AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
PEER REVIEWED, REFERRED & INDEXED JOURNAL
E-ISSN: 3069-0102
VOL.5, NO. 2(2025)

A Predictive Analytics Approach for Detecting Fraudulent Financial Transactions Using Supervised Learning

G U K Bhargava[1*], Vemireddy Bhargavi[1], Guddala Bhavya Sri[1], Maragani Navya Sri[1], Gaddala Deepika[1], Nagalla Akshaya[1]

[1]Department of Electronics and Communication Engineering, Mother Teresa Institute of Science and Technology, Sanketika Nagar, Kothuru, Sathupally, Khammam, 507303, Telangana, India

*Correspondence: G U K Bhargava

ABSTRACT

The rapid growth of digital banking and online financial transactions has significantly increased the risk of fraudulent activities, posing serious threats to both customers and financial institutions in terms of monetary loss and reputational damage. As banking systems become more complex and interconnected, identifying fraudulent behavior at an early stage has become essential for maintaining security and trust. This study presents an intelligent fraud detection framework based on Machine Learning (ML) techniques to effectively identify suspicious transactions and minimize financial risks. The proposed system utilizes Artificial Intelligence (AI) to automate and accelerate processes such as transaction verification and anomaly detection, thereby reducing manual effort and improving response time. Various supervised learning algorithms are trained on a publicly available dataset to analyze patterns and uncover hidden relationships associated with fraudulent activities. A major challenge addressed in this work is class imbalance, where fraudulent cases are significantly fewer than legitimate ones; to overcome this, data resampling techniques are applied to ensure balanced learning and improved model performance. Additionally, data preprocessing and feature engineering are performed to enhance the quality and relevance of input data. The experimental evaluation demonstrates that the proposed ML-based framework achieves reliable and accurate fraud detection. This approach contributes to strengthening fraud prevention systems, enabling faster decision-making, and ensuring secure and efficient banking operations.

Key words: Fraud Detection, Banking Transactions, Machine Learning (ML), Ensemble Learning, Real-Time Detection, Transaction Analysis

1. INTRODUCTION

Fraud, as defined in standard lexicons such as the The American Heritage Dictionary of the English Language, refers to intentional deception carried out to obtain unlawful or unfair advantage, and in the financial domain it manifests in multiple forms including insurance fraud, credit card misuse, and accounting manipulation. Fraud detection, therefore, involves identifying anomalous patterns or behavioral inconsistencies even when there is no prior indication of malicious intent. In modern banking ecosystems, fraud has transitioned from isolated incidents by individuals to highly organized operations conducted by coordinated crime groups employing advanced techniques to compromise accounts and execute unauthorized transactions. Reports from the Nigeria Inter-Bank Settlement System highlight the significant rise in fraudulent activities within banking systems, while studies by Javelin Strategy & Research indicate that millions of users have been affected by card-related fraud, resulting in multi-billion-dollar financial losses. To address this growing challenge, advanced computational approaches leveraging Machine Learning (ML) and distributed processing frameworks have been introduced. Technologies such as Apache Spark, integrated with scalable programming environments like Java 8, enable high-speed processing of large transactional datasets through parallel computation, significantly improving detection efficiency. One of the core challenges in fraud detection is the severe class imbalance, where fraudulent transactions represent only a small fraction of total data, leading to biased model predictions and increased false positives. Techniques such as stratified sampling and data

Received: 02-04-2025 | Accepted: 12-05-2025 | Published: 19-05-2025 | www.ajaccm.com

resampling are employed to balance the dataset and enhance model reliability. Additionally, feature standardization is essential to eliminate scale variations among variables, ensuring stable model performance. The system architecture further incorporates secure user authentication, where new users register by providing essential identity details, followed by controlled access to banking services. During transaction processing, an additional security layer is introduced through dynamic verification mechanisms such as security questions, which help distinguish legitimate users from fraudulent actors. Overall, this integrated approach combines data-driven intelligence, scalable computation, and multi-level authentication to build a robust and efficient fraud detection system capable of minimizing financial risks and enhancing trust in digital banking environments.

2. LITERATURE SURVEY

S. M. Darwish, et al. [1] proposed an intelligent credit card fraud detection framework that combined statistical analysis with ML-based classification techniques. Their study focused on analyzing the statistical distribution of transaction datasets to identify anomalous patterns indicative of fraudulent behavior. They utilized Linear Discriminant Analysis and Logistic Regression as core classification methods and further enhanced performance through semantic fusion of multiple classifiers. This hybrid approach enabled the system to capture both linear separability and probabilistic relationships within the data. Their work demonstrated that combining statistical modeling with ML techniques significantly improves detection accuracy and reduces false positives in real-time fraud detection systems. A. Eshghi, et al. [2] introduced a novel fraud detection methodology that integrated the KNN algorithm with an outlier detection mechanism to identify suspicious banking transactions. Their model focused on detecting deviations from normal transaction patterns by analyzing similarity measures between data points. The approach effectively handled uncertainty in financial data by incorporating evidence fusion techniques, allowing the system to make more reliable decisions even in ambiguous scenarios. Their findings highlighted that combining distance-based learning with anomaly detection improves the identification of rare fraudulent events that are often overlooked by traditional classification models. S. Hossain, et al. [3] developed a fraud detection system using ensemble learning techniques that combined Random Forest and Neural Networks for improved classification performance. Their study utilized spatio-temporal data to capture both transactional behavior and time-based patterns, which are critical in identifying fraud trends. By comparing multiple models, they demonstrated that ensemble approaches provide higher accuracy and robustness compared to single classifiers. Their work emphasized that integrating multiple learning models enhances the system's ability to generalize across diverse datasets and improves detection of complex fraud patterns. M. Zamini, et al. [4] conducted a comprehensive survey