IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning–Driven Credit Card Fraud Detection

[1]Vipula Shamrao Chavan, [2]Dr. Sachin Sukhadeo Bere, [3]Dr. Dinesh Bhagwan Hanchate

[1]ME Student, [2]Associate Professor, [3]Professor
Department of Computer Engineering,
Dattakala Group of Institution Faculty of Engineering, Swami Chincholi, Savitribai Phule Pune University, Maharashtra, India

Abstract : The growing dependence on credit cards for digital transactions has led to an increase in fraudulent activities, posing serious concerns for the financial industry. This paper explores the application of machine learning (ML) and deep learning (DL) techniques in detecting credit card fraud. By reviewing existing research, it highlights effective methods for tackling issues such as class imbalance, evolving fraud tactics, and false positive rates. The study underscores the significance of real-time fraud detection systems that leverage advanced ML and DL models to enhance transaction security.

IndexTerms **- Credit Card Fraud, Deep Learning, Fraud Detection, Machine Learning**

INTRODUCTION

The rapid growth of digital transactions has amplified the risk of credit card fraud, causing substantial financial losses globally. Fraudulent activities, such as unauthorized transactions and identity theft, exploit vulnerabilities in payment systems. Machine learning (ML) offers a promising solution by enabling systems to learn patterns from historical data and detect anomalies in real-time [2]. Unlike rule-based systems, ML models adapt to evolving fraud tactics, improving detection accuracy. This paper proposes a machine learning-driven framework for credit card fraud detection, focusing on supervised and unsupervised techniques. The study aims to evaluate model performance, address challenges like imbalanced datasets, and propose a scalable system architecture [7].

REVIEW OF LITERATURE

The literature on machine learning for credit card fraud detection highlights a range of methodologies and challenges. Ghanem et al. (2022) evaluated supervised learning techniques, including Decision Trees and Logistic Regression, using a synthetic dataset. Their findings showed high accuracy but underscored the challenge of imbalanced datasets, where fraudulent transactions are significantly outnumbered by legitimate ones, leading to biased models [1]. Alarfaj et al. (2022) conducted a comprehensive study on advanced algorithms, such as XGBoost and Deep Neural Networks (DNNs). They demonstrated that DNNs excel in capturing intricate fraud patterns, particularly when combined with feature engineering techniques like transaction aggregation [2].

Tressa et al. (2023) focused on ensemble methods, specifically Random Forest, which achieved a low false positive rate due to its ability to handle high-dimensional data effectively. Their work emphasized the importance of preprocessing steps, such as normalization and outlier removal, to enhance model performance [3]. Doshi et al. (2024) proposed a hybrid framework combining supervised and unsupervised learning. They utilized clustering techniques to identify anomalies in unlabeled data, followed by classification to confirm fraudulent transactions, addressing the issue of limited labeled datasets [4].

Jaswant et al. (2024) provided a detailed review of recent advancements, highlighting the critical role of feature selection in improving model robustness. They noted that features like transaction time, location, and user behavior patterns significantly enhance detection accuracy [5]. Gupta (2023) explored cost-sensitive learning, prioritizing high-value transactions to minimize financial losses. Their approach adjusted model thresholds to focus on impactful fraud cases, improving practical applicability [6].

Dastidar et al. (2024) conducted a survey of ML methods, emphasizing deep learning's ability to model temporal dependencies in transaction sequences. They highlighted the potential of Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks for real-time detection [7]. Sagar (2024) investigated lightweight neural architectures for resource-constrained environments, achieving comparable accuracy with reduced computational overhead [8]. Similarly, Sumedha et al. (2023) explored deep learning models, advocating for transfer learning to leverage pre-trained networks for fraud detection, particularly in scenarios

with limited training data [9]. Collectively, these studies underscore the need for adaptive, scalable, and computationally efficient solutions to combat evolving fraud patterns.

SYSTEM ARCHITECTURE

Detecting fraudulent transactions in credit card data involves a systematic series of steps designed to effectively distinguish between genuine and suspicious activity. The approach adopted in this study can be broadly categorized into the following key stages, as depicted in Fig. 1.

