



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. Se pide:

1. Resolver el problema de clasificación de transacciones en fraude / no fraude. Sabemos que por cada transacción 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 pequeño informe de la solución propuesta:
 - Hipótesis (Introducción y desarrollo)
 - Análisis y transformaciones del dataset
 - Modelos utilizados
 - Evaluación
 - Conclusión

El informe debe ser entregado en formato PDF.

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 prolijo y comentado.

- El código debe poder ser ejecutado por el corrector siguiendo los pasos indicados, obteniendo resultados similares a los indicados en el informe.

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.