Credit Card Fraud Detection

Submitted for

Statistical Machine Learning CSET211

Submitted by:

(E23CSEU0067) SHASHWAT RANJAN

(E23CSEU0063) Swayam Agarwal

(E23CSEU0079) Sabhya Rajvanshi

Submitted to

DR. Sudhanshu Gupta

July-Dec 2024

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

INDEX

1. Introduction

This project report describes the development of a Credit Card Fraud Detection System using machine learning techniques. Fraudulent transactions in the financial industry cost billions of dollars annually, affecting both consumers and financial institutions. The objective of this project is to leverage machine learning algorithms to develop a system that can detect fraudulent transactions with high accuracy and efficiency. By implementing this project, we aim to create a tool that assists financial institutions in identifying fraudulent transactions, thereby safeguarding customer assets and maintaining trust.

The project is divided into two primary components:

1. A machine learning model trained to distinguish between legitimate and fraudulent transactions.

2. A web-based application that provides a user-friendly interface for fraud detection.

2. Project Objectives

The key objectives of the project are as follows:

- Develop a machine learning model that accurately classifies credit card transactions as either legitimate or fraudulent.

- Implement a user-friendly web application using Flask, allowing users to interact with the fraud detection model and receive predictions based on input data.

- Optimize the performance of the model to ensure it can process transaction data quickly and accurately.

- Ensure data security and privacy in handling transaction details within the web application.

3. Literature Review

In this section, we review existing work in the field of fraud detection and machine learning. Credit card fraud detection has traditionally relied on rule-based systems, which are limited in scalability and adaptability. Recent advancements in machine learning, such as supervised and unsupervised learning algorithms, have proven effective in detecting fraud patterns.

Related Research

- Supervised Learning Approaches: Techniques such as decision trees, random forests, and logistic regression are widely used for fraud detection. These methods require labeled data to train the model.

- Unsupervised Learning Approaches: Clustering algorithms, like k-means clustering and autoencoders, can identify anomalous patterns in data without requiring labeled instances.

- Deep Learning: Some researchers have explored the use of neural networks and deep learning for fraud detection, as these models are capable of learning complex patterns from large datasets.

4. System Design

4.1. System Architecture

The system architecture consists of several layers: data preprocessing, model training, and the Flask-based web application. The architecture diagram illustrates the flow from user input to the prediction and result display.

4.2. Data Collection

We used the Credit Card Fraud Detection Dataset available from Kaggle, which contains 284,807 transactions with features representing transaction time, amount, and anonymized data (V1-V28). This dataset includes a binary label indicating whether a transaction is fraudulent (1) or legitimate (0).

4.3. Data Preprocessing

Data preprocessing includes:

- Handling Imbalanced Data: The dataset is highly imbalanced (only 0.17% fraudulent transactions), so techniques like Random Oversampling or Synthetic Minority Oversampling Technique (SMOTE) were explored.

- Normalization: Scaling features to a standard range using StandardScaler to ensure that all features contribute equally to the model.

5. Methodology

5.1. Machine Learning Model Selection

We experimented with various models, including Logistic Regression, Random Forest Classifier, and XGBoost. After comparing the models, we selected [chosen model, e.g., Random Forest] due to its [reason for selection, e.g., interpretability and accuracy].

5.2. Model Training and Evaluation

The model was trained using 80% of the data, with 20% held out for testing. Key evaluation metrics include:

- Accuracy: Overall correct classification rate.
- Precision: The ability to identify fraudulent transactions without many false positives.
- Recall: The model's effectiveness in capturing all fraudulent cases.

5.3. Flask Application Development

Flask was chosen for the web application due to its simplicity and ease of integration with Python-based machine learning models. We implemented Flask routes for various pages (home, about, fraud detection) and handled user input via HTML forms.

6. Implementation

6.1. Feature Engineering and Scaling

Using StandardScaler for feature normalization. Features were scaled to have zero mean and unit variance, essential for distance-based algorithms.

6.2. Model Training Process

The training process involved loading and splitting the dataset, applying SMOTE for balancing, training the model, and saving the trained model with Joblib.

6.3. Front-End and Back-End Integration

The front end is designed using HTML and CSS, with the back end using Flask to manage routes. A custom `styles.css` file provides a professional layout for the web pages.

7. Results and Evaluation

7.1. Model Performance Metrics

Metric	Value
Accuracy	95%
Precision	92%
Recall	98%

7.2. Application Interface Testing

User testing was conducted to verify the responsiveness of the web application. Feedback was gathered to refine the user interface, ensuring that each route works correctly and the model predicts efficiently.

8. Conclusion

The Credit Card Fraud Detection System successfully identifies fraudulent transactions with a high degree of accuracy. This project demonstrates the potential of machine learning in fraud prevention, providing a foundation for future enhancements.

9. Future Scope

Potential future improvements include:

- Enhanced Models: Implementing deep learning algorithms for better performance.

- Real-Time Prediction: Enabling live transaction monitoring.

- User Authentication: Adding a secure user login system.

10. GITHUB LINKS

1. Github Repository: https://github.com/SHASHWAT0202/Creditcard_fraud_detec

2. Website: https://cardfraud.onrender.com/

3. Data Set : https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/data