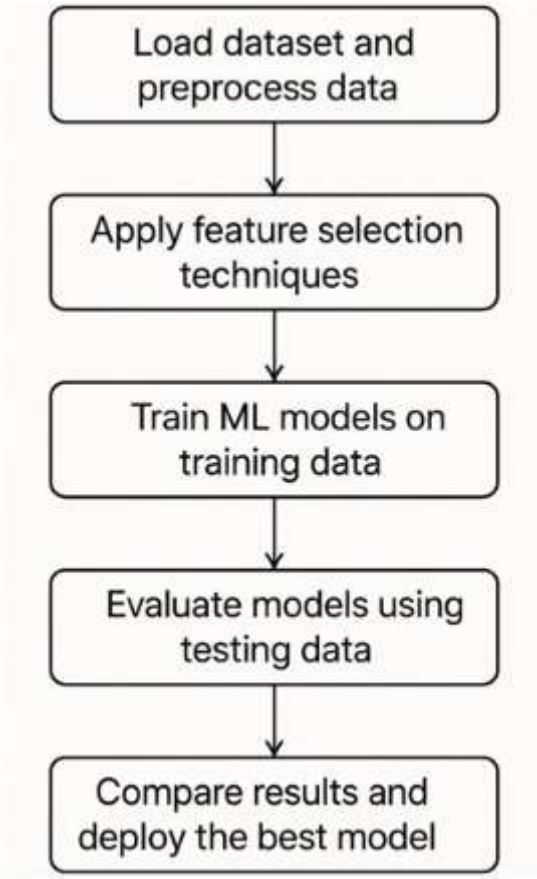

Fig. 1 Flow diagram of Credit Card Fraud Detection

1. Data Acquisition and Preprocessing

The process initiates with the acquisition of a comprehensive dataset comprising historical credit card transactions. Prior to model development, the raw data is subjected to preprocessing operations such as data cleaning to remove inconsistencies, imputation of missing values, and normalization of numerical attributes to ensure uniform scaling. A significant challenge in fraud detection is the pronounced class imbalance, where fraudulent records are far fewer than legitimate ones. To address this, techniques like Synthetic Minority Over-sampling Technique (SMOTE) are employed to rebalance the dataset and ensure more effective model training.

2. Data Splitting

Following preprocessing, the dataset is divided into two main subsets: the Training Set and the Test Set. The training set is utilized to develop predictive models, while the test set is reserved for performance validation on unseen data. This partitioning simulates real-world conditions, allowing for a fair evaluation of each model's generalization capability.

3. Feature Engineering

To enhance model accuracy and detection capability, feature engineering is conducted. This step involves deriving meaningful variables from the original data using statistical transformations and domain-specific knowledge. The goal is to extract characteristics that are strong indicators of fraudulent behavior, thereby enriching the dataset with more predictive features.

4. Model Development and Training

A range of machine learning and deep learning models are explored to identify the most effective fraud detection solution. Traditional ML techniques such as Support Vector Machine (SVM) and Random Forest (RF) are evaluated alongside DL models like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. Each model is trained on the training data, and its performance is benchmarked to determine its suitability for detecting fraudulent activity.

RESULT AND DISCUSSION

A comprehensive evaluation of the machine learning and deep learning models was conducted using key performance metrics, including accuracy and AUC-ROC. The findings provide important insights into the strengths and limitations of each model:

Model	Accuracy	AUC-ROC	Recall (Class 0)	Recall (Class 1)	Key Observations
Logistic Regression	58.06%	0.6425	50%	Moderate	Weak precision-recall balance; high false negatives for Class 0
Random Forest	92.43%	0.9726	High	High	Best overall performance; strong precision and recall; well-balanced classification
XGBoost	48.81%	0.7059	25%	96%	High false positives; severe class imbalance impact
Deep Learning Model	68.39%	0.7469	Moderate	Moderate	Balanced recall; needs further tuning for improved fraud detection accuracy

Here is the detailed analysis of each model based on experimentation.

