

# Predicción de Transacciones Fraudulentas utilizando Aprendizaje Automático

Universidad Autónoma de Nuevo León  
Facultad de Ciencias Físico Matemáticas  
Autor: Damián Atilano Martínez Alvarado

## Introducción

El aumento constante de las transacciones financieras ha llevado a un incremento paralelo de las amenazas de fraude. La detección eficiente de transacciones fraudulentas es esencial para garantizar la seguridad de las instituciones financieras y la confianza de los usuarios. En este contexto, este póster presenta un enfoque de predicción de transacciones fraudulentas utilizando técnicas de aprendizaje automático. El objetivo es desarrollar un modelo efectivo para identificar transacciones sospechosas y minimizar el impacto económico del fraude. En este conjunto de datos específico, el comportamiento fraudulento de los agentes tiene como objetivo obtener ganancias tomando el control de las cuentas de los clientes e intentar vaciar los fondos transfiriéndolos a otra cuenta y luego cobrándolos del sistema y también tenemos que observar las transferencias masivas de una cuenta a otra y detectar intentos ilegales.

## Objetivos

- Desarrollar un modelo de predicción:** Implementar un modelo de aprendizaje automático que sea capaz de predecir transacciones fraudulentas en función de patrones y características identificables en los datos transaccionales.
- Evaluar el rendimiento del modelo:** Utilizar métricas de evaluación como la precisión, la sensibilidad y la especificidad para medir la eficacia del modelo en la detección de transacciones fraudulentas.

## Metodología

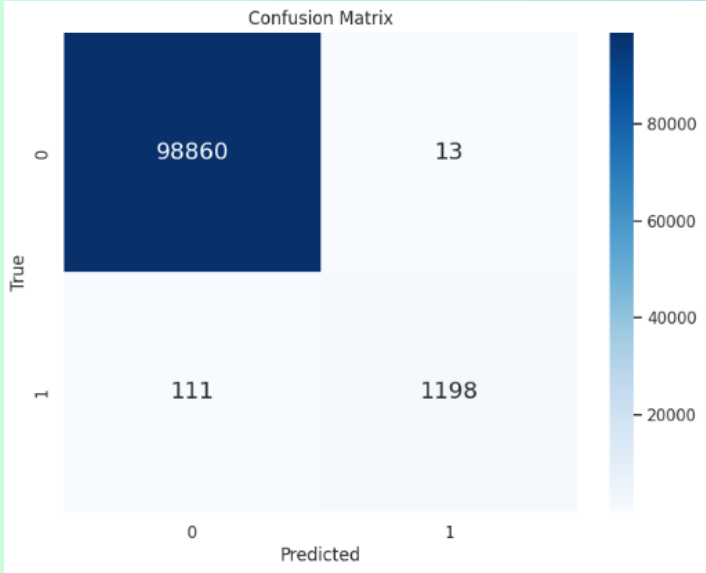
- Recopilación de Datos:** Se recopilan datos de transacciones financieras, que incluyen información sobre el monto, la cuenta de origen y destino, y otras variables relevantes.
- Preprocesamiento de Datos:** Se limpian los datos, buscamos valores atípicos y revisamos valores faltantes. Se normalizan y estandarizan los datos para su uso en el modelo.
- Desarrollo del Modelo de Predicción:** Se implementa un modelo de aprendizaje automático. Se entrena el modelo utilizando los datos de entrenamiento.
- Evaluación del Modelo:** Se evalúa el rendimiento del modelo utilizando métricas. Se ajustan los parámetros del modelo según sea necesario.
- Modelo árbol de decisión:** Los árboles de decisión son como expertos que pueden comprender tanto números como palabras. En el caso de nuestras transacciones, que involucran diferentes tipos de información, como cantidades de dinero y tipos de transacciones, estos árboles pueden entender todo sin la necesidad de realizar procedimientos complicados. Esto es crucial porque simplifica el proceso y facilita la identificación de patrones en los datos financieros de manera más directa.

