***Fraud Detection in Credit Card Transactions***

***Overview of the Project***

**Introduction**

In the dynamic landscape of digital finance, the increase in fraudulent credit card transactions poses a significant challenge for Mastercard, a leading global credit card provider in Europe. The Fraud Detection Department, led by Chief Risk Officer Sam Miller, faces the critical task of minimizing fraud losses while maintaining customer trust and operational efficiency.

**Challenges**

The existing rule-based system used by the department results in high false positives, leading to customer dissatisfaction and increased service calls. Additionally, sophisticated frauds often evade detection, causing false negatives and substantial financial losses.

**Proposed Solution**

To address these challenges, the department aims to leverage machine learning (ML) models capable of learning from historical transaction data. The objective is to develop a predictive model that enhances fraud detection accuracy compared to the current rule-based system.

Brief Conclusion

By implementing a robust ML solution, Mastercard anticipates enhancing customer experience, strengthening security measures, reducing operational costs, and upholding its reputation as a secure credit card provider.

**Problem Statement**

**Objectives**

- Reduce False Positives: Minimize legitimate transactions mistakenly flagged as fraudulent.

- Reduce False Negatives: Increase the detection rate of actual fraudulent transactions.

***Business and Data Understanding***

**Data Source**

The dataset comprises credit card transactions from European cardholders over two days in September 2013, obtained from Kaggle. It contains 284,807 transactions, with only 492 being fraudulent, indicating a severe class imbalance.

**Data Description**

- ***Features:*** Numerical PCA-transformed variables (V1 through V28), 'Time', and 'Amount'.

- ***Target Variable***: 'Class' indicating fraud (1) or non-fraud (0).

**Data Analysis**

Initial insights revealed the rarity of fraudulent transactions, the uniform distribution of 'Time', and a wide range of transaction amounts.

**Modeling**

The department considered various ML models, namely:

- Logistic Regression

- Random Forest

- Gradient Boosting

**Evaluation**

**Logistic Regression Results**

- ***Accuracy***: 97.32%

- ***Precision:*** 5.23%

- ***Recall:*** 92.65%

- ***F1 Score***: 9.91%

- ***AUPRC***: 79.5%

**Random Forest Results**

- Accuracy: 99.95%

- Precision: 83.22%

- Recall: 87.50%

- F1 Score: 85.30%

-AUPRC: 88.35%

**Gradient Boosting Results**

- Accuracy: 99.35%

- Precision: 18.56%

- Recall: 91.18%

- F1 Score: 30.85%

- AUPRC: 78.14%

**Recommendations**

- Deploy Logistic Regression as an initial filter followed by Random Forest for better precision and recall.

- Explore unsupervised learning and deep learning techniques for improved detection.

- Continuous model monitoring and updates are crucial for adapting to evolving fraud patterns.

**Evaluation**

The Logistic Regression and Random Forest models were selected for deployment due to their superior performance.

**Deployment**

**Pickling Models**

- Logistic Regression Model: Saved as 'logistic\_regression\_model.pkl'.

- Random Forest Model: Saved as 'random\_forest\_model.pkl'.