Team Power Rangers

FRAUD DETECTION FOR ONLINE PAYMENT PLATFORM

Outline

- Introduction
- Data Exploration and Feature Engineering
- Model Evaluation

Start Page ›

Introduction

Context

The online payment platform processes millions of transactions daily, making it vulnerable to various types of fraudulent activities. These activities pose a significant threat to both our business and our customers. To safeguard the platform and enhance user experience, we aim to leverage the power of data science and machine learning to proactively detect and prevent fraudulent transactions.

Goals for the EDA

To investigate:
Which are the top 10 transaction types by transaction amounts?

Which are the longest average transaction times?

Who are the user's with the highest credit scores?

Frequently used browser type for transactions among users?

Prevailing transaction types characterized by amounts?

Dominant users' occupation?

Summary of steps

The Steps taken are:

Understanding Business Problem

Data Cleaning

Exploratory Data Analysis

Model Building

Presentation

Data Exploration

Dataset Overview

The dataset has 6000000 rows (for observations) and 32 columns (for variables).

The variables are of different data types, including strings, floats, and numerics.

The dataset did not contain null values and duplicate values.

Column Titles

The column titles include;

Transaction ID; User ID; Transaction Amount; Transaction Date and Time; Merchant ID; Payment Method; Country Code; Transaction Type; Device Type; IP Address; Browser Type; Operating System; Merchant Category; User Age; User Occupation; User Income; User Gender; User Account Status; Transaction Status; Location Distance; Time Taken for Transaction; Transaction Time of Day; 'User's Transaction History; Merchant's Reputation Score; 'User's Device Location; Transaction Currency; Transaction Purpose; User's Credit Score; User's Email Domain; Merchant's Business Age; Transaction Authentication Method; Fraudulent Flag

Data Exploration

Data Exploration

The Python programming language was used to handle the EDA cleaning processes, and steps were documented in a Python notebook. Because not all columns are required to answer the questions in this task, some columns will be removed. Removing these unnecessary variables will help focus on the variables that will focus on the business objective in this task.

Feature Engineering

This involved the transforming of features (variables) to improve the machine learning model in detecting fraudulent transactions. It aimed to provide the model with more relevant and discriminating information to make accurate predictions.

Model Evaluation

The model was trained using Random Forest Classifier. PCA was used to reduce computation time due to the large size of the dataset.

The model had an accuracy of 0.50 suggesting the model’s performance was moderate, with accuracy and precision, recall and F1-score all around 0.50, which indicates that the model is not performing significantly better than random chance.

The model was improved using Hyperparameter tuningto optimize the hyper parameters of the model.