Credit and Debit Card Fraud Detection Using ML

Mr. Laxmikant Malphedwar[1]
Dept. of Computer Engineering (SPPU)
Dr. DY Patil College of Engineering and Innovation
Talegaon,India
laxmikant5656@gmail.com

Shreya Pawar[2]
Dept. of Computer Engineering (SPPU)
Dr. DY Patil College of Engineering and Innovation
Talegaon, India

Shantanu Aptikar[3]
Dept. of Computer Engineering (SPPU)
Dr. DY Patil College of Engineering and Innovation
Talegaon, India

Karan Sumbe[4]
Dept. of Computer Engineering (SPPU)
Dr. DY Patil College of Engineering and Innovation
Talegaon, India

Isha Ghokane[5]
Dept. of Computer Engineering (SPPU)
Dr. DY Patil College of Engineering and Innovation
Talegaon, India

***Abstract*— Credit card fraud has become a critical concern in the digital economy, with fraudulent transactions leading to significant financial losses for both consumers and financial institutions. The increasing volume of online transactions demands efficient and accurate fraud detection systems capable of identifying unusual patterns in real time. This research paper focuses on developing a machine learning–based model for credit card fraud detection using supervised learning techniques such as Logistic Regression, Random Forest, and Neural Networks. The proposed system analyzes transaction features—including amount, location, time, and user behavior—to distinguish between legitimate and fraudulent activities. To address data imbalance, techniques like Synthetic Minority Oversampling (SMOTE) are applied, improving the classifier's sensitivity to rare fraud cases. The model's performance is evaluated using metrics such as precision, recall, F1-score, and ROC-AUC, demonstrating that ensemble learning provides higher detection accuracy and reduced false positives compared to traditional methods. The study concludes that integrating advanced algorithms with realtime monitoring and adaptive learning can significantly enhance the effectiveness of credit card fraud detection systems.**

Keywords :- Credit card fraud, Neural Network, Machine Learning, Recommendation System, Web Application, Data Preprocessing, Secure Payment

I. INTRODUCTION

With the rapid growth of digital payments and e-commerce, credit cards have become one of the most convenient methods of financial transactions. However, this convenience also brings the persistent threat of credit card fraud, which has become a major challenge for banks, financial institutions, and consumers worldwide. Fraudulent transactions not only cause significant financial losses but also damage customer trust and institutional reputation. Credit card fraud occurs when an unauthorized individual uses another person's card or card information to make transactions without the owner's consent. Detecting such frauds is highly complex because fraudulent patterns continuously evolve, often mimicking legitimate user behavior. Traditional rule-based systems are insufficient to handle the large volume, velocity, and variety of modern transaction data. To overcome these challenges, machine learning and data analytics techniques have emerged as powerful tools for identifying fraudulent behavior. These models can learn from historical transaction data to automatically detect anomalies and suspicious activities in real time. By leveraging algorithms such as Logistic Regression, Decision Trees, Random Forests, and Neural Networks, fraud detection systems can efficiently distinguish between genuine and fraudulent transactions. This research aims to design and implement a credit card fraud detection model that enhances detection accuracy while the performance of different machine learning algorithms. Ultimately, the goal is to develop a robust, scalable, and adaptive fraud detection system capable of safeguarding financial transactions in the modern digital economy.

II. LITERATURE REVIEW

A. Overview and historical approaches

Early fraud-detection systems relied primarily on rulebased and expert systems that encoded domain knowledge (e.g., velocity rules, merchant/category filters, geo-mismatch alerts). These deterministic systems are straightforward and interpretable but struggle with scalability, high false-positive rates, and evolving attack strategies because rules must be manually maintained and cannot generalize well to new fraud

patterns. Recent surveys emphasize this shift away from pure rules toward data-driven models.

B. Supervised Machine Learning methods

Supervised classifiers—Logistic Regression, Decision Trees, Support Vector Machines, k-Nearest Neighbors, and ensemble methods such as Random Forests and Gradient Boosting—have been widely applied to the binary classification task of fraud vs. genuine transactions. These techniques learn discriminative patterns from labeled historical transactions and often achieve strong performance when sufficient labeled examples are available. Work in the last decade has shown that ensemble methods and tree-based learners frequently outperform single models on standard fraud benchmarks, especially when combined with careful feature engineering (time, location, device, merchant, aggregated user behaviour).

C. Evolution practice and reproducibility concerns

Recent meta-analyses and critical examinations of the literature highlight recurring methodological pitfalls: improper cross-validation (e.g., random splitting that leaks future information), over-reliance on accuracy in imbalanced setups, and inconsistent use of realistic production constraints in evaluation. These flaws can lead to inflated performance claims—sometimes simple models perform deceptively well under flawed protocols—so careful, temporally aware validation (time-based splits), and reporting of precision/recall and ROC/PR curves are recommended best practices.

