**Credit Card Fraud Detection Design and Innovation**

|  |  |
| --- | --- |
| **Date** | **31-10-2023** |
| **Team ID** | **3890** |
| **Project Name** | **Credit Card Fraud Detection** |

**Table of Contents**

|  |  |
| --- | --- |
| 1 | Introduction |
| 2 | Problem Statement |
| 3 | Literature Survey |
| 4 | Design Thinking |
| 5 | Phase Development |
| 5.1 | Phase 1 |
| 5.2 | Phase 2 |
| 5.3 | Phase 3 |
| 5.4 | Phase 4 |
| 5.5 | Phase 5 |
| 6 | Conclusion |

**1. Introduction**

The project aims to develop a machine learning-based system that analyses transaction data in real-time, effectively detecting credit card fraud while minimizing false positives. This solution will help financial institutions protect against fraudulent transactions, reducing financial losses and ensuring customer trust.

**2. Problem Statement**

Develop an innovative and highly accurate credit card fraud detection system that addresses the evolving nature of fraud in the digital age. The primary objective is to minimize financial losses for both cardholders and financial institutions while maintaining a seamless user experience.

**3. Literature Survey**

**Detection of Credit Card Detection using Machine Learning:**

The results of a study that compares the performance of different machine learning algorithm and neural networks in detecting credit card fraud. The study evaluates the accuracy of various techniques such as KNN, Naïve Bayes, Logistic Regression. The study also finds that MLP performs better than CFLANN in terms of accuracy and time required to converge to a solution.

**Review of Machine Learning Approach:**

The paper provides a comparative analysis of various techniques such as Logistic Regression, K-Nearest Neighbour, Naïve Bayes, Decision Trees, and Neural Network Algorithms. The study aims to identify the most effective technique for detecting fraudulent transactions and to help financial organizations protect their customers from credit card theft.

**A Novel Approach for Detecting Credit Card Fraud Detection using Machine Learning**:

Decision tree and random forest algorithms are machine learning techniques used in detecting credit card fraud. The decision tree algorithm is a predictive modelling approach that groups uncommon events in a business from an accredited customer. It enforces the consideration of all the probable outcomes of a decision and creates a comprehensive analysis of the consequences. On the other hand, the random forest algorithm is an improved version of the decision tree algorithm that uses a combination of decision trees to give better results. Every single decision tree draft for the diverse condition and works on arbitrary data sets and on the decision trees. Every tree gives the possibility of the scam business and non-scam as well.

**Unsupervised Machine Learning Techniques for Behavorial-Based Credit Card Users Segmentation:**

This process includes data pre-processing, modelling, and generating user profiles as outputs. The implementation is illustrated through a concrete case using credit card usage data from the Commercial International Bank of Egypt. The PDF file also recommends updating the segmentation model on a quarterly basis to capture new behaviours from customers and suggests further investigations using monthly, bi-annual, and annual data to understand the similarities and dissimilarities of card customers' behaviours across Africa.

**4. Design Thinking Process:**

**Data Collection:**

The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable, and it takes value 1 in case of fraud and 0 otherwise. **Dataset Link:**[**https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud**](https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud)

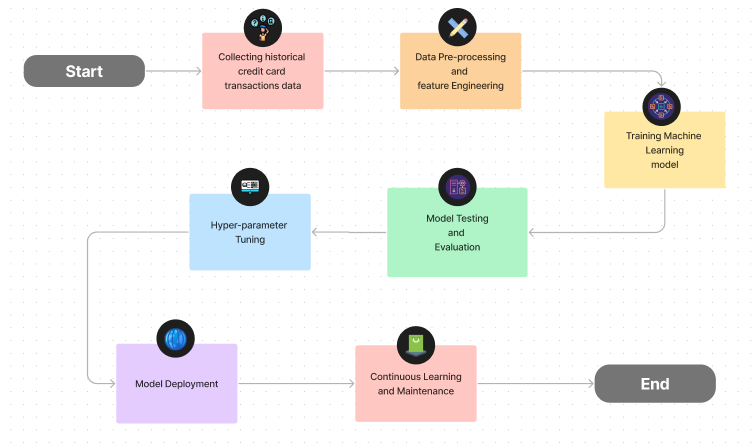
**Model Selection and Training:**

**Innovation**: Ensemble Methods and Classification Methods.

The most powerful model for classification is Logistic Regression and Decision Tree to classify the results, improve the model by changing the various feature selection. Developed the model using ensemble techniques includes Random Forest Classifier and Gradient Boosting Variants etc.

