Credit Card Fraud Detection Project

**Introduction**

Credit card fraud is a significant concern for financial institutions, leading to substantial financial losses annually. This project aims to develop a machine learning model to detect fraudulent transactions accurately. By leveraging logistic regression, a widely used classification algorithm, the model identifies patterns that distinguish legitimate transactions from fraudulent ones. The dataset used includes transactions labeled as either legitimate (Class 0) or fraudulent (Class 1).

**Analysis**

**Data Overview**

The dataset contains various features representing transaction details. The 'Class' feature is the target variable, with '0' indicating legitimate transactions and '1' indicating fraudulent ones. Key observations about the data include:

The dataset is highly imbalanced, with a significantly larger number of legitimate transactions compared to fraudulent ones.

Legitimate transactions: 284,315

Fraudulent transactions: 492

**Data Preprocessing**

To address the class imbalance, we employed an undersampling technique, where we randomly sampled 492 legitimate transactions to match the number of fraudulent transactions. This balanced the dataset, allowing the model to train more effectively on both classes.

**Model Development**

We utilized logistic regression due to its simplicity and effectiveness in binary classification tasks. The model was trained on 80% of the dataset, with the remaining 20% reserved for testing. Key steps included:

Splitting the data into training and testing sets using an 80/20 split, with stratification to maintain class balance.

Training the logistic regression model on the training data.

Evaluating model performance on both training and testing datasets.

**Results**

The logistic regression model achieved the following accuracy scores:

Training Data Accuracy: 93.91%

Testing Data Accuracy: 92.27%

These results demonstrate that the model effectively identifies fraudulent transactions with a high degree of accuracy, making it a valuable tool for real time fraud detection.

**Conclusion**

This project successfully implemented a logistic regression model to detect credit card fraud. Despite the initial class imbalance, the undersampling technique allowed the model to learn patterns in both legitimate and fraudulent transactions effectively. The model's high accuracy on both training and testing data suggests it is wellsuited for deployment in realworld scenarios, where early detection of fraudulent transactions is crucial for minimizing financial losses.

**Future Work**

To further improve the model's performance, future work could include exploring more sophisticated techniques, such as ensemble methods or deep learning, and testing the model on larger, more diverse datasets. Additionally, implementing realtime monitoring and alert systems could enhance the model's practical application in detecting fraudulent transactions as they occur.