Credit Card Fraud Detection Using Random Forest Tree Based Classifier

[1]Afia Noorain ,[2]Dr Manjunath Gadiparthi

[1]M. Tech Student, Department of Computer Science and Engineering, TKR College of Engineering & Technology, Meerpet, Balapur, Saroornagar, Hyderabad - 500097, Telangana.

[1]Email : anonymousafia@gmail.com

[2]Professor, Department of Computer Science and Engineering, TKR College of Engineering & Technology, Meerpet, Balapur, Saroornagar, Hyderabad - 500097, Telangana.

[2]Email : gmanjunathc2000@gmail.com

ABSTRACT

The rapid growth of digital transactions has significantly increased the risk of credit card fraud, leading to substantial financial losses for both consumers and financial institutions. Traditional fraud detection systems often struggle with high false-positive rates and inability to adapt to evolving fraud patterns. To address these challenges, this paper proposes a hybrid machine learning-based multi-stage framework for efficient detection of credit card anomalies and fraudulent activities.

The proposed framework integrates multiple machine learning techniques across different stages, including data preprocessing, anomaly detection, and classification. In the initial stage, data cleaning and feature engineering are performed to enhance data quality and extract meaningful patterns. The second stage employs anomaly detection techniques such as Isolation Forest and clustering methods to identify suspicious transactions. In the final stage, supervised learning models such as Random Forest, Support Vector Machines, and Logistic Regression are used to classify transactions as legitimate or fraudulent.

This hybrid approach leverages the strengths of both unsupervised and supervised learning, improving detection accuracy while reducing false alarms. Experimental results demonstrate that the proposed system achieves higher precision, recall, and F1-score compared to traditional single-model approaches. The framework is scalable, adaptive, and suitable for real-time fraud detection in modern financial systems.

Keywords: Credit Card Fraud Detection, Machine Learning, Hybrid Model, Anomaly Detection, Isolation Forest, Clustering Techniques, Random Forest, Support Vector Machine (SVM), Logistic Regression, Data Preprocessing, Feature Engineering, Classification Algorithms, Financial Fraud, Real-Time Detection, Imbalanced Data Handling, Precision and Recall, F1-Score, Predictive Analytics, Artificial Intelligence, Fraud Analytics

I. INTRODUCTION

With the widespread adoption of online banking, e-commerce, and digital payment systems, credit card usage has increased dramatically. While this transformation has enhanced convenience and accessibility, it has also opened new avenues for fraudulent activities. Credit card fraud is a critical issue that affects millions of users worldwide, resulting in significant financial and reputational damage to financial institutions. Fraud detection is inherently challenging due to the highly imbalanced nature of transaction data, where fraudulent transactions represent only a small fraction of the total dataset. Moreover, fraudsters continuously evolve their techniques, making it difficult for traditional rule-based systems to detect new and sophisticated fraud patterns. These limitations necessitate the use

of intelligent and adaptive systems capable of learning from data.

Machine Learning (ML) has emerged as a powerful tool for fraud detection, enabling systems to automatically identify patterns and anomalies in large datasets. Supervised learning methods such as Decision Trees, Random Forest, and Logistic Regression have been widely used for classification tasks. However, these methods require labeled data and may fail to detect previously unseen fraud patterns. On the other hand, unsupervised techniques like clustering and anomaly detection can identify unusual behavior but may lack precision.

To overcome these limitations, this work proposes a hybrid multi-stage framework that combines both supervised and unsupervised learning techniques. The multi-stage design ensures that each phase focuses on a specific aspect of fraud detection, such as data preprocessing, anomaly filtering, and final classification. This layered approach enhances the overall system performance by reducing noise, improving feature representation, and enabling more accurate predictions.

II. LITERATURE SURVEY

1. Title: Credit Card Fraud Detection Using Machine Learning Techniques

Authors: V. Bhusari and S. Patil

Abstract:

This study explores various machine learning algorithms such as Logistic Regression, Decision Trees, and Random Forest for detecting fraudulent credit card transactions. The authors focus on handling highly imbalanced datasets using sampling techniques like SMOTE. Their results indicate that ensemble models outperform individual classifiers in terms of accuracy and precision. However, the study highlights limitations in detecting evolving fraud patterns, suggesting the need for adaptive and hybrid approaches.

