

Modelo de detección de fraudes

Descripción: El [conjunto de datos](#) que se le proporciona para este ejercicio, contiene datos de transacciones realizadas con Tarjetas de Crédito (TC) de clientes que pagan a comerciantes. En esencia, una transacción con TC consiste en cualquier monto pagado a un comerciante por un cliente en un momento determinado. El dataset contiene un total de 23 variables y más de 200.000 transacciones. Del conjunto de variables se describen a continuación las más relevantes para el análisis:

- ID de la transacción: un identificador único para la transacción.
- Fecha y hora: fecha y hora en la que se produce la transacción.
- ID del cliente: el identificador del cliente. Cada cliente tiene un identificador único.
- ID del terminal: El identificador del comerciante (o más precisamente del terminal). Cada terminal tiene un identificador único
- Monto de la transacción: el monto en dólares de la transacción.
- Etiqueta de fraude: una variable binaria, con el valor 0 para una transacción legítima o el valor 1 para una transacción fraudulenta.

Estas características, en el dataset, se denominan: TRANSACTION_ID, TX_DATETIME, CUSTOMER_ID, TERMINAL_ID, TX_AMOUNT y TX_FRAUD. El resto de las variables son menos importantes, pero deben formar parte de la etapa de pre-procesamiento de datos, modelado y evaluación.

Se le pide que, con el dataset en cuestión, Ud. pueda construir un modelo de detección de fraude. Para ello deberá implementar las siguientes etapas:

1. Ingeniería de atributos (5 puntos)
2. Entrenamiento del modelo (2 puntos)
3. Validación del modelo (3 puntos)

La entrega se debe hacer en un notebook de google colab que tenga una buena documentación de cada uno de los pasos realizados, las decisiones de diseño y las

Minería de Datos I

Examen final de curso

conclusiones. En la etapa de validación se espera que pueda proveer un análisis de los resultados y comparación entre diferentes modelos.

El notebook debe compartirlo con el Profesor para poder ser evaluado.

La fecha/hora tope de entrega será el miércoles 13-12-2023 a las 23:59hs.

Link al conjunto de datos:

https://drive.google.com/file/d/1cWvkOy0Jy6Q5ICx_2xAVEH7x1zrj8jd/view?usp=drive_link

Suerte