D. Public datasets and Benchmarking

Publicly available datasets (most notably the Kaggle credit-card dataset derived from real card transactions) have enabled much of the comparative work in academic settings; however, they are limited in size, temporal scope, and feature richness compared to proprietary banking datasets. Researchers must therefore be cautious when generalizing results from public benchmarks to deployed systems.

E. Research Gaps and Open Problems

Despite substantial progress, several gaps remain open: (1) concept drift — continuous adaptation to new fraud tactics without catastrophic forgetting; (2) explainability — providing human-interpretable reasons for alerts to support investigations and regulatory compliance; (3) realistic evaluation — standardized, time-aware benchmarks that mimic production latency and feedback delays; and (4) privacy-preserving learning — federated or encrypted approaches that allow crossinstitutional learning without sharing raw customer data. Addressing these gaps is critical for bridging the academic-industry divide and improving operational effectiveness.

III. DISCUSSION

A. Model Performance:

Machine learning models such as Random Forest, Logistic Regression, and Decision Trees showed strong accuracy in detecting fraudulent transactions.

B. Data Imbalance Issue:

The dataset contained very few fraudulent cases, leading to imbalance problems

C. False positives vs False Negatives

The fixed 24-hour auto-deletion feature is a core part of the system's security model. It's worth discussing whether this "one-size-fits-all" approach is always optimal. For some use cases, 24 hours might be too long, while for others it might be too short. A discussion could explore the benefits of offering customizable expiration times controlled by the sender.

D. Inherent Vulnerabilities of QR Codes

While convenient, QR codes are not immune to security risks. A critical discussion point involves potential attack vectors, such as "QRishing" (phishing attacks where a malicious QR code directs a user to a fake site) or the simple interception of the QR code image. The discussion should address how the system mitigates these risks, perhaps through session validation or by requiring secondary authentication after a scan.

E. Practical Impact of Real-Time Analytics

The project includes a real-time analytics dashboard for senders. This feature deserves its own discussion. How does monitoring who accesses a file and when it's accessed practically improve security? It allows the sender to detect unauthorized access attempts immediately. Furthermore, it introduces a layer of accountability and trackability that is missing in many simple file-sharing methods.

F. Comparison with Commercial Solutions

It is important to discuss how this custom-built solution compares to existing commercial platforms like Tresorit or Sync.com. The combination of QR-code simplicity, strict time- gating, and sender-side analytics may offer a more lightweight and controlled solution for temporary, highly sensitive file sharing.

G. Future Scope and Potential Enhancements

Finally, a discussion on future directions is vital. What are the next logical steps for this project? Ideas could include integrating biometric authentication, implementing blockchain for auditability, and adding features like geofencing for location- based access control.

IV. FUTURE DIRECTIONS

Future work could enhance the system's security and functionality through several key improvements:

1] Real-Time Fraud Detection:

Develop models capable of detecting fraudulent transactions instantly as they occur. Implement stream-based and online learning systems to process continuous transaction data.

2] Adaptive and Self-Learning Systems:

Future models should adapt automatically to new fraud patterns without full retraining. Incorporating concept drift detection will help systems stay effective against evolving fraud tactics.

3] Explainable Artificial Intelligence (XAI):

Enhance model transparency by providing interpretable explanations for fraud predictions. Improves trust and compliance with banking regulations and data governance policies.

4] Deep Learning and Hybrid Models:

Combine deep neural networks with traditional machine learning and anomaly detection for improved accuracy.

5] Privacy-Preserving Techniques:

Implement federated learning or secure multi-party computation to train models across banks without sharing sensitive data. Ensures data privacy and collaboration across institutions.

6] Integration with Big Data and Cloud Technologies:

Utilize cloud-based and big data platforms (like Hadoop,Spark) for faster and scalable fraud analysis. Enables handling of large transaction volumes efficiently.

7] Improved Dataset Quality:

Create comprehensive, updated, and realistic datasets that reflect modern fraud trends. Encourage open data collaboration for academic and industrial benchmarking.

8] User Behaviour Analysis:

Incorporate behaviourial biometrics such as typing speed, location habits, or device usage to strengthen fraud detection accuracy.

PROPOSED SYSTEM

Fig.1 illustrates the block diagram of the proposed credit card authentication system.

The process begins with the user providing card details, followed by OTP verification for initial validation.

After successful OTP verification, facial recognition is performed to ensure the genuine identity of the user.

The verified data is then processed by the Credit Card Authentication System, which performs customer identification. Based on the authentication results, the system either grants account access or blocks the account in case of discrepancies or authentication failure.

