Guatemala 02/06/2025

Alejandro Martínez – 21430

**Security Data Science**

Proyecto final

Detectar fraudes en entretenimiento

Optimizar el modelo para fraudes en lugares como cines, conciertos, bares, etc.

# 1. Resumen

Este proyecto aborda el reto de detectar fraudes en tarjetas de crédito a partir de un conjunto de datos altamente desbalanceado, donde las transacciones fraudulentas representan una fracción muy pequeña del total. Se exploran y comparan diversas estrategias de modelado y métricas de evaluación, con el objetivo principal de maximizar la tasa de detección de fraudes (recall) y minimizar los falsos positivos, especialmente en usuarios que consumen en el sector de entretenimiento.

A lo largo del desarrollo, se aplicaron técnicas avanzadas de ingeniería de características y se diseñaron métricas personalizadas alineadas con los objetivos del negocio, priorizando la minimización del impacto negativo sobre la experiencia del usuario legítimo. Dado que los falsos positivos pueden llevar a la suspensión injustificada de tarjetas y pérdida de confianza del cliente, se enfatizó el desarrollo de soluciones que identificaran patrones reales de fraude sin castigar excesivamente las variaciones normales del comportamiento de consumo.

Se hizo especial énfasis en detectar transacciones anómalas dentro de contextos específicos, como compras en horarios atípicos, días festivos o establecimientos de entretenimiento. Esta focalización respondió a una necesidad del negocio de mejorar la precisión en este segmento, donde los hábitos de consumo pueden ser altamente variables.

# 2. Metodología

El análisis comenzó con una exploración exhaustiva de los datos para comprender la magnitud del desbalanceo (menos del 0.2% de las transacciones eran fraudulentas) y para identificar las variables que podrían ser más predictivas. Se analizaron tendencias temporales, frecuencias de compra, importes medios por categoría y patrones geográficos, con especial atención a las transacciones vinculadas al entretenimiento.

Durante la etapa de ingeniería de variables, se diseñaron características personalizadas basadas en el comportamiento histórico de cada usuario. Por ejemplo, se identificaron patrones como la frecuencia de compras en horario nocturno, gasto promedio en fines de semana, y la regularidad con la que un usuario realiza compras en fechas festivas o en locales de ocio. Estas variables permitieron enriquecer la representación del comportamiento del cliente y mejorar la discriminación entre transacciones normales y potencialmente fraudulentas.

Para el modelado, se empleó LightGBM por su eficiencia y capacidad para manejar grandes volúmenes de datos con características categóricas y continuas. El modelo fue optimizado utilizando una métrica de costo personalizada (fp\_ratio) que penaliza más severamente los falsos positivos, buscando el equilibrio entre la sensibilidad del modelo y la experiencia del usuario. Se aplicaron técnicas de re-muestreo como el undersampling para manejar el desbalance de clases sin comprometer la información crítica de la clase minoritaria.

Finalmente, en el análisis de errores se revisaron manualmente los casos de falsos positivos, lo que permitió identificar patrones comunes e inconsistencias en las predicciones. Esta revisión retroalimentó el proceso de ingeniería de características, permitiendo la incorporación de nuevas variables que mejoraron la capacidad del modelo para distinguir entre comportamientos atípicos legítimos y fraudulentos.

# 3. Implementación

La implementación técnica del proyecto siguió una secuencia estructurada, comenzando con la limpieza y preprocesamiento del dataset original. Se identificaron y eliminaron registros duplicados y se imputaron valores faltantes cuando fue necesario, asegurando así la calidad de los datos de entrada. Se evaluaron y eliminaron variables redundantes o irrelevantes para reducir la dimensionalidad y evitar el sobreajuste.

Posteriormente, se implementaron scripts para la creación automatizada de nuevas variables basadas en reglas de comportamiento, incluyendo ventanas móviles de gasto y contadores temporales por categoría. Se aplicaron transformaciones logarítmicas y escalado cuando fue requerido para ciertas variables numéricas con distribución sesgada.

El conjunto de datos limpio y enriquecido se dividió en subconjuntos de entrenamiento y prueba, asegurando una representación balanceada en la muestra de prueba. Se aplicó undersampling al conjunto mayoritario para facilitar el entrenamiento del modelo, sin comprometer la información representativa del patrón normal de transacciones.

Una vez entrenado el modelo, se aplicó una métrica personalizada dentro de la función de pérdida, lo que permitió priorizar la reducción de falsos positivos, especialmente en las categorías sensibles. Las predicciones del modelo fueron almacenadas y evaluadas utilizando tanto métricas estándar (accuracy, precision, recall, F1) como las personalizadas para evaluar el rendimiento real en el contexto del negocio.

# 4. Análisis de Resultados

El modelo especializado logró una mejora del **9.36%** en recall comparado con un modelo base sin ingeniería de variables ni métrica personalizada. Esta mejora, aunque modesta en términos absolutos, representa una optimización significativa en un entorno altamente restrictivo por el bajo volumen de fraude y la variabilidad del comportamiento de los usuarios.

A pesar de la baja correlación directa entre transacciones fraudulentas y el sector de entretenimiento, la estrategia de segmentación permitió aislar ciertos patrones útiles. Por ejemplo, se identificó que varias transacciones fraudulentas simulaban patrones normales pero eran ejecutadas con tarjetas nuevas o en ubicaciones atípicas. La inclusión de variables geográficas y de historial transaccional fue crucial para estas detecciones.

El modelo logró un accuracy general superior al 99.9%, y en particular, de **5784 transacciones en entretenimiento, solo 29 resultaron en falsos positivos**, lo que representa una tasa de error de tan solo 0.5% en esa categoría. Esta precisión es altamente valiosa en contextos donde una mala clasificación puede llevar a la cancelación innecesaria de servicios para clientes fieles.

Sin embargo, también se detectaron limitaciones importantes, especialmente en casos donde no existía suficiente historial del usuario o cuando los hábitos de consumo eran altamente erráticos. Estas observaciones indican la necesidad de integrar fuentes externas (por ejemplo, geolocalización en tiempo real o redes de dispositivos) en futuras iteraciones del modelo.

# 5. Conclusiones

* El desbalance extremo de clases representa un desafío crítico al momento de construir modelos de clasificación efectivos para la detección de fraude. Técnicas como el undersampling y la creación de métricas personalizadas resultaron esenciales para superar esta barrera.
* La detección de fraude en el sector de entretenimiento presenta retos particulares debido a la alta variabilidad y espontaneidad del comportamiento de los usuarios. Es difícil establecer perfiles estables para cada cliente, lo que complica la diferenciación entre comportamiento anómalo legítimo y fraudulento.
* La ingeniería de características fue el componente más determinante del éxito del modelo. Variables derivadas del comportamiento del usuario, contextualizadas con temporalidad y categoría, permitieron incrementar significativamente la precisión del sistema.
* A futuro, se recomienda explorar técnicas de aprendizaje semi-supervisado o por refuerzo, además de incorporar información contextual externa y arquitecturas de aprendizaje profundo que puedan identificar patrones complejos no lineales.