**Continuous Learning:**

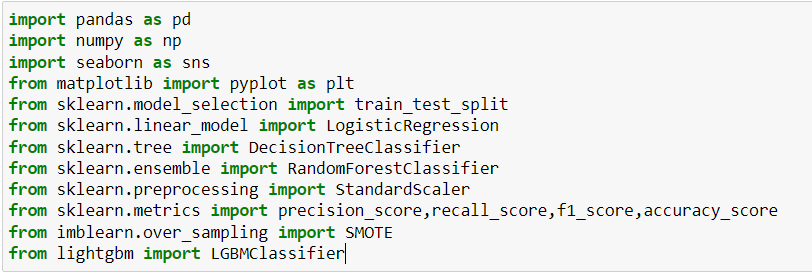
After deploying the model into a production environment, continuous monitoring is essential. It involves real-time tracking of model performance, setting up alerts to detect potential issues or drift in data patterns and regular updates.



**5.Phases of Development:**

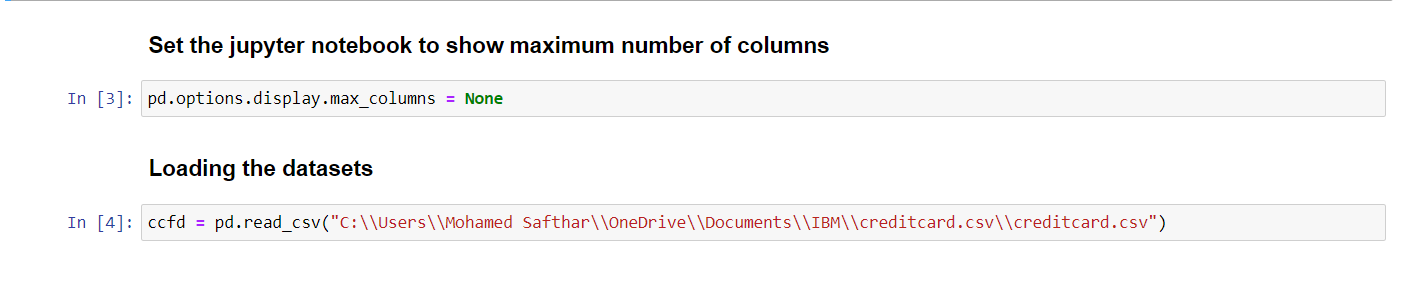
**5.1 Phase 1: Importing Dependencies**

This phase involves importing necessary python libraries and modules. These libraries are required for data processing, visualization and various machine learning tasks.