Logistic Regression
Accuracy: 58.06%
AUC-ROC: 0.6425
The logistic regression model exhibited limited classification performance, particularly in identifying non-fraudulent transactions (Class 0), with a recall of only 50%. This indicates a high rate of false negatives. The overall precision-recall trade-off was suboptimal across both classes, highlighting the model's inadequacy in handling class imbalance effectively.

Random Forest
Accuracy: 92.43%
AUC-ROC: 0.9726
The random forest model outperformed all other approaches, demonstrating high accuracy and excellent AUC-ROC. It achieved strong precision and recall across both classes, indicating a well-balanced classification performance. The model was particularly effective in minimizing both false positives and false negatives, making it the most reliable among the tested algorithms.

XGBoost
Accuracy: 48.81%
AUC-ROC: 0.7059
XGBoost showed a high recall for fraudulent transactions (Class 1) at 96%, but performed poorly for non-fraudulent cases (Class 0), with a recall of only 25%. This led to a significant number of false positives. The results suggest that the model is highly sensitive to class imbalance, which adversely affects its overall classification reliability.

Deep Learning Model
Accuracy: 68.39%
AUC-ROC: 0.7469
The deep learning model achieved moderately balanced recall across both classes, with a notable improvement over logistic regression and XGBoost. However, it still struggled with accurately detecting fraudulent transactions. The model demonstrates potential but requires further optimization, particularly in terms of hyperparameter tuning and class balancing strategies.

CONCLUSION

This study demonstrates that Random Forest outperforms Logistic Regression, XGBoost, and Deep Learning models in credit card fraud detection, achieving the highest accuracy and AUC-ROC. The proposed system supports real-time detection and scalability, addressing key fraud prevention challenges. While supervised models excel, future research should explore hybrid approaches, explainable AI for transparency, and federated learning for privacy. Overall, Random Forest proves to be an effective solution for safeguarding financial transactions against fraud.

REFERENCES

1] Ghanem, M., Elkaffas, S., & Madbouly, M. (2022). Machine Learning Technique for Credit Card Fraud Detection. 2022 32nd International Conference on Computer Theory and Applications (ICCTA), 82-89. https://doi.org/10.1109/ICCTA58027.2022.10206291
2] Alarfaj, F., Malik, I., Khan, H., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit Card Fraud Detection Using State-of-the-art Machine Learning and Deep Learning Algorithms. IEEE Access, PP, 1-1. https://doi.org/10.1109/ACCESS.2022.3166891
3] Tressa, N., Asha, V., M, G., Padanoor, S., Tabassum, R., Dharmesh, D., & Saju, B. (2023). Credit Card Fraud Detection Using Machine Learning. 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), 1-6. https://doi.org/10.1109/ASIANCON58793.2023.10270805
4] Doshi, N., Chillarge, G., & Shekapure, S. (2024). Enhanced Detection of Credit Card Fraud Using Machine Learning Techniques. International Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2024.62899

5] Jaswant, T., Manoj, G., Vamisdhar, V., Aravind, A., Reddy, S., Patni, J., Bahadure, N., & Kumar, R. (2024). Credit Card Fraud Detection Using Machine Learning- A Comprehensive Review. 2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON), 1-4. https://doi.org/10.1109/DELCON64804.2024.10866830
6] Gupta, J. (2023). Credit Card Fraud Detection Using Machine Learning Algorithms. International Journal of Science and Research (IJSR). https://doi.org/10.21275/sr231123121203
7] Dastidar, K., Caelen, O., & Granitzer, M. (2024). Machine Learning Methods for Credit Card Fraud Detection: A Survey. IEEE Access, 12, 158939-158965. https://doi.org/10.1109/ACCESS.2024.3487298
8] Sagar, V. (2024). CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. https://doi.org/10.55041/ijsrem32382
9] M., Sumedha, K., & Samhitha, V. (2023). Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms. International Journal For Multidisciplinary Research. https://doi.org/10.36948/ijfmr.2023.v05i03.2926