**Credit Card Fraud Detection Project**

**Table of Contents**

1. **Introduction**
2. **Dataset Information**
3. **Objectives**
4. **Methodology**
   * Data Preprocessing
   * Exploratory Data Analysis (EDA)
   * Model Selection
   * Model Training
   * Model Evaluation
5. **Results**
6. **Conclusion**
7. **References**

**1. Introduction**

In recent years, credit card fraud has become a significant issue for financial institutions and consumers alike. The ability to detect fraudulent transactions in real-time is crucial for minimizing losses and maintaining customer trust. This project aims to develop a machine learning model capable of identifying fraudulent transactions using a dataset containing credit card transactions.

**2. Dataset Information**

The dataset utilized for this project consists of transactions made by credit cards in September 2013 by European cardholders. It includes:

* **Total Transactions**: 284,807
* **Fraudulent Transactions**: 492 (0.172% of total transactions)
* **Features**:
  + Numerical input variables resulting from PCA transformation (V1, V2, …, V28)
  + 'Time': Seconds elapsed between each transaction and the first transaction
  + 'Amount': Transaction amount
  + 'Class': Response variable (1 for fraud, 0 for non-fraud)

Given the highly imbalanced nature of the dataset, accuracy is not a suitable metric; instead, the Area Under the Precision-Recall Curve (AUPRC) will be used for evaluation.

**3. Objectives**

* To preprocess the dataset for analysis.
* To perform exploratory data analysis to understand transaction patterns.
* To implement machine learning algorithms for fraud detection.
* To evaluate model performance using appropriate metrics.

**4. Methodology**

**Data Preprocessing**

* Import necessary libraries (e.g., pandas, numpy , seaborn, matplotlib).
* Load the dataset and check for missing values.
* Normalize or scale features if necessary.
* Split the dataset into training and testing sets.

**Exploratory Data Analysis (EDA)**

* Analyze the distribution of the target variable (Class).
* Visualize the relationship between features and the target variable.
* Identify any patterns or anomalies in the data.

**Model Selection**

* Choose suitable machine learning algorithms (e.g., Logistic Regression, Random Forest, Gradient Boosting).
* Consider techniques for handling class imbalance (e.g., SMOTE, undersampling ).

**Model Training**

* Train the selected models using the training dataset.
* Use cross-validation to ensure the model's robustness.

**Model Evaluation**

* Evaluate model performance using metrics such as AUPRC, confusion matrix, precision, recall, and F1-score.
* Compare the performance of different models and select the best one.

**5. Results**

* Present the results of the model evaluations.
* Include visualizations such as ROC curves and confusion matrices.
* Discuss the findings and how the model can be utilized in a real-world scenario.

**6. Conclusion**

Summarize the project, highlighting the importance of credit card fraud detection and the effectiveness of the developed model. Discuss potential future work, such as improving model accuracy or applying the model to other datasets.

**7. References**

* Kaggle dataset link: https://www.youtube.com/redirect?event=video\_description&redir\_token=QUFFLUhqa1BSSVMyaWlHYTBqek40Zm9lcHU3cDRkMTVOZ3xBQ3Jtc0ttTmN0ZnlXenhndUpETEJTdk9LdkZpekdGTHVoZmRpd1ZwRHFvSHlya3pHM3kyX2pUNm9US2VIOU9Vc2d5TDFTdkg1TmJzSl9Ja3pHdjh3Um1SMTFxazU1S0l5a0Jzb2dBODlSWmtZZHoxbGxWRFhNUQ&q=https%3A%2F%2Fwww.kaggle.com%2Fmlg-ulb%2Fcreditcardfraud&v=NCgjcHLFNDg