# DATASCIENCE PHASE 2

# CREDIT CARD FRAUD DETECTION

## PROBLEM STATEMENT:

The problem is to develop a machine learning-based system for real-time credit card fraud detection. The goal is to create a solution that can accurately identify fraudulent transactions while minimizing false positives. This project involves

- data preprocessing
- feature engineering
- model selection
- training
- evaluation to create a robust fraud detection system.

## Main challenges involved in credit card fraud detection:

- ➢ Enormous Data is processed every day and the model build must be fast enough to respond to the scam in time.
- ➢ Imbalanced Data i.e most of the transactions (99.8%) are not fraudulent which makes it really hard for detecting the fraudulent ones
- ➢ Data availability as the data is mostly private.
- ➢ Misclassified Data can be another major issue, as not every fraudulent transaction is caught and reported. Adaptive techniques used against the model by the scammers.

## SOLUTION:

A model that will provide the best results in revealing and preventing fraudulent transactions. This is achieved through bringing together all meaningful features of card users' transactions, such as Date, User Zone, Product Category, Amount, Provider, Client's Behavioral Patterns, etc.

The information is then run through a subtly trained model that finds patterns and rules so that it can classify whether a transaction is fraudulent or is legitimate.

**STEPS TO BE FOLLOWED:**

**1.Exploratory Data Analysis**

**Dataset Link: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud**

The dataset that is used with this proposed approach is a real-world dataset obtained from Kaggle . It contains transactions made by credit cards  The raw dataset taken for the study was sorted and pre-processed for the sole intention of improving the performance of the classifiers and reducing their training and operating time.

**2.Tools and libraries :**

- Numpy – 1.19.2
- Python – 3.x
- Scikit-learn – 0.24.1
- Matplotlib – 3.3.4
- Imblearn – 0.8.0
- Collections, Itertools

**3.Algorithm**

Implementing anomaly detection techniques to identify potentially fraudulent credit card transactions.

One-class SVM is algorithm for anomaly detection that is based on the concept of maximum margin hyperplanes. It works by creating a hyperplane that separates the normal data points from the anomalies and identifying points that lie on the wrong side of the hyperplane as anomalies.

## 4.Train-test split

The train data can be used to train our model whereas the unseen data can be used to test our model. we are familiar with the train/test split, which we can perform in order to check the performance of our models with unseen data.

- x_train: It is used to represent features for the training data
- x_test: It is used to represent features for testing data
- y_train: It is used to represent dependent variables for training data
- y_test: It is used to represent independent variable for testing data

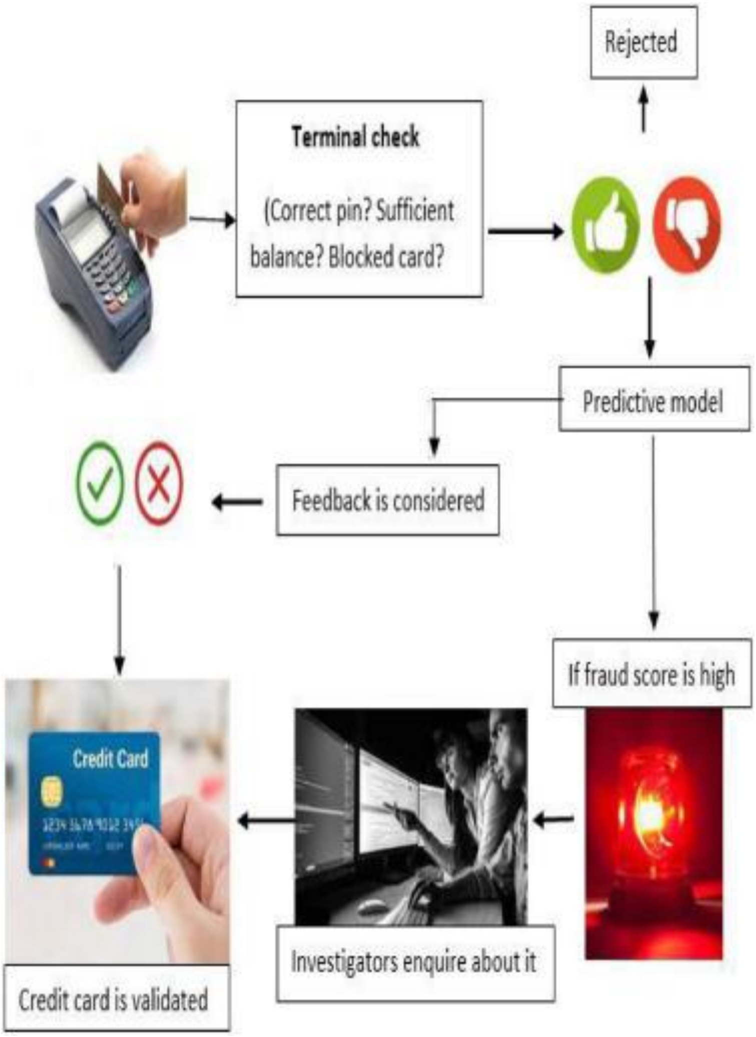
## 5.Modeling

This is the final step at which we can try different models and fine-tune their hyperparameters until we get the desired level of performance on the given dataset. We should try and see if we get a better model by the various sampling techniques.

## 6.Evaluating Final Model Performance

Evaluate the models using appropriate evaluation metrics. Note that since the data is imbalanced it is is more important to identify which are fraudulent transactions accurately than the non-fraudulent. We need to choose an appropriate evaluation metric which reflects this business goal.

## FLOWCHART:



**NAME : Akshya S**
**REG NO: 420421104005**