



Fraud Detection Using Machine Learning

Barkat Abdulhaq Z. Khokhar Afzal Y. Parmar Neha L. Parmar Nikita L. Prof. Janki Tejas Patel

Department of Engineering Ahmedabad - 380060

ABSTRACT :

Fraud detection has become a crucial area in financial institutions, e-commerce, and digital transactions. With the exponential increase in online transactions, detecting fraudulent activities has become challenging due to complex behavioral patterns and data imbalance. This research paper presents an analytical study on machine learning algorithms used for fraud detection, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM). The aim is to compare these algorithms based on accuracy, precision, recall, and F1-score to identify the most effective model for real-time fraud detection.

Keywords: Fraud Detection, Machine Learning, Classification, Random Forest, SVM, Data Imbalance.

Introduction

In the modern digital world, online financial transactions have significantly increased, which has led to a parallel rise in fraudulent activities. Fraud detection refers to identifying and preventing such malicious actions before they cause major financial loss. Traditional rule-based systems are inefficient in handling large-scale data and adaptive fraud patterns. Therefore, machine learning techniques have gained prominence for detecting fraudulent activities automatically based on patterns learned from data. This paper focuses on implementing and analyzing multiple machine learning algorithms for fraud detection and evaluating their performance based on key metrics.

Methodology / Experimental Setup

The experimental setup involves preprocessing the dataset, handling data imbalance, training models, and evaluating performance. The dataset used in this study represents credit card transactions, with labels indicating fraudulent and non-fraudulent records. The following algorithms were implemented:

- Logistic Regression – a simple yet effective linear model for binary classification.
- Decision Tree – a non-linear classifier that splits data based on feature information gain.
- Random Forest – an ensemble of decision trees providing robustness and higher accuracy.
- Support Vector Machine (SVM) – a powerful classifier that finds optimal hyperplanes to separate classes.

Data preprocessing included removing null values, normalizing features, and applying SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset. Each model was trained on 70% of the data and tested on 30%. Performance was measured using accuracy, precision, recall, and F1-score metrics.

Result and Discussion

The experimental results reveal that ensemble learning models outperform single classifiers. The Random Forest model achieved the highest accuracy of 96%, followed by SVM with 93%, Decision Tree with 91%, and Logistic Regression with 89%. Precision and recall values also indicated Random Forest's superior ability to identify fraudulent cases while minimizing false positives.

Algorithm | Accuracy | Precision | Recall | F1-Score

-----|-----|-----|-----|-----

Logistic Regression | 89% | 0.88 | 0.84 | 0.86

Decision Tree | 91% | 0.90 | 0.87 | 0.88

Random Forest | 96% | 0.95 | 0.94 | 0.94

SVM | 93% | 0.92 | 0.90 | 0.91

The Random Forest model's robustness is due to its ensemble nature, which reduces overfitting and increases generalization. These results highlight the potential of machine learning in reducing manual fraud analysis and improving transaction security.

Conclusion

This study demonstrates that machine learning algorithms play a crucial role in detecting fraudulent transactions effectively. Among the tested models, Random Forest achieved the highest accuracy and reliability. Machine learning approaches can significantly minimize human intervention and improve detection efficiency in real-time environments. However, continuous updates and retraining are required as fraud patterns evolve with time.

Future Scope

In the future, deep learning models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) can be explored for enhanced feature extraction and sequential data analysis. Additionally, integrating artificial intelligence with blockchain and big data analytics may further improve fraud prevention systems. Research can also focus on explainable AI (XAI) techniques to enhance transparency and interpretability in fraud detection decisions.

REFERENCES

1. J. West and M. Bhattacharya, 'Intelligent Financial Fraud Detection: A Comprehensive Review,' Computers & Security, 2021.
2. R. Randhawa et al., 'Credit Card Fraud Detection Using Machine Learning: A Comparative Analysis,' IEEE Access, 2022.
3. W. Stallings, 'Data and Computer Communications,' 11th Edition, Pearson, 2023.
4. K. Patel and S. Sharma, 'Ensemble Models for Financial Fraud Detection,' Expert Systems with Applications, 2024.
5. N. Verma, 'Machine Learning in Financial Fraud Prevention,' Journal of Applied Computing, 2025.