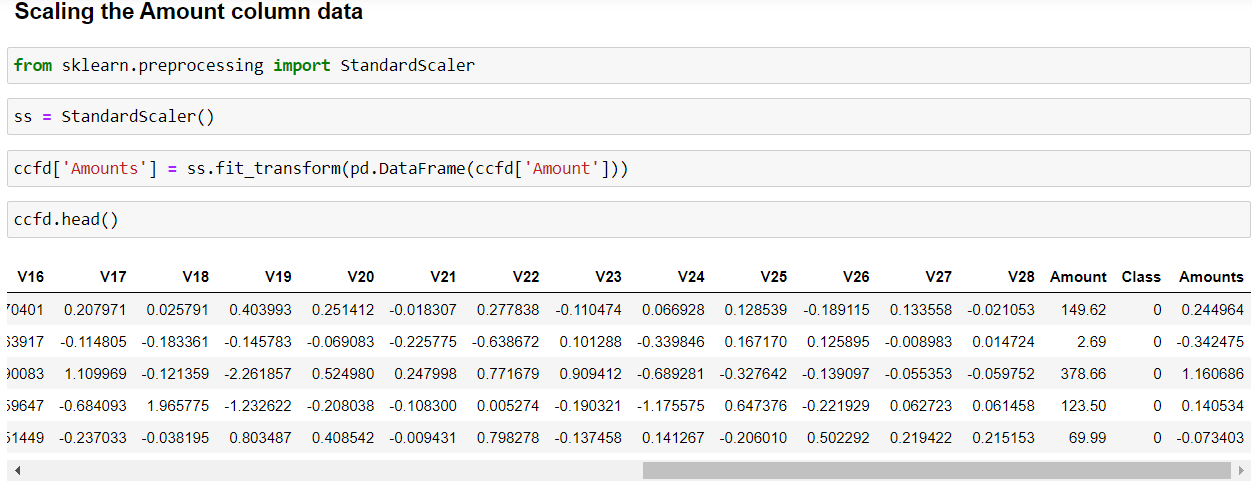
**5.2 Phase 2: Reading and Viewing Data**

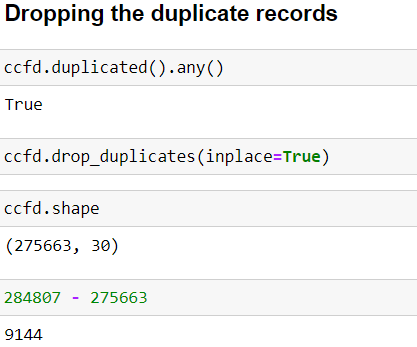
In this phase, we are importing the whole dataset into the Jupyter Notebook.

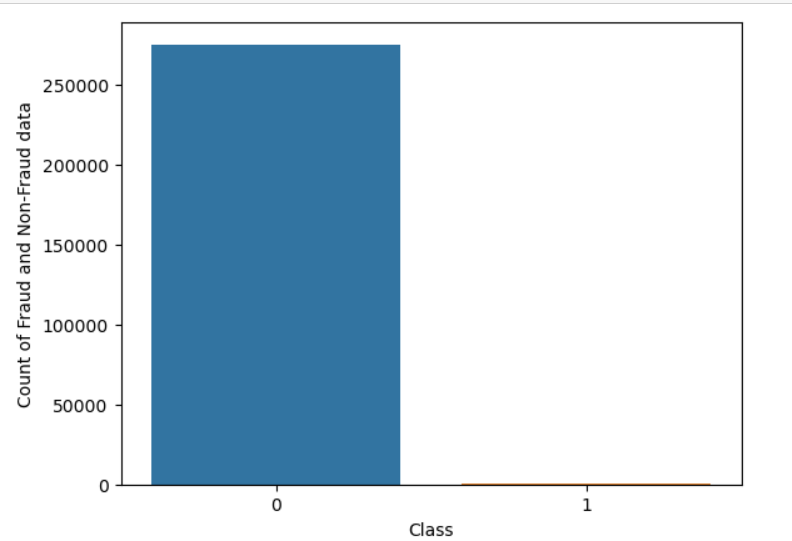


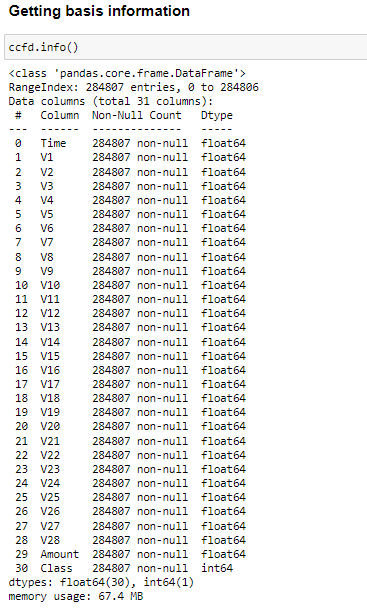
**5.3 Phase 3: Data Preprocessing and Exploration**

Here data preprocessing and exploration occur, including column selection, column transformation, creating a new feature called Amounts. Since, in given dataset there is no outliers etc.

