Credit Card Fraud Detection System

Credit Card Fraud Detection System

Internship Completion Report

Submitted By: Ayesha Siddiqa
Organization: Growthlink
Project Title: Credit Card Fraud Detection System
Duration: [Insert Duration]

Abstract

This project report presents the development of a machine learning-based Credit Card Fraud Detection System. A Random Forest Classifier was implemented to detect fraudulent transactions with a focus on maximizing model performance and interpretability.

Acknowledgment

I would like to express my sincere gratitude to Growthlink for providing me with the opportunity to work on this project. Special thanks to my mentor and the entire team for their invaluable guidance and support throughout the internship.

Table of Contents

Credit Card Fraud Detection System

Introduction

Credit card fraud is a significant issue in the financial sector, causing substantial monetary losses worldwide. Early and accurate detection of fraudulent transactions is crucial for maintaining customer trust and financial stability.

Problem Statement

Given the increasing number of credit card fraud cases, there is a critical need for an automated system capable of detecting fraudulent transactions accurately and efficiently.

Objective

The main objective is to design a machine learning model capable of accurately classifying credit card transactions as either fraudulent or non-fraudulent while minimizing false positives.

Literature Review

1. Bolton, R. J., & Hand, D. J. (2002). 'Statistical Fraud Detection: A Review.'
2. Whitrow, C., Hand, D., Juszczak, P., Weston, D., & Adams, N. (2009). 'Transaction aggregation

as a strategy for credit card fraud detection.'
3. Bahnsen, A. C., Aouada, D., Stojanovic, J., & Ottersten, B. (2016). 'Cost sensitive credit card fraud detection using Bayes minimum risk.'
4. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). 'Calibrating Probability with Undersampling for Unbalanced Classification.'

Methodology

Dataset Description:
- 10,000 simulated transactions.
Preprocessing:
- Extracted hour, day_of_week.
- Frequency encoding.
- Standardization.
Handling Imbalance:
- Used class_weight='balanced'.

Model Building

Random Forest Classifier selected for its robustness and ability to handle imbalanced data.
Parameters: n_estimators=100, random_state=42, class_weight='balanced'.

Model Training Process

Stratified train-test split. Model trained and evaluated based on multiple metrics.

Model Evaluation

Metrics: Precision, Recall, F1-Score.
Visual tools: Confusion Matrix and Precision-Recall Curve.

Credit Card Fraud Detection System

Feature Importance Analysis

Calculated feature importance scores to identify key attributes influencing fraud detection.

Misclassification and Error Analysis

Analyzed false positives and false negatives to understand model shortcomings.

Use Case Diagrams

11.1 Customer:

- Login
- View Transactions
- Receive Fraud Alerts

11.2 Admin:

- Monitor System Logs
- Review Flagged Transactions
- Retrain Models

(Diagrams attached in appendix.)

Challenges Faced

1. Handling imbalanced data.
2. Feature engineering.
3. Achieving high recall without sacrificing precision.

Future Scope

1. Applying SMOTE or ADASYN techniques.
2. Using deep learning for anomaly detection.
3. Real-time integration.

Credit Card Fraud Detection System

Conclusion

The project successfully implemented a machine learning-based fraud detection system, demonstrating promising results.

References

1. Bolton & Hand (2002)
2. Whitrow et al. (2009)
3. Bahnsen et al. (2016)
4. Dal Pozzolo et al. (2015)
5. Scikit-learn Documentation

Appendix

- Code snippets
- Evaluation graphs
- Use case diagrams