Table of Contents

I.   Problem Description

Detecting fraud for transactions in a payment gateway

A new disruptive payment gateway start-up, 'IndAvenue', has started gaining traction due to its extremely low processing fees for handling online vendors' digital payments. This strategy has led to very low costs of acquiring new vendors.

Unfortunately, due to the cheap processing fees, the company was not able to able to build and deploy a robust and fast fraud detection system. Consequently, a lot of the vendors have accumulated significant economic burden due to handling fraudulent transactions on their platforms. This has resulted in a significant number of current clients leaving IndAvenue's payment gateway platform for more expensive yet reliable payment gateway companies.
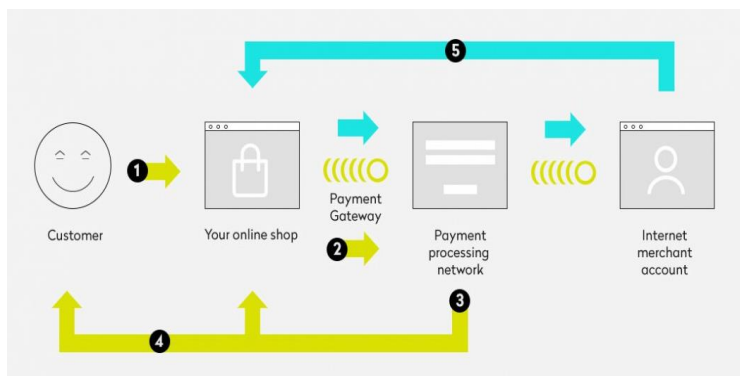
The company's data engineers curated a dataset that they believe follows the real world distribution of transactions on their payment gateway. The company hired Insofe and provided it with the dataset, to create a fast and robust AI based model that can detect and prevent fraudulent transactions on its payment gateway.

They have provided you with the dataset that has the `**is_fraud**` column, which encodes the information whether a transaction was fraudulent or not.

In this hackathon, you will now have to use this curated data to create a machine learning model that will be able to predict the `**is_fraud**` column.

II. Dataset Description

Target attribute: "is_fraud" (discrete variable: 2 classes)

There are 3 CSV files provided to us, they are described below:

- **train_data.csv**: Every transaction in the database has a unique *'transaction_number'* and an associated *'is_fraud'* target label along with several features associated with the transaction.

- **test_data.csv**: The test data has all the columns provided in the train dataset other than the target column.

- **sample_submission.csv**: The format of CSV file required for submission to the evaluation backend

Note: Negative values in the money_transacted column relate to amount credited, while positive values signify amount debited

III. Hackathon Tasks

As part of this hackathon, you are expected to complete three tasks: i) Exploratory Analysis ii) ML Modelling iii) Recommendations and Deployment Strategies

Exploratory analysis

Exploratory Data Analysis using visualizations, numerical analysis, and describing the findings.

- List down the insights/patterns observed from the visualizations

- Explain the impact of the most important attributes on the target attribute observed from data visualizations.

ML Modelling

You are expected to create a robust fraud detection framework by engineering new features, tuning, and improving the baseline ML model performance.

Recommendations to the business

- What are your recommendations for IndAvenue?

- Any final visualizations you would use to convey your recommendations?

- Can you explain your ML model using non-technical terms?

- Which simple strategies would you use to ensure fast customer checkout?

IV. Evaluation Metric

- The evaluation metric for this hackathon is the **F1 Score**