) para garantizar confidencialidad sobre los datos originales.

Estas variables no tienen una interpretación semántica directa, pero representan proyecciones que capturan la mayor varianza de los datos originales. El objetivo de esta transformación fue anonimizar la información financiera sin comprometer su estructura estadística útil para el modelado.

2.2.1 Características clave de la base de datos:

- **Fuerte desbalance de clases**: solo 0.172% de las transacciones están etiquetadas como fraudulentas (492 casos), lo cual genera un problema serio de desbalance para los modelos de clasificación. Este desequilibrio obliga a considerar técnicas de muestreo, generación de datos sintéticos o enfoques robustos como los modelos basados en árboles con ajustes de pesos.
- **Ausencia de valores faltantes**: ninguna columna contiene datos nulos, lo que permite avanzar directamente con las fases de normalización, selección de variables y modelado.

2.2.2 Distribución de variables:

- Las variables **Amount** y **Time** no están transformadas ni estandarizadas, lo que puede afectar algoritmos sensibles a la escala (como redes neuronales o regresiones). Por tanto, se deben normalizar antes del entrenamiento.
- Algunas variables transformadas mediante PCA (por ejemplo, V10, V14, V17) presentan distribuciones notablemente diferentes entre las clases legítima y fraudulenta, lo que sugiere su utilidad como predictores.
- **Baja multicolinealidad**: al haber sido procesadas por PCA, la mayoría de las variables no están altamente correlacionadas entre sí. Esto sugiere que no hay redundancia severa y permite construir modelos sin necesidad inmediata de técnicas adicionales de reducción de dimensionalidad.

Durante la exploración de los datos realizada por el equipo, se evidenció que algunas variables tenían mayor concentración de valores extremos en las

transacciones fraudulentas, lo cual respalda la hipótesis de que existen patrones numéricos distinguibles. Se destaca también que, en general, las transacciones fraudulentas tienden a involucrar montos inferiores a los de transacciones legítimas, lo cual es coherente con el comportamiento evasivo de los defraudadores.

2.3 Paradigma de aprendizaje adoptado

Dado que el conjunto de datos contiene información etiquetada de forma binaria sobre el carácter fraudulento o no de cada transacción, se adopta un enfoque de aprendizaje supervisado, en el que el modelo aprende a clasificar nuevas instancias a partir de un conjunto de entrenamiento con salidas conocidas.

Este paradigma permite utilizar algoritmos que han sido ampliamente probados en contextos similares, con la posibilidad de cuantificar el rendimiento mediante métricas específicas (precisión, sensibilidad, especificidad, F1-score, AUC-ROC) y aplicar metodologías de validación adecuadas como hold-out o validación cruzada.

Además, se contempla el uso de técnicas de balanceo de clases, como:

- Submuestreo de la clase mayoritaria (undersampling)
- Sobre-muestreo sintético (SMOTE)
- Ajuste de los pesos de clase en la función de pérdida

Este problema se caracteriza por tener costos asimétricos: es más costoso dejar pasar un fraude (falso negativo) que clasificar incorrectamente una transacción válida como fraudulenta (falso positivo). Por ello, el modelo deberá ser especialmente sensible a la clase minoritaria, incluso a costa de una menor precisión global.

3 Estado del arte

La detección de fraude con tarjetas ha sido abordada con diversos enfoques de ML. Analizamos cuatro estudios clave que utilizan el mismo dataset de Kaggle:

3.1 MLP + SMOTE (Zhang & Gong, 2024)

- **Enfoque:** Red neuronal (16-8-1) con sobremuestreo sintético
- **Resultados:** F1=0.84, Recall=92% (mejoró detección de minoría)
- **Conclusión:** SMOTE es efectivo para desbalance

3.2 Supervisado vs No-supervisado (Niu et al., 2019)

- **Comparación:** XGBoost (AUC=0.989) vs Autoencoder (AUC=0.961)
- **Hallazgo:** Métodos supervisados superan en precisión cuando hay etiquetas

3.3 Detección de anomalías (Porwal, 2024)

- **Técnicas:** Isolation Forest (Recall=75%), LOF y One-Class SVM
- **Aplicación:** Útiles cuando el etiquetado es limitado o costoso

3.4 Transformers para datos tabulares (Yu et al., 2024)

- **Innovación:** Adaptación de arquitectura Transformer (AUC=0.993)
- **Ventaja:** Detecta mejor patrones raros vs XGBoost (F1=0.91 vs 0.89)

Tendencia general: Los enfoques supervisados (especialmente ensembles y redes neuronales con balanceo) ofrecen mejores resultados cuando se dispone de datos etiquetados, mientras los métodos no supervisados son útiles como sistemas complementarios.