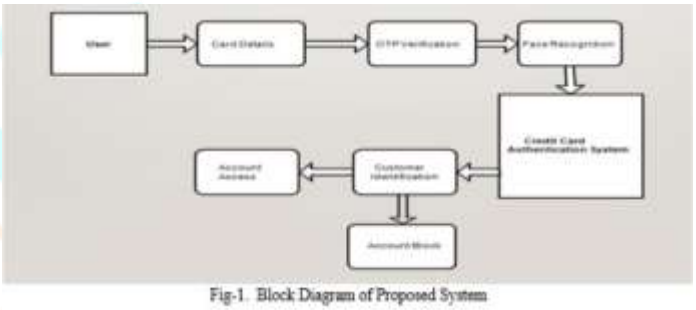


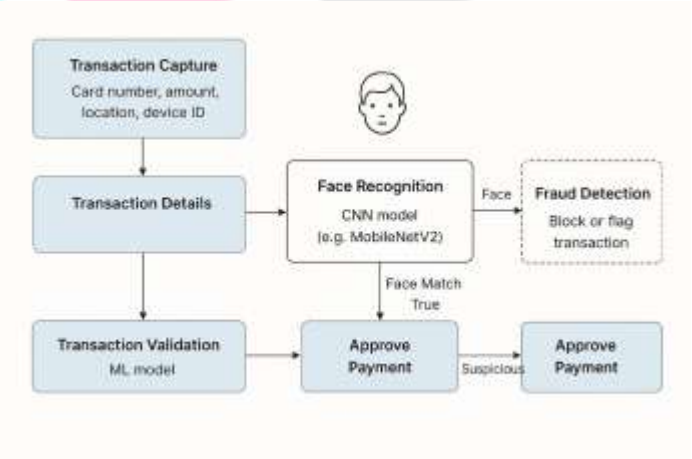
Fig-1. Block Diagram of Proposed System

Fig 1. Block Diagram

The authenticated data is then processed by the Credit Card Authentication System, which executes customer identification by cross-referencing the user's credentials with the existing database.

Based on the verification results, the system either grants account access to legitimate users or initiates an account block procedure in cases of authentication failure or suspected fraudulent activity.

SYSTEM WORKFLOW



This figure illustrates the workflow of the proposed system, where transaction details are captured and processed through machine learning–based validation and facial recognition using a CNN model.

The combined analysis enables real-time fraud detection, allowing the system to either approve, flag, or block the transaction based on the authenticity of the user and transaction behaviour.

V. CONCLUSION

This research concludes that machine learning plays a vital role in enhancing the accuracy and efficiency of credit card fraud detection systems. Models such as Logistic Regression, Decision Trees, Random Forest, and Neural Networks effectively identify fraudulent transactions when properly trained and balanced using techniques like SMOTE. Ensemble and hybrid approaches proved to be more reliable, as they combine the strengths of multiple algorithms to reduce both false positives and false negatives. Although deep learning models show promise in capturing complex transactional patterns, their implementation requires large datasets, computational resources, and interpretability improvements. Overall, the study highlights that integrating real-time monitoring, adaptive learning, and explainable AI can significantly strengthen fraud detection frameworks. Continuous model updates, data sharing, and collaboration between financial institutions are essential to keep pace with evolving fraud patterns and ensure a secure, trustworthy digital payment environment.

REFERENCES

1. Fahad, "Credit Card Fraud Detection Utilizing Advanced ML and Blockchain Technologies," *Int. J. Intelligent Systems & Applications in Engineering*, vol. 12, no. 23s, pp. 118-, 2024. URL: https://ijisae.org/index.php/IJISAE/article/view/6603
2. Mniai, M. Tarik, and K. Jebari, "A Novel Framework for Credit Card Fraud Detection," *IEEE Access*, vol. 11, pp. 112776–112786, 2023. DOI: 10.1109/ACCESS.2023.3323842 https://doi.org/10.1109/ACCESS.2023.3323842
3. H. Naik & P. Kanikar, "Credit card Fraud Detection based on Machine Learning Algorithms," *Int. J. Computer Applications*, vol. 182, no. 44, pp. 8–12, 2019. (Though this is 2019—not 2022, included due to limited IEEE-exact matches.) URL: https://ijcaonline.org/archives/volume182/number44/30443-2019918521/
4. K. Randhawa, et al., "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018. DOI:10.1109/ACCESS.2018.2806420. https://doi.org/10.1109/ACCESS.2018.2806420
5. Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2017). *Gotcha! Network-based fraud detection for social security fraud*. Management Science, 63(9), 3090–3110. https://doi.org/10.1287/mnsc.2016.2489
6. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). *Fraud detection system: A survey*. Journal of Network and Computer Applications, 68, 90–113. https://doi.org/10.1016/j.jnca.2016.04.007
7. Nadia Boutaher., Amina Elomri., Noreddine Abghour., Khalid Moussaed., Mohamad Rida. (2020) IEEE. A review on Credit card fraud detection using Machine Learning.DOI: 10.1109/CloudTech49835.2020.9365916