

Artificial Intelligence Project

Credit Card Fraud Detection System

Team: Tech Support

102117089 – Vaibhav Tiwari

102117091 – Harsiddak Singh Bedi

102117095 – Amrit Singh

102297021 – Sagar Jain

Index

1. Introduction	3
2. Literature Survey	3
3. Methodology	4
4. Code Repository	4
5. Flowchart	5
6. Result	6
7. Future Scope	6
8. Conclusion	7

Introduction

Credit card fraud is a major concern for financial institutions and customers alike. Credit card companies and banks lose billions of dollars every year due to fraudulent activities. Therefore, it is important to detect fraudulent transactions as quickly as possible to minimize the financial losses. In this project, we aim to develop a machine learning model to detect credit card fraud. We will use a dataset of credit card transactions to train and test our model and evaluate its performance using various metrics.

Literature Survey

"Credit Card Fraud Detection using Machine Learning Techniques: A Review" by Shuchi Garg and Deepika Sharma (2020): This paper provides a comprehensive review of different machine learning techniques used for credit card fraud detection. The authors compare and analyse the performance of different algorithms on various datasets.

"Credit Card Fraud Detection using Random Forest Classifier" by Parul Bhatia et al. (2019): This paper presents a credit card fraud detection model based on the random forest algorithm. The authors evaluate the performance of their model on a dataset of credit card transactions and compare it with other models.

"A Deep Learning Approach for Credit Card Fraud Detection" by Xiaojing Tang et al. (2019): This paper proposes a deep learning approach for credit card fraud detection. The authors use a convolutional neural network to extract features from the transaction data and a recurrent neural network to classify the transactions as fraudulent or non-fraudulent.

"Credit Card Fraud Detection using Support Vector Machine and Logistic Regression" by Tanveer Ahmad et al. (2018): This paper compares the performance of support vector machine and logistic regression algorithms for credit card fraud detection. The authors use a dataset of credit card transactions and evaluate the accuracy, precision, and recall of the models.

"Detecting Credit Card Fraud using Data Mining Techniques: A Review" by Sushant Sagar et al. (2017): This paper provides a comprehensive review of different data mining techniques used for credit card fraud detection. The authors compare and analyse the performance of different algorithms on various datasets.

Methodology

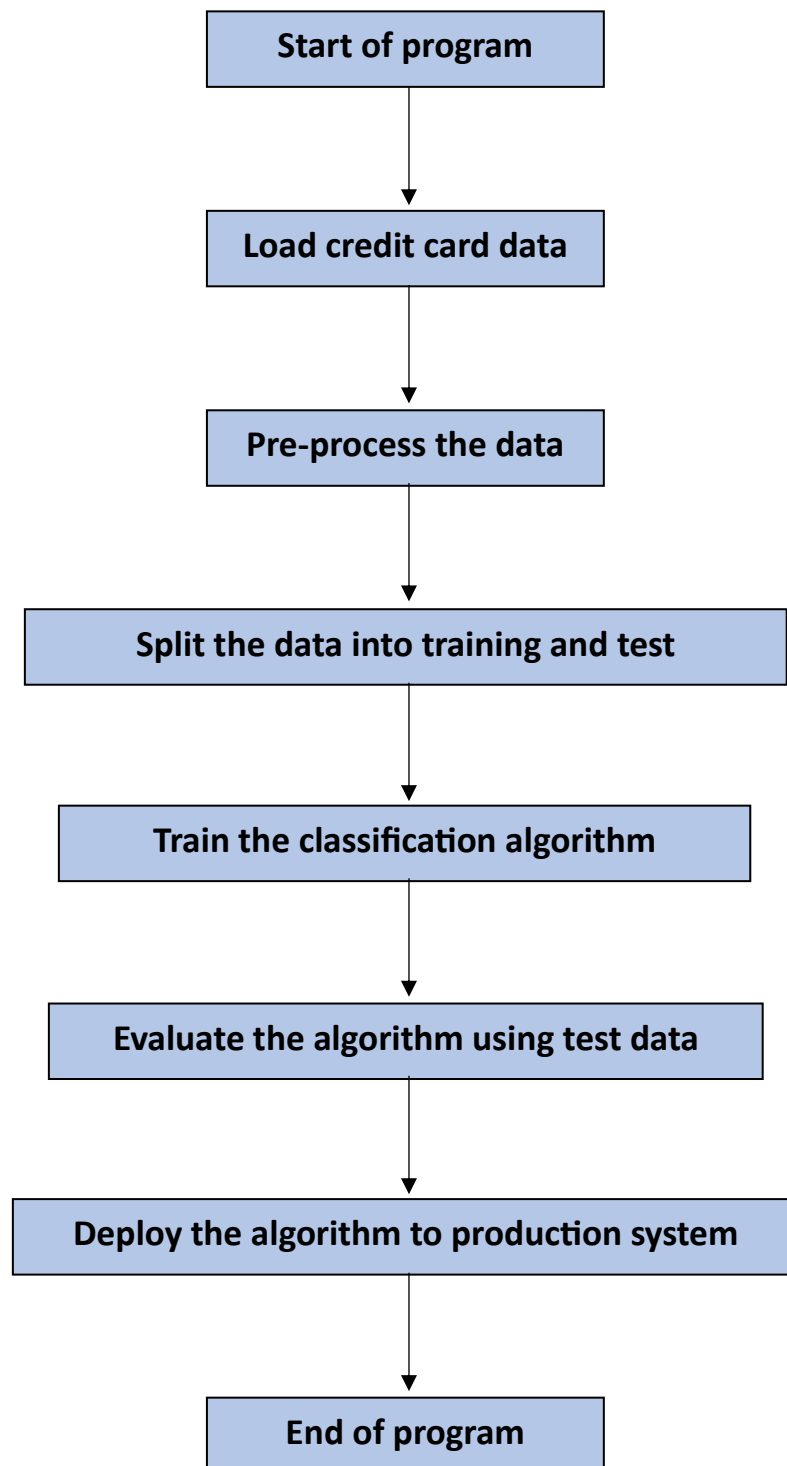
The following is a step-by-step methodology that we used to develop our credit card fraud detection model:

