

Meti Kebede

DATA 3421-001

05-01-2023

### Credit card Fraud Detection

The detection of credit card fraud is a crucial issue in the financial industry, and it is essential to prevent financial losses by detecting fraudulent transactions. This report outlines a credit card fraud detection project that utilizes machine learning algorithms in Python.

The dataset used in this project comprises credit card transactions that took place over a two-day period in September 2013. The dataset contains a total of 284,807 transactions, of which only 492 are fraudulent, making it an imbalanced dataset. To address this issue, I implemented a down sampling technique, where we randomly selected 1040 normal and 393 fraudulent transactions from the majority class to ensure a balanced dataset.

Visualizing the data was a challenging task due to the imbalanced nature of the dataset. However, I overcame this issue by using various types of histograms and scatter plots to effectively visualize the data and gain insights into the features' correlation and identify any patterns.

Seven machine learning algorithms were utilized to develop the model: Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier,

AdaBoost Classifier, KNeighbors Classifier, and Support Vector Machine. Each model was trained ten times, and the average of the results was taken to ensure the model's robustness and reliability.

The results demonstrated that all models performed well in detecting credit card fraud, with Random Forest outperforming the others. It achieved an accuracy of 0.9996, F1-score of 0.8605, and recall of 0.7714. In contrast, Gradient boost was the least performing model, with an accuracy of 0.9089, F1-score of 0.6629, and recall of 0.6020. These results indicate that machine learning algorithms are effective in detecting fraudulent transactions in credit card transactions.

The primary challenge faced in this project was dealing with the imbalanced dataset. Down sampling was used to address this issue, which may result in the loss of data. Another challenge was visualization, which was addressed by using different types of histograms and scatter plots to visualize the data effectively.

In conclusion, this report outlines the development of a credit card fraud detection model using Python. The results demonstrate the effectiveness of machine learning algorithms in detecting fraudulent transactions in credit card transactions. However, there is still room for further research to explore other techniques that can improve the model's performance, such as using complex deep learning model. And also using different dataset that is balanced could help the project even more. Overall, this project highlights the importance of developing robust and reliable machine learning models to detect credit card fraud and prevent financial losses.