Credit Card Fraud Detection Report

Student: Devansh Patel (C0928483)

Course: AML-3104 Neural Networks and Deep Learning

Submitted to: Ishant Gupta

1. Introduction

This project's objective is to use the Credit Card Fraud Detection dataset from Kaggle to develop a machine learning model that can identify fraudulent credit card transactions. Due to the small proportion of fraudulent transactions compared to legitimate ones, the dataset is extremely unbalanced. Data preprocessing, class imbalance management, logistic regression model training, and performance evaluation are all part of the project.

---

2. Overview of the Dataset

Credit card transactions performed by European cardholders in September 2013 are included in the dataset. Just 492 of the 284,807 transactions in it are fake, accounting for 0.172% of the total. There are 31 features in the dataset:

- **Time:** The amount of time that passed between the current transaction and the initial one.
- **V1-V28:** Principal components derived from PCA (anonymised characteristics) are V1–V28.
- **Amount:** The sum of the transaction.
- **Class:** Target variable (0 for valid transactions, 1 for fraud).

---

3. Preprocessing Data

3.1 Taking Care of Missing Values

When the dataset was examined for missing values, none were discovered.

3.2 Eliminating Replicas

To guarantee data quality, 1081 duplicate entries were eliminated from the dataset. The size of the dataset was lowered to 283,726 rows after duplicates were eliminated.

3.3 Distribution of Classes

The dataset has a significant imbalance.

Transactions that are legitimate: **283,253**
Transactions involving fraud: **473**

3.4 Dealing with Unbalanced Classes

The SMOTE (Synthetic Minority Oversampling Technique) was used to rectify the class imbalance. To balance the dataset, this method creates artificial samples of the minority class (fraudulent transactions).

The dataset was balanced with **473 valid** and **473 fraudulent** transactions following the application of **SMOTE.**

---

4. Training Models

4.1 Data Splitting

Training and testing sets were created from the dataset:

80% of the data (756 samples) is the training set.

20% of the data (190 samples) is the testing set.

4.2 Logistic Regression Model

This categorisation task was assigned to a Logistic Regression model.

The balanced training dataset was used to train the model.

---

5. Model Evaluation

5.1 Accuracy of Training

On the training data, the model's accuracy was **93.39%.**

5.2 Accurate Testing

On the test data, the model's accuracy was **92.63%.**

---

6. Model Deployment

6.1 Locally Saving the Model

The joblib package was used to save the trained logistic regression model locally. As a result, the model may be easily included into applications and reused without requiring retraining.

6.2 Building a Front-End Application with Streamlit

A Streamlit web application was created to enable end users to access the model. A strong framework for creating dynamic web apps with little coding is called Streamlit.

Users can enter transaction information (such as V1–V28, amount, etc.) into the application, which then determines if the transaction is fraudulent or not.

Key Features of the Streamlit App:

- **User Input Form:** A form allows users to enter transaction information.
- **Prediction:** To determine if the transaction is fake, the software makes use of the saved logistic regression model.
- **Result Display:** The app shows the outcome of the prediction (such as "Legitimate" or "Fraudulent").

---

7. Conclusion

- With an accuracy of 92.63% on the test dataset, the logistic regression model demonstrated strong performance in identifying fraudulent transactions.
- The class imbalance problem was successfully resolved by using SMOTE, which made it possible for the model to learn equally from both classes.
- Experiments with additional machine learning algorithms (such Random Forest and XGBoost) and hyperparameter adjustment to boost performance could be part of future developments.

Dataset Link: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud