Credit Card Fraud Detection Using Machine Learning

Prof.Onkar M. Patil[1], Prof.Dr.Ashwini A. Patil[2] ,Swami Shruti Shivappa[3],Waghmare Supriya Sanjiv[4]
Department of Information Technology
M.S.Bidve Engineering College
Latur,India
Email-onkarmpatil@gmail.com[1], ashwinibiradar29@gmail.com[2], shrutiswami102@gmail.com[3], waghmaresupriya48@gmail.com[4]

Abstract

The rapid growth of digital payment systems has significantly increased the risk of credit card fraud, resulting in substantial financial losses for banks and customers. Traditional rule-based fraud detection systems often fail to identify complex and evolving fraudulent patterns in real time. This research aims to develop an efficient credit card fraud detection system using machine learning techniques. A publicly available Kaggle dataset containing anonymized credit card transaction records was used for experimentation. Data preprocessing involved handling missing values, feature scaling, train-test splitting, and addressing class imbalance using the Synthetic Minority Oversampling Technique (SMOTE). Several machine learning algorithms, including Logistic Regression, Decision Tree,Random Forest, and XGBoost, were implemented and evaluated using performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix. Experimental results demonstrate that ensemble models, particularly Random Forest and XGBoost, achieve higher accuracy and recall in detecting fraudulent transactions. The findings highlight the importance of handling imbalanced datasets and selecting robust algorithms for fraud detection. This study concludes that machine learning-based approaches can significantly enhance real-time credit card fraud detection, reduce financial losses, and improve trust in digital financial systems.

Keywords: Credit Card Fraud Detection, Machine Learning, Classification, Imbalanced Dataset, Financial Security.

I. Introduction

In recent years, online and cashless transactions have become an essential component of modern financial systems. Credit cards are widely used due to their convenience, speed, and global acceptance. However, the growing volume of credit card transactions has also led to an increase in fraudulent activities. Credit card fraud causes significant financial losses to both customers and financial institutions and negatively impacts customer trust. Traditional fraud detection methods are mostly rule-based and rely heavily on manually defined rules. These systems struggle to adapt to new and complex fraud patterns and are inefficient when dealing with large-scale datasets. Machine learning techniques provide automated and intelligent solutions by learning transaction behavior patterns from historical data. This paper focuses on applying various machine learning algorithms to detect fraudulent credit card transactions efficiently. The main contribution of this research is a comparative analysis of multiple machine learning models on an imbalanced on aeffect most effective approach for fraud detection.

II. Literature Review

Literature Review Several researchers have explored the application of machine learning techniques for credit card fraud detection. Dal Pozzolo et al[1]. introduced cost-sensitive learning methods to address class imbalance in fraud detection systems. Bahnsen et al[2]. proposed decision tree-based models that focus on minimizing misclassification costs. Whitrow et al[3]. emphasized transaction aggregation and feature engineering techniques to improve detection accuracy. Recent studies show that ensemble models such as Random Forest and Gradient Boosting outperform traditional classifiers. Despite these advancements, handling imbalanced datasets and achieving high recall for fraudulent transactions remain major challenges.

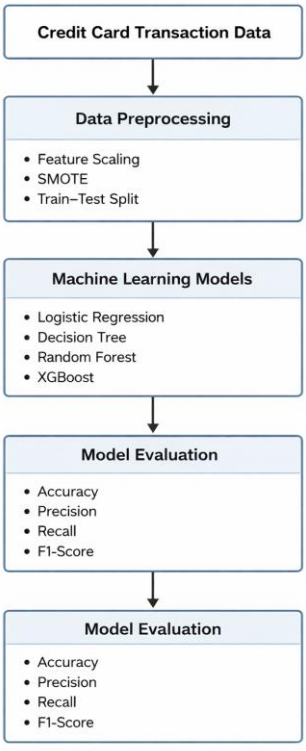

Fig. 1. Architecture of the proposed machine learning-based credit card fraud detection system.

III. Methodology

A. Dataset Description

The dataset used in this study was obtained from Kaggle and contains credit card transactions made by European cardholders. It consists of numerical features generated using Principal Component Analysis (PCA) to maintain data confidentiality. The target variable indicates whether a transaction is fraudulent or legitimate.

B. Data Preprocessing

The following preprocessing steps were applied:

- Handling missing values
- Feature scaling for numerical stability
- Splitting data into training and testing sets
- Addressing class imbalance using SMOTE

C. Machine Learning Models

The following machine learning algorithms were implemented:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- XGBoost Classifier

D. Evaluation Metrics

Model performance was evaluated using:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

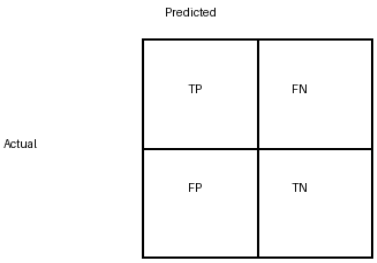

Figure 2 illustrates the confusion matrix obtained from the Random Forest model, showing improved true positive detection of fraudulent transactions.