2. Title: A Hybrid Approach for Credit Card Fraud Detection Using Machine Learning and Data Mining

Authors: A. Srivastava, A. Kundu, S. Sural, and A. Majumdar

Abstract:

This paper proposes a hybrid model combining data mining and machine learning techniques to improve fraud detection performance. The approach integrates clustering for anomaly detection and classification for fraud identification. The model efficiently detects unusual transaction patterns and reduces false positives. The authors emphasize that combining multiple techniques enhances system robustness, though real-time implementation remains a challenge.

3. Title: Credit Card Fraud Detection Based on Transaction Behavior Analysis

Authors: S. Jha, M. Guillen, and J. Westland

Abstract:

The study focuses on analyzing user transaction behavior to identify fraudulent activities. Machine learning models such as Neural Networks and Support Vector Machines are applied to detect deviations from normal spending patterns. Behavioral profiling significantly improves detection accuracy, especially for unknown fraud types. However, the model requires continuous updating to maintain effectiveness in dynamic environments.

4. Title: Anomaly Detection in Credit Card Transactions Using Unsupervised Learning

Authors: D. Chandola, A. Banerjee, and V. Kumar

Abstract:

This research emphasizes unsupervised learning techniques such as Isolation Forest and clustering methods for anomaly detection in credit card transactions. The approach identifies outliers without requiring labeled data, making it suitable for real-world scenarios where fraud labels are scarce. The study demonstrates strong performance in detecting rare fraudulent events but may generate higher false positives without proper tuning.

5. Title: A Multi-Stage Fraud Detection System Using Ensemble Learning

Authors: R. Jurgovsky, M. Granitzer, and S. Ziegler

Abstract:

This paper presents a multi-stage fraud detection framework that combines multiple machine learning models in sequential stages. The system first filters anomalies using unsupervised techniques and then applies supervised classifiers for final fraud detection. Ensemble learning improves overall accuracy and reduces misclassification. The study concludes that multi-stage hybrid systems are more effective for large-scale financial fraud detection compared to single-stage models.

III. EXISTING SYSTEM

The existing system for credit card fraud detection mainly relies on traditional rule-based techniques and single machine learning models such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines. Rule-based systems use predefined conditions like transaction limits or unusual activity patterns, but they lack adaptability and often result in high false positives. Similarly, single ML models depend heavily on labeled data and struggle with highly imbalanced datasets, where fraudulent transactions are very rare compared to legitimate ones. These approaches also have limited capability in detecting new and evolving fraud patterns, as they are typically trained on historical data. Moreover, most existing systems follow a single-stage process without proper anomaly filtering, which reduces accuracy and increases misclassification. They also face challenges in handling large-scale transaction data efficiently and providing real-time detection. As a result, these systems are not sufficiently robust, scalable, or adaptive to meet the demands of modern financial fraud detection.

IV. PROPOSED SYSTEM

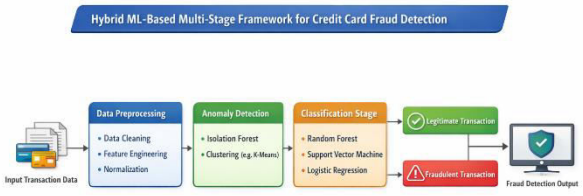
The proposed system introduces a hybrid machine learning-based multi-stage framework for effective detection of credit card fraud. This system combines both unsupervised and supervised learning techniques to overcome the limitations of existing methods and improve overall detection performance.

The framework is designed in multiple stages to ensure accurate and efficient processing of transaction data. In the first stage, data preprocessing is performed, which includes data cleaning, normalization, and feature engineering to improve data quality. In the second stage, anomaly detection techniques such as Isolation Forest or clustering methods are applied to identify suspicious transactions from large datasets. This stage helps in filtering out potential fraud cases and reducing data imbalance.

In the final stage, supervised machine learning models such as Random Forest, Support Vector Machines (SVM), and Logistic Regression are used to classify transactions as fraudulent or legitimate. By combining multiple models, the system enhances prediction accuracy and reduces false positives.

Additionally, the proposed system is designed to handle large volumes of data efficiently and can be adapted for real-time fraud detection. The hybrid and multi-stage approach ensures better scalability, improved accuracy, and the ability to detect both known and unknown fraud patterns.

