Credit Card Fraud Detection using Machine Learning

# 1. Problem Statement

Fraudulent credit card transactions are a growing concern for financial institutions and customers. This project aims to build a machine learning-based solution that detects fraudulent transactions using a publicly available dataset. The objective is to analyze patterns in the data and develop a model that can predict fraudulent activity with high accuracy, helping prevent financial losses and improving security.

# 2. Statistical Methods Used

We used statistical methods such as correlation analysis and distribution plots to understand the data. The correlation matrix helped identify relationships between features and the target class, while histograms provided insight into the distribution of key numerical features like transaction Amount and Time.

# 3. Data Preparation Techniques

Data preprocessing included checking for missing values, scaling the 'Amount' and 'Time' features using StandardScaler, and handling class imbalance with SMOTE. We also created a new 'Hour' feature from 'Time' to analyze temporal patterns.

# 4. Predictive Model(s) Used

We implemented a Random Forest Classifier due to its ability to handle imbalanced datasets and its robustness. The model was trained on SMOTE-balanced data. We used scikit-learn for modeling and imbalanced-learn for resampling.

# 5. Evaluation Metrics

Evaluation was based on precision, recall, F1-score, and ROC-AUC score. These metrics were selected to address the imbalanced nature of the dataset, where accuracy alone would be misleading.

**7. Conclusion & Insights**

The model achieved a high ROC-AUC score, indicating strong predictive performance in distinguishing fraudulent transactions from legitimate ones. The use of SMOTE helped balance the dataset and improve recall. Future improvements could involve testing other ensemble models like XGBoost or incorporating real-time data streams for detection.