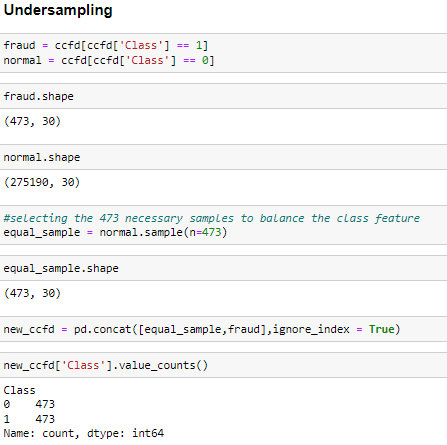




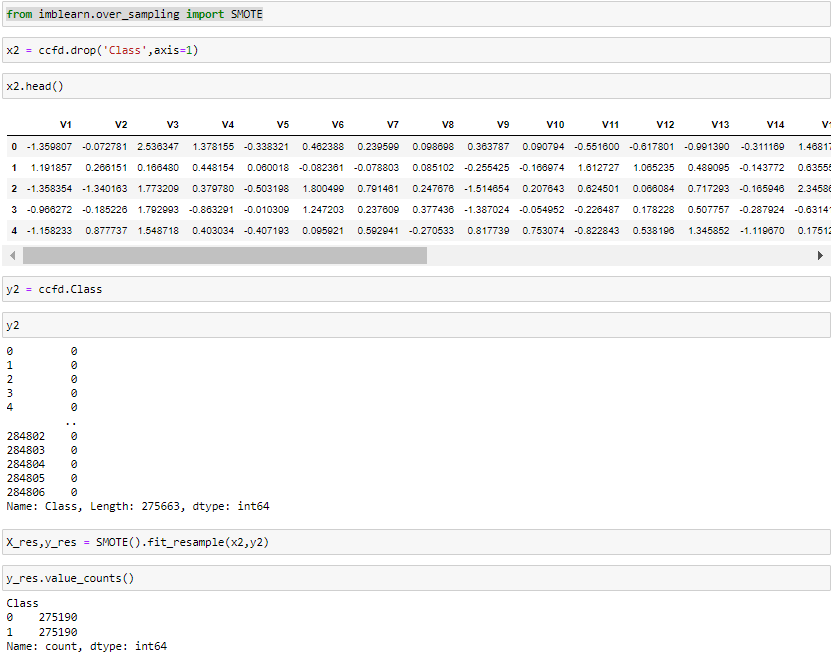


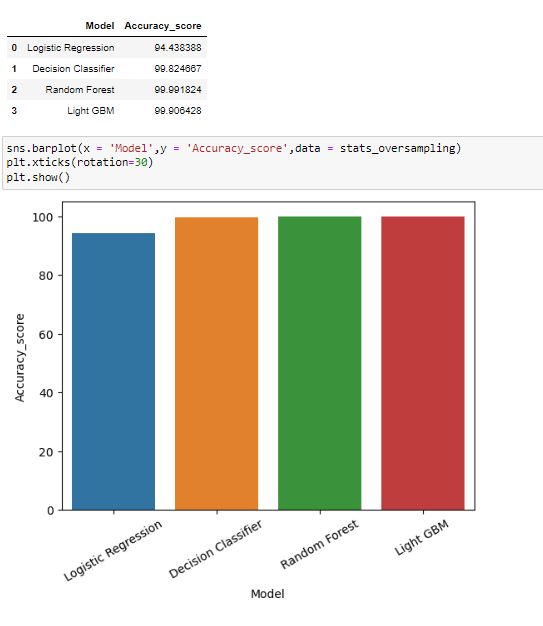


**5.4 Phase 4: Model Training**

This phase includes splitting the dataset into training and testing sets and preparing the features and target values for machine learning models. Apply under sampling and oversampling technique in this phase. 