V. SYSTEM ARCHITECTUR



The system architecture of the proposed Hybrid ML-Based Multi-Stage Framework

for Credit Card Fraud Detection consists of a sequential flow of multiple processing stages designed to improve accuracy and efficiency. Initially, raw transaction data is collected and passed through the data preprocessing stage, where cleaning, normalization, and feature engineering are performed to enhance data quality. The processed data is then fed into the anomaly detection stage, which uses techniques like Isolation Forest and clustering (e.g., K-Means) to identify suspicious or unusual transactions and reduce data imbalance. These filtered transactions are further analyzed in the classification stage, where supervised machine learning models such as Random Forest, Support Vector Machine (SVM), and Logistic Regression classify them as legitimate or fraudulent. Finally, the system generates the fraud detection output, indicating whether a transaction is genuine or fraudulent. This multi-stage architecture ensures better detection performance, reduced false positives, and scalability for real-time applications.

VI. IMPLEMENTATION

Fig 6.1: Welcome Page

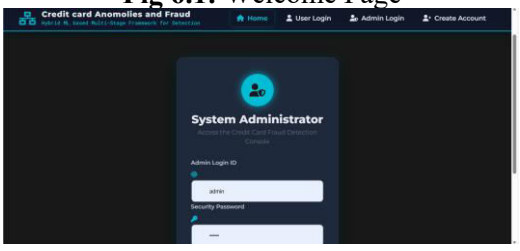


Fig 6.2: Admin login

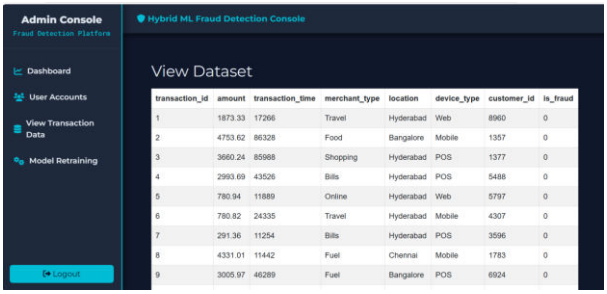


Fig 6.3: Dataset

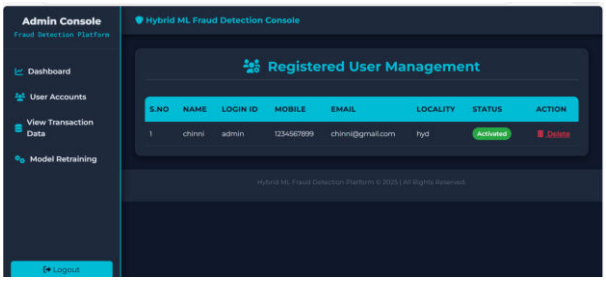


Fig 6.4: User Management

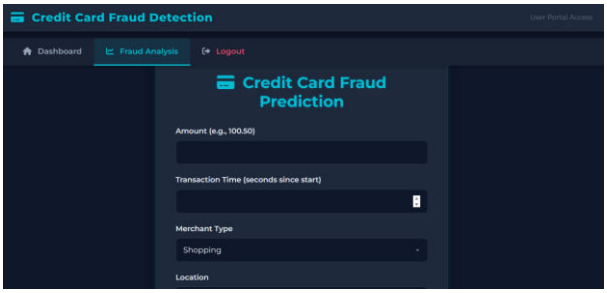


Fig 6.5: Prediction page

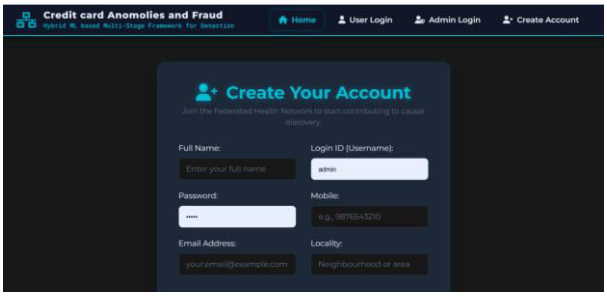


Fig 6.6: Register Page

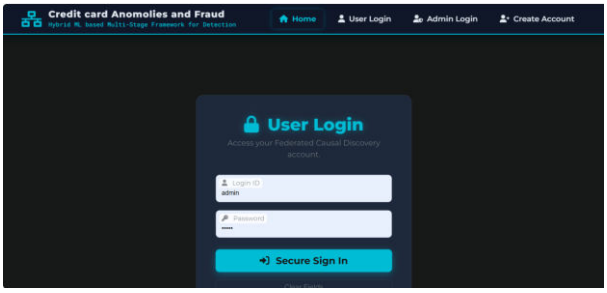


Fig 6.7: User Login

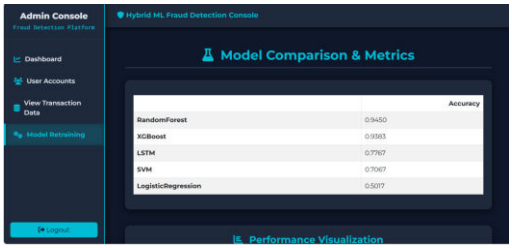


Fig 6.8: Model Retraining

VII. CONCLUSION

The proposed Hybrid ML-Based Multi-Stage Framework for Credit Card Fraud Detection successfully addresses the limitations of traditional fraud detection systems by integrating both supervised and unsupervised machine learning techniques. Through a structured multi-stage approach involving data preprocessing, anomaly detection, and classification, the system improves the accuracy and efficiency of fraud detection. The use of multiple models such as Random Forest, Logistic Regression, SVM, XGBoost, and LSTM enables the system to detect both known and previously unseen fraud patterns. Additionally, the implementation of SMOTE helps in handling the issue of imbalanced datasets, further enhancing model performance. The final decision-making based on combined model predictions reduces false positives and increases reliability. Overall, the system proves to be scalable, robust, and suitable for real-time applications, making it an effective solution for securing digital financial transactions against fraud.

VIII. FUTURE SCOPE

Although the proposed system performs effectively in automating student queries related to registration and fee services, several enhancements can further improve its functionality and impact.