## Procedimiento

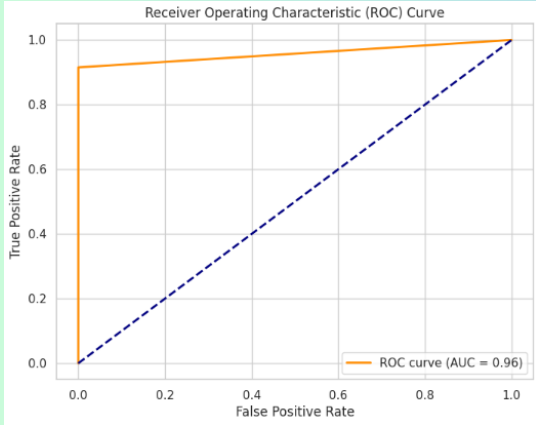
- Preprocesamiento de datos:** Se cargaron los datos y se identificaron las variables relevantes para el modelo de árbol de decisión. Se realizaron algunas visualizaciones y análisis exploratorio para comprender la distribución de las variables y la presencia de posibles valores atípicos. Dado que el modelo de árbol de decisión requiere datos numéricos, se realizó la codificación adecuada de las variables categóricas presentes en el conjunto de datos. Esto facilita la interpretación y procesamiento por parte del algoritmo de aprendizaje automático.
- Entrenamiento del modelo de Árbol de Decisión:** Se dividieron los datos en conjuntos de entrenamiento y prueba. Se entrenó el modelo de árbol de decisión utilizando el conjunto de entrenamiento.
- Evaluamos el modelo:** Se evaluó el modelo utilizando métricas, incluyendo accuracy, precision, recall, y F1-score. Se examinó una métrica adicional como la Curva ROC para una evaluación más detallada.

## Resultados

- Accuracy (Exactitud):** El modelo alcanzó un 99.88% de exactitud, lo que significa que fue altamente preciso en la clasificación general de transacciones como fraudulentas o no fraudulentas.
- Precision (Precisión):** La precisión es del 99% para transacciones no fraudulentas y del 99% para transacciones fraudulentas. Esto indica la proporción de transacciones clasificadas correctamente como fraudulentas entre todas las transacciones clasificadas como fraudulentas por el modelo.
- Recall (Recuperación o Sensibilidad):** El recall es del 100% para transacciones no fraudulentas y del 92% para transacciones fraudulentas. Indica la proporción de transacciones fraudulentas correctamente identificadas por el modelo con respecto al total de transacciones fraudulentas en el conjunto de datos.
- F1-Score:** El F1-Score, que combina precisión y recuperación, es del 99% para transacciones no fraudulentas y del 95% para transacciones fraudulentas.



- True Positives (TP):** 1198 - Transacciones fraudulentas que fueron correctamente clasificadas como fraudulentas por el modelo.
- False Negatives (FN):** 111 - Transacciones fraudulentas que fueron incorrectamente clasificadas como no fraudulentas por el modelo.
- True Negatives (TN):** 98860 - Transacciones no fraudulentas que fueron correctamente clasificadas como no fraudulentas por el modelo.
- False Positives (FP):** 13 - Transacciones no fraudulentas que fueron incorrectamente clasificadas como fraudulentas por el modelo.
- La tasa de falsos positivos (FPR) es prácticamente nula, lo que indica que el modelo cometió muy pocos errores al clasificar transacciones no fraudulentas como fraudulentas.
- La tasa de verdaderos positivos (TPR), es del 91.52%. Esto significa que el modelo identificó correctamente el 91.52% de todas las transacciones fraudulentas presentes en el conjunto de datos.



El AUC-ROC Score es del 95.75%. Esta métrica proporciona una evaluación global del rendimiento del modelo, siendo 1 el rendimiento perfecto. Un valor alto como este indica una buena capacidad del modelo para distinguir entre transacciones fraudulentas y no fraudulentas.

## Conclusión

El modelo de árbol de decisión tiene un desempeño excelente al identificar transacciones fraudulentas. Su gran precisión y pocas predicciones incorrectas lo convierten en una herramienta efectiva para detectar fraudes en operaciones financieras.

Cuando vemos los resultados, notamos que el modelo puede prever con certeza y en su totalidad (100%) las transacciones normales. Sin embargo, es importante señalar que la capacidad para prever transacciones fraudulentas, aunque alta (92%), podría mejorarse. Esto sugiere que realizar algunos ajustes al modelo podría hacerlo más capaz de descubrir transacciones fraudulentas. La baja tasa de predicciones incorrectas, respaldada por la Curva ROC y el AUC-ROC Score del 95.75%, dice que el modelo puede equilibrar muy bien la precisión y la capacidad para encontrar fraudes.

Este análisis indica que, aunque el modelo es bastante preciso en general, hay maneras de hacerlo mejor en la detección de transacciones fraudulentas. Se sugiere probar algunos cambios y ajustes adicionales, así como usar métodos para equilibrar mejor los datos. Además, es importante tener en cuenta características específicas cuando los conjuntos de datos tienen más ejemplos de una clase que de otra. Estas mejoras podrían hacer que el modelo sea aún más efectivo para prevenir fraudes en el ámbito financiero.