



Data scientist technical challenge

Las compras en Mercado Libre han crecido exponencialmente los últimos años, al punto tal de que hoy en día el volumen de compras supera las miles de transacciones por segundo.

Este crecimiento ha traído también desafíos en la creación de sistemas de prevención de fraude que permitan garantizar una excelente experiencia para los usuarios y reducir las pérdidas económicas por fraudes.

El fraude en plataformas de pagos es una problemática mundial que representa pérdidas económicas a gran escala. Mercado Libre, por su dimensión y presencia regional, es blanco de este tipo de ataques diariamente.

Frente a esta realidad, la construcción de 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 Mercado Libre.

Se pide:

1. Entrene un algoritmo con el fin de predecir el fraude sabiendo 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
3. ¿Qué pasos puedo seguir para intentar asegurar que la performance del modelo en laboratorio será similar a la de producción?



4. Suponiendo que la performance predictiva en producción es muy diferente a la esperada, ¿Cuáles cree que son las causas más probables?
5. ¿Qué pasos debería seguir para poner el nuevo modelo en producción?

El dataset se encuentra disponible en la siguiente url:

[MercadoLibre Data Scientist Technical Challenge - Dataset.csv](#)

El informe debe ser entregado en formato PDF.

Disponibilizar el código fuente utilizado para el análisis en Github o Colab, junto con los pasos necesarios para su ejecución. Ten 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.