

IEEE - Detección de fraude en transacciones electrónicas

**Evelyn Zharick Saez Gallego
Aura Luz Moreno Díaz**

INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

Raúl Ramos Pollan

UNIVERSIDAD DE ANTIOQUIA

2023

DESCRIPCIÓN DEL PROBLEMA PREDICTIVO

El instituto de Ingenieros en Electricidad y Electrónica es la organización técnica profesional más grande y prestigiosa del mundo, fue fundada en New York, el 13 de Mayo de 1884 por un grupo de profesionales entre los cuales están: Thomas Alva Edison, Alexander Graham Bell y Frankin Leonard Pope. (<https://www.ieee.org/>,2023)

IEEE ha realizado un convenio con la empresa VESTA quien ha facilitado los Dataset. Vesta es una empresa dedicada a detectar fraudes electrónicos desde 1.995. Presta sus servicios a nivel internacional. Para más información: [Payment Fraud Prevention Services | Vesta](https://www.vesta.io) (<https://www.vesta.io>)

En esta competición elegida en el sitio de Kaggle, deseamos predecir la probabilidad de que una transacción online sea fraudulenta, esto, usando una variable booleana (o binaria) llamada **IsFraud**.

El dataset está dividido en dos archivos: **Identity** y **Transaction**, los cuales están cruzados por **TransactionID**

DATASET

https://www.kaggle.com/competitions/ieee-fraud-detection/data?select=test_transaction

Tenemos las siguientes variables categóricas:

IDENTITY

- DeviceType
- DeviceInfo
- id_12 - id_38

TRANSACTION

- ProductCD
- card1 - card6
- addr1, addr2

- P_emaildomain
- R_emaildomain
- M1 - M9

La Variable **TransactionDT** es una **es un delta de tiempo** de una referencia de variable de tiempo (no es un timestamp)

Tenemos los siguientes archivos:

- **train_{transaction, identity}.csv** - the training set
- **test_{transaction, identity}.csv** - the test set (you must predict the `isFraud` value for these observations)
- **sample_submission.csv** - a sample submission file in the correct format

Tenemos esta base de datos alternativa en caso de que nuestro dataset elegido no funcione para fines académicos: <https://www.kaggle.com/competitions/amex-default-prediction/code>

MÉTRICAS DE DESEMPEÑO (De machine learning y de negocio)

Estamos ante un problema de clasificación binaria, ya que queremos determinar si una transacción financiera es fraude o no. Es por esto que podríamos utilizar algunas como:

Accuracy: Podríamos determinar la proporción de predicciones correctas con relación al total de predicciones.

Recall: Medir la proporción de falsos positivos con relación al total de falsos positivos

F1-Score: Combinando las dos anteriores

Curva ROC y AUC: para determinar los falsos positivos y los verdaderos positivos a medida que variamos el umbral de clasificación del modelo. También determinando el área bajo la curva, para el AUC, para determinar desde 0 (clasificador aleatorio) hasta el 1 (clasificador ideal).

Entre las métricas de negocio tenemos:

Costo total del fraude: Podríamos detectar el número de transacciones fraudulentas vs el número total de transacciones, determinando su costo.

Falsos positivos: Son los números de transacciones clasificadas como fraudes que en realidad no son. Esto podría acarrear problemas ya que el banco podría abrir procesos disciplinarios o legales a entidades/personas que en realidad no están haciendo fraude y esto supone un problema de servicio al cliente que podría traer pérdidas económicas

Tiempo de respuesta: Cuánto se demora la entidad en detectar un fraude electrónico.

DESEMPEÑO DESEABLE

Lo que esperamos con este análisis es poder determinar la cantidad de fraudes dentro de un conjunto de transacciones establecido de manera tal que podamos hacer campañas de prevención y detección a tiempo para evitar pérdidas económicas dentro del negocio a futuro.

Acortando los tiempos de espera mediante la detección temprana se ahorraría dinero, así también detectar los falsos positivos mediante el accuracy (o precisión) de manera tal de ser más eficientes al momento de determinar si una transacción es fraudulenta o no.

BIBLIOGRAFIA

IEEE – CIS Fraud Detection | Kaggle (2020). Retrieved 12 March 2023, from: <https://www.kaggle.com/competitions/ieee-fraud-detection/>