

EVERTEC

Challenge Equipo Data Science

Descripción:

Como integrante del equipo Data Science dentro de Evertec, se le solicita genere un primer acercamiento sobre un nuevo proyecto, donde el objetivo es desarrollar un modelo de Machine Learning aplicado a la prevención de fraude para una institución que cuenta con sucursales en varios pases de LATAM.

Dentro de las tareas esperadas se encuentran:

- Limpieza de datos
- Análisis de los datos
- Ingeniería de atributos
- Modelado
- Comparativa y selección final

Para esto, se le brinda un dataset reducido de 40.000 transacciones aproximadamente con sus correspondientes marcas de fraude.

Dado que este es su primer proyecto, también se le facilita un diccionario con una breve descripción de las variables, con la particularidad de que, al ser un cliente nuevo, aún no se cuenta con la identificación total de los códigos y categorías dentro de las variables, pero de todas formas se le indican cuales son numéricas y cuales categóricas para poder efectuar correctamente el análisis y modelado.

El objetivo de esta primera instancia es conocer los datos y generar insights en cuanto al comportamiento del fraude y las características que lo identifican, para poder generar variables que logren capturarlo.

En cuanto al modelado, se busca encontrar el que mejor resultados obtenga, por lo que se deberá seleccionar la métrica que mejor sirva para comprar de forma objetiva los algoritmos implementados.

Entregables:

- Notebook con todo el código desarrollado
- Un documento explicando el proceso realizado, los criterios aplicados y análisis de los resultados obtenidos.

Extras

- Utilizar librería MLflow
- Implementación algoritmo CatBoost
- Optimización de hiperparámetros mediante Optuna o Hyperopt

Features

ID_TRX: Identificador único de transacción

S_PAN: Número de tarjeta original

S_ENCRYPTED_PAN: número de tarjeta encriptado. Los últimos 4 dígitos coinciden con lo del S_PAN

F_AMOUNT_TRANSACTION: monto original de la transacción

N_CURRENCY_CODE_TRANSACTION: moneda de la transacción - *categorica*

F_DOLLAR_AMOUNT: monto expresado en dólares

N_TRANSMISSION_DATE_AND_TIME: fecha de la transacción expresada de forma numérica

N_MERCHANT_TYPE: clasificación de comercio - *categorica*

N_ACQ_INSTITUTION_COUNTRY_CODE: código del país donde se realizó la transacción - *categorica*

N_ENTRY_MODE: forma en que se ingresó la información de la tarjeta al momento de realizar la compra – *categorica*

N_POINT_OF_SERV_COND_CODE: información sobre el POS donde se realizó la transacción – *categorica*

S_MERCHANT_LEGAL_NAME: nombre legal del comercio

FRAUDE: indica si la transacción fue o no fraude - OBJETIVO