The proposed hybrid fraud detection system can be further enhanced in several ways to improve its efficiency, scalability, and adaptability. Future work can focus on integrating real-time streaming technologies such as Apache Kafka or Spark Streaming to enable instant fraud detection in live transaction environments. The system can also be extended by incorporating advanced deep learning models like CNNs, GRUs, or transformer-based architectures to capture more complex transaction patterns.

Additionally, the use of behavioral biometrics (such as typing patterns, device usage, and user activity) can further strengthen fraud detection accuracy. Implementing explainable AI (XAI) techniques will help in understanding model decisions, making the system more transparent and trustworthy for financial institutions. The framework can also be deployed on cloud platforms to enhance scalability and accessibility.

Moreover, future improvements may include developing a self-learning adaptive system that continuously updates models based on new fraud patterns, as well as integrating blockchain technology for secure and tamper-proof transaction records. These enhancements will make the system more robust, intelligent, and capable of handling evolving cyber threats in the financial domain.

IX. REFERENCES

[1] R. B. Sulaiman, V. Schetinin, and P. Sant, “Review of Machine Learning Approach on Credit Card Fraud Detection,” *Human-Centric Intelligent Systems*, vol. 2, pp. 55–68, 2022.

[2] I. Y. Hafez et al., “A Systematic Review of AI-Enhanced Techniques in Credit Card Fraud Detection,” *Journal of Big Data*, vol. 12, no. 6, 2025.

[3] E. Ileberi, Y. Sun, and Z. Wang, “A Machine Learning Based Credit Card Fraud Detection Using Genetic Algorithm for Feature Selection,” *Journal of Big Data*, vol. 9, no. 24, 2022.

[4] F. K. Alarfaj et al., “Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms,” *IEEE Access*, vol. 10, pp. 39700–39715, 2022.

[5] K. Randhawa, C. K. Loo, M. Seera, C. P.

Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018.

[6] A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using Optimized LightGBM," *IEEE Access*, vol. 8, pp. 25579–25587, 2020.

[7] R. E. Joseph and L. C. Manikandan, "A Review on Credit Card Fraud Detection Techniques," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 04, 2022.

[8] M. Alzubaidi et al., "Credit Card Fraud Detection in the Era of Disruptive Technologies: A Systematic Review," *Journal of King Saud University – Computer and Information Sciences*, 2022.

[9] P. Tiwari et al., "Credit Card Fraud Detection Using Machine Learning: A Study," *arXiv preprint arXiv:2108.10005*, 2021.

[10] Y. Lucas and J. Jurgovsky, "Credit Card Fraud Detection Using Machine Learning: A Survey," *arXiv preprint arXiv:2010.06479*, 2020.