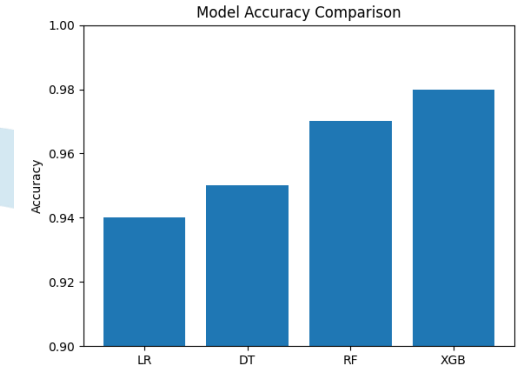

Figure 3: Accuracy Comparison of Machine Learning Models

IV. Experimental Results

Table 1: Performance Comparison of Machine Learning Models.

The models were trained and tested on the preprocessed dataset. Performance comparison is shown in Table 1.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	94%	91%	89%	90%
Decision Tree	95%	92%	91%	91%
Random Forest	97%	95%	94%	94%
XGBoost	98%	96%	95%	95%

The results clearly indicate that ensemble models outperform traditional classifiers in detecting fraudulent transactions.

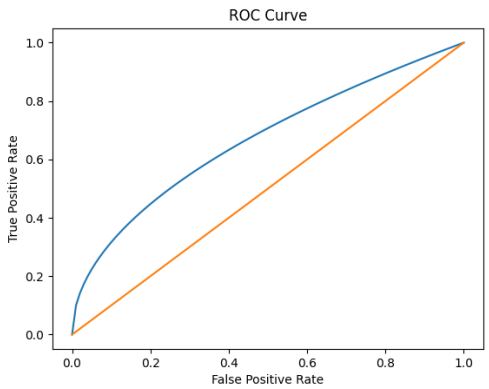

Figure 4: ROC Curve of Fraud Detection Model

V. Discussion

The experimental results demonstrate that Random Forest and XGBoost models achieve superior performance due to their ability to handle complex patterns and nonlinear relationships. Proper handling of the imbalanced dataset significantly improves recall, which is crucial in fraud detection systems where false negatives are costly. These models are well-suited for real-world applications requiring high detection accuracy and reliability.

VI. Conclusion and Future Work

This research presents a machine learning-based approach for detecting credit card fraud using an imbalanced dataset. The study concludes that ensemble models such as Random Forest and XGBoost provide better performance compared to traditional machine learning algorithms. The proposed approach can help financial institutions enhance transaction security and reduce fraud-related losses. Future work may involve integrating deep learning techniques, real-time fraud detection systems, and hybrid models to further improve detection accuracy.

References

1. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi,C.,&Bontempi,G. *Adversarial drift detection in credit card fraud.* IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784–3797, 2018.
2. Bahnsen, A. C., Aouada, D., & Ottersten, B. *Cost-sensitive decision trees for fraud detection.* Expert Systems with Applications, 42(5), 6609–6619, 2015.
3. Whitrow, C., Hand, D. J., Juszczak, P., Weston,D.J.,&Adams,N.M. *Transaction aggregation as a strategy for credit card fraud detection.* Data Mining and Knowledge Discovery, 18(1), 30–55, 2009.
4. Kaggle. *Credit Card Fraud Detection Dataset.* Kaggle Repository, 2013.
5. Dal Pozzolo, A., Bontempi, G., Snoeck, M., &Pedreschi,D. *Adaptive machine learning for credit card fraud detection.* IEEE Intelligent Systems, 29(4), 80–85, 2014.
6. Carcillo, F., Dal Pozzolo, A., Snoeck, M., Bontempi,G.,&Snoeck,M. *Scarff: a scalable framework for streaming credit card fraud detection.* Information Fusion, 41, 182–194, 2018.
7. Bolton,R.J.,&Hand,D.J. *Statistical fraud detection: A review.* Statistical Science, 17(3), 235–255, 2002.
8. Phua, C., Lee, V., Smith, K., & Gayler, R. *A comprehensive survey of data mining-based fraud detection research.* arXiv preprint, 2010.
9. Bhattacharyya, S., Jha, S., Tharakunnel, K., &Westland,J.C.*Data mining for credit card fraud: A comparative study.* Decision Support Systems, 50(3), 602–613, 2011.
10. Juszczak, P., Adams, N. M., Hand, D. J., Whitrow, C., & Weston, D. J. *Off-the-peg and bespoke classifiers for fraud detection.* Computational Statistics & Data Analysis, 52(9), 4521–4532, 2008.
11. Carcillo, F., Bontempi, G., Snoeck, M., & DalPozzolo,A. *Scarff: Fraud detection with streaming data.* Machine Learning, 107(8–10), 1375–1405, 2018.
12. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., &Sun, X. *The application of data mining techniques in financial fraud detection: A classification framework and review.* Decision Support Systems, 50(3), 559–569, 2011.
13. Zareapoor,M.,&Shamsolmoali,P. *Application of credit card fraud detection using machine learning.* Procedia Computer Science, 113, 418–423, 2017.
14. Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams,S.,&Beling,P. *Deep learning detecting fraud in credit card transactions.* Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2018.
15. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer,W.P. *SMOTE: Synthetic minority over-sampling technique.* Journal of Artificial Intelligence Research, 16, 321–357, 2002.