Credit Card Fraud Detection Report

# Introduction

This report outlines the process of detecting credit card fraud using machine learning methods. We utilize a dataset containing transaction records labeled as either "Normal" or "Fraud," and apply various machine learning algorithms to identify fraudulent activities efficiently.

# Data Overview

The dataset is loaded using pandas and examined for any missing values. We also explore the distribution of transactions to get a better understanding of the data. The dataset includes features derived from credit card transactions, such as transaction amount and time.

# Data Preprocessing

- Missing values are identified and handled.  
- A bar chart is generated to compare the number of fraudulent versus normal transactions.  
- Summary statistics for both fraudulent and normal transactions are calculated to understand the data better.

# Data Visualization

- A histogram is created to show how transaction amounts are distributed across fraud and normal transactions.  
- A scatter plot visualizes how transaction time and amount relate to one another for both types of transactions.  
- A heatmap is produced to display the correlations between various features in the dataset.

# Machine Learning Models Used

- \*\*Isolation Forest\*\*: This unsupervised learning model detects anomalies by looking at how isolated each data point is.  
- \*\*Local Outlier Factor (LOF)\*\*: This method identifies outliers based on the density of data points around them, helping pinpoint unusual transactions.  
- \*\*One-Class SVM\*\*: A support vector machine trained on only normal transactions to detect fraudulent activity by recognizing deviations from typical patterns.

# Model Training and Evaluation

- A random sample (10%) of the dataset is used for analysis.  
- Fraudulent and legitimate transactions are separated to calculate the proportion of outliers.  
- Relevant features are selected, and the independent (X) and dependent (Y) variables are prepared for training the models.  
- The models' performance is assessed based on accuracy and other evaluation metrics.

# Conclusion

This analysis demonstrates how machine learning techniques can be applied effectively to detect fraudulent credit card transactions. By using models such as Isolation Forest, LOF, and One-Class SVM, we can identify patterns indicative of fraud. To further enhance the models' accuracy, future work could involve optimizing hyperparameters and incorporating additional features for a more comprehensive analysis.