



Data scientist technical challenge

Los pagos realizados a través de Mercadopago han crecido exponencialmente mes a mes los últimos años, al punto tal de que hoy en día el volumen de transacciones realizadas por fuera del marketplace superan a las realizadas internamente.

Parte de este crecimiento está en la prevención de fraude en los distintos productos de MercadoPago.. Esto es vital para garantizar una excelente experiencia para los usuarios y reducir las pérdidas económicas por fraudes.

Contexto

El fraude en plataformas de pagos, es una problemática mundial que representa pérdidas económicas a gran escala. MercadoPago, por su dimensión y presencia regional, es blanco de este tipo de ataques diariamente. Construir modelos de machine learning para prevención de fraude robustos y escalables que analicen todo el flujo de pagos es indispensable para la seguridad de nuestros usuarios.

Enunciado

Se desea construir una solución basada en Machine Learning para **predecir si una transacción es fraudulenta o no**, en base a un conjunto de datos proveniente del sistema de prevención de fraude de MercadoLibre.

Está permitida la utilización de herramientas de AI (como ChatGPT, Gemini u otras) para generar ideas, código, análisis o sugerencias en cualquier etapa del desafío. Caso lo hagas, será obligatorio:

1. Documentar explícitamente en tu reporte el uso de AI, incluyendo:
 - a. El/Los prompts exactos que utilizaste
 - b. El resultado generado por la IA (por ejemplo, el texto o código devuelto)
 - c. Cómo integraste o iteraste sobre ese resultado (por ejemplo, qué modificaciones hiciste o qué conclusiones sacaste)
2. Asegurarte de que el código presentado sea claro y pueda ejecutarse. Sugerimos que agregues los comentarios pertinentes para que se pueda seguir la lógica subyacente. Es importante que puedas explicar y justificar cada paso del material que presentes
3. Que expliques el razonamiento detrás del uso de IA y cómo contribuyó al resultado final.

Con estas consideraciones, se pide:

1. Resolver el problema de **clasificación de transacciones en fraude / no fraude**. Sabemos que **por cada transacción aprobada** el porcentaje de ganancia es de un 25%, y **por cada fraude aprobado** se pierde el 100% del dinero de la transacción. Realizar un análisis y determinar un modelo que permita maximizar la ganancia de la empresa.

2. Realizar un informe de la solución propuesta, abordando al menos los siguientes tópicos:
 - Hipótesis: Introducción y desarrollo
 - Análisis y transformaciones del dataset
 - Modelos utilizados
 - Evaluación
 - Conclusión

El informe debe ser presentado en PDF, o en una notebook específica desarrollada para tal fin. Recomendamos que incluyas en el informe una sección dedicada a la utilización de AI, con los detalles que se mencionan anteriormente.

3. Disponibilizar el código fuente utilizado para el análisis en Github o [Colab](#), junto con los pasos necesarios para su ejecución. Tené en cuenta que:
 - La calidad del código será tomada en cuenta para la evaluación. Recomendamos que escribas un código prolífico y comentado.
 - El código debe poder ser ejecutado por los revisores siguiendo los pasos indicados, obteniendo resultados similares a los indicados en la evaluación y conclusión.

El dataset “MercadoLibre Inc. Data Scientist Hiring Test - Fraud Dataset - Data.csv” se encuentra adjunto en el presente mail.

ACLARACIÓN: El formato de los números se encuentra con el delimitador . (punto) para la separación de decimales y , (coma) para miles.