Problem Statement of Credit Card Fraud Detection

Objective:

The primary goal of this project is to develop a machine learning model to detect fraudulent credit card transactions. By analysing transaction data, the model aims to accurately classify transactions as either fraudulent or non-fraudulent.

Problem Description:

Financial fraud is a significant challenge in the modern digital economy. Fraudulent credit card transactions can lead to substantial financial losses for banks, merchants, and customers. Timely detection of such transactions is essential to minimize financial risks and protect consumers.

Challenges:

1. Class Imbalance: The dataset is highly imbalanced, with a significantly lower number of fraudulent transactions compared to non-fraudulent ones.
2. Feature Understanding: Most of the features are anonymized using Principal Component Analysis (PCA), making interpretability difficult.
3. Real-Time Detection: Fraud detection requires models that can provide predictions in real-time to prevent further financial losses.
4. False Positives and False Negatives: Minimizing false positives (flagging legitimate transactions as fraud) and false negatives (failing to detect fraud) is crucial.

Solution Approach:

1. Data Exploration and Visualization: Perform exploratory data analysis (EDA) to understand the distribution and characteristics of transactions.
2. Data Preprocessing: Apply feature scaling and handle missing data if necessary. Address class imbalance using resampling techniques like SMOTE.
3. Model Selection: Implement and evaluate classification models, including a Random Forest classifier.
4. Evaluation: Assess model performance using appropriate metrics such as Accuracy, Precision, Recall, and F1-Score.
5. Visualization: Visualize results to gain insights into fraudulent transaction patterns.