Credit card fraud detection

Amit Magar 200567979

Aryan Radadiya 200570960

Biplove Jaisi 200572872

Dinesh Saud 200570137

**Introduction**

As new generation comes to the market more and more individuals start using credit cards. As more individuals start using credit cards the amount of transactions also increases, and the likely hood of fraud transactions also increases. Credit card fraud detection is an important part in the financial market to protect the privacy of customers. Fraudulent activities can lead to series trust issues from the customer, monetary loss and damaging the reputation of financial institutions. The traditional methods of fraud handling are becoming useless against the evolving methods of committing frauds. This project aims to develop a code sufficient to predict the fraud transactions and assess the model through relevant metrics.

**Scope and Objectives**

* Developing a machine learning model to successfully differentiate between fraud and valid transactions.
* Trying to optimize the model to quickly analyze real-time transactions.
* Trying to reduce the number of valid transactions marked as fraud to enhance user experience.
* Trying to ensure user privacy throughout the process.

**Methodology**

**Data Collection and Preparation:**

* We tried to collect real time credit card data from Kaggle.
* Data link: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>
* Then we imported the data into our Jupiter notebook for further analysis.
* Then we performed data cleaning and preprocessing to make sure there are no missing values or duplicates.

**Exploratory Data Analysis (EDA):**

* Then we performed EDA process on the data by summarizing the cleaned data.
* As well as we visualize some encoded columns and created correlations heat map.

**Model Selection and Building:**

* After the EDA process we will split the data into two sets train and test for evaluation.
* We will use Logistic regression and Random Forest classification for the model.

**Hyperparameter Tuning:**

* We will be using either GridsearchCV or RandomizedSearchCV for tuning the model.
* As well as we will record the performance to compare the performance.

**Model Evaluation and Validation:**

* We will evaluate the model accuracy and performance through relevant metrics.

**Model Deployment:**

* We will set up an REST API to deploy the model, using either TesorFlow or Flask.

**Conclusion**