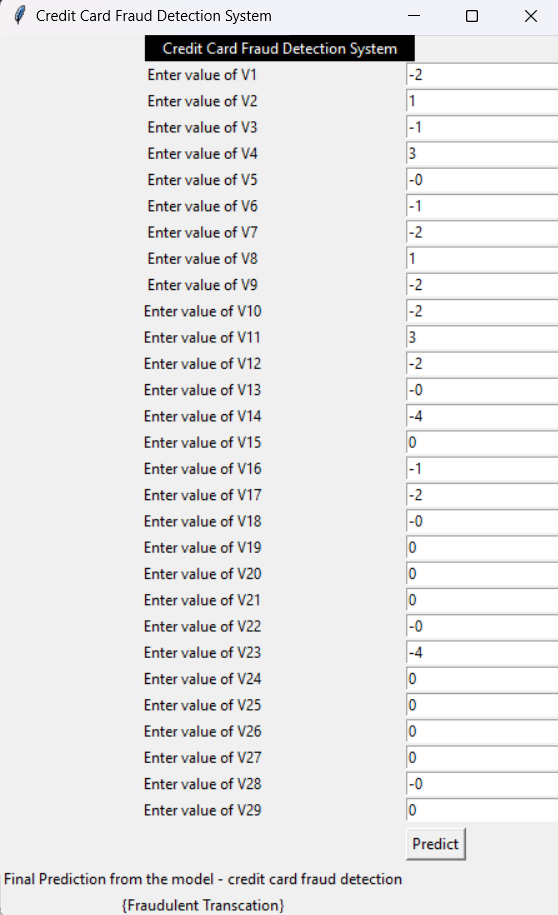
**Oversampling Technique:**

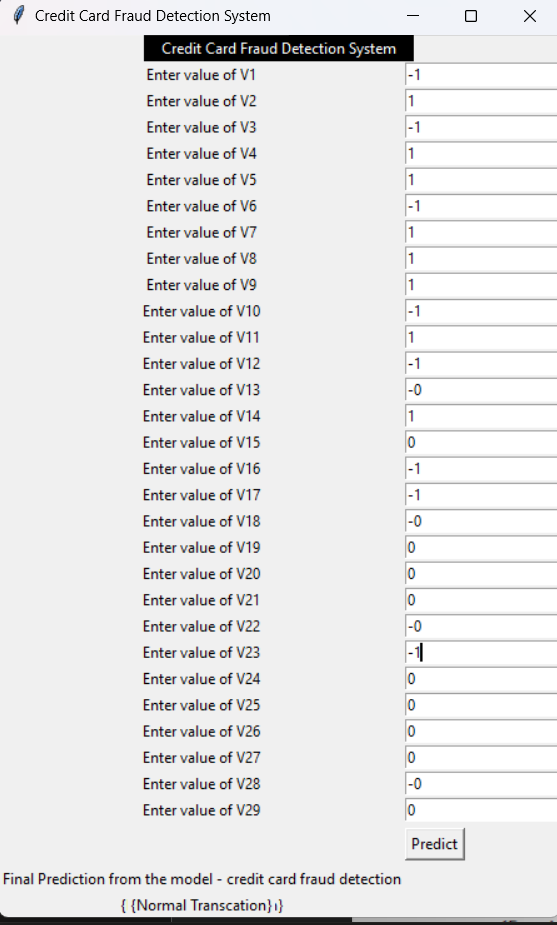




**5.5 Phase 5: Model Deployment**

For better user experience, the model has been deployed on desktop application, this app is created by tkinter module in python.





**6. Conclusion:**

In conclusion, our credit card fraud detection project demonstrates the power of machine learning in safeguarding financial transactions. By harnessing advanced data pre-processing techniques and a diverse range of classification algorithms, we've developed a robust model capable of identifying fraudulent activities. Our innovative approach, which includes ensemble methods and Light GBM, enhances the accuracy and reliability of our predictions. This project not only protects consumers and financial institutions from fraudulent activities but also sets a strong foundation for further advancements in the field of fraud detection and prevention through machine learning.