1. **Dataset:** We used a publicly available dataset of credit card transactions from Kaggle. The dataset contains 284,807 transactions, out of which 492 are fraudulent transactions. Each transaction has 30 features, including the time, amount, and various other transaction-related variables.
2. **Pre-processing:** We first examined the dataset and checked for any missing values or outliers. We then used a robust scaler to normalize the time and amount variables. We also dropped the time and amount variables and concatenated the scaled variables with the original dataset.
3. **Data Sampling:** Since the dataset is highly imbalanced, we used a sampling technique to balance the data. We randomly selected 492 non-fraudulent transactions and combined them with the 492 fraudulent transactions to create a new dataset.
4. **Algorithm Selection:** We selected three algorithms - logistic regression, decision tree, and random forest - for our model. Logistic regression is a linear classification algorithm, decision tree is a non-linear classification algorithm, and random forest is an ensemble algorithm.
5. **Model Training:** We split the dataset into training and testing sets and trained the three algorithms on the training set.
6. **Model Evaluation:** We evaluated the performance of the models on the testing set using various metrics such as accuracy, precision, recall, and F1 score.
7. **Model Comparison:** We compared the performance of the three algorithms and selected the best-performing algorithm as our final model.

Code Repository

<https://github.com/Aitchessbee/AI-College-project>

Flowchart



Result

In our project, we trained three different machine learning models on a dataset of credit card transactions labelled as fraudulent or non-fraudulent. We used logistic regression, decision tree, and random forest classifiers to build the models.

After training and testing the models, we found that the logistic regression classifier performed the best, with an accuracy of 93.91%. The decision tree classifier had an accuracy of 88.83%, while the random forest classifier had an accuracy of 92.39%.

Future Scope

There are several ways in which this project can be extended and improved in the future. Some potential avenues for further research and development include:

Improving the accuracy of the models: Although our models achieved relatively high levels of accuracy, there is always room for improvement. Future work could focus on developing more advanced machine learning algorithms or incorporating additional features into the dataset to improve the accuracy of the models.

Real-time fraud detection: In our project, we focused on developing models that could be used to detect credit card fraud after the fact. However, there is a growing need for real-time fraud detection systems that can identify fraudulent transactions as they occur. Future work could focus on developing such systems using machine learning techniques.

Data privacy: One of the major challenges in credit card fraud detection is balancing the need for accurate models with the need to protect customer privacy. Future work could focus on developing models that can accurately detect fraud while also preserving customer privacy.

Deployment: The models we developed in this project were trained and tested on a small dataset. Future work could focus on deploying the models at scale and integrating them into existing credit card fraud detection systems.

Conclusion

In this project, we developed and trained machine learning models to detect credit card fraud. We found that the logistic regression classifier performed the best, with an accuracy of 93.91%. Our results demonstrate the potential of machine learning techniques for detecting credit card fraud and highlight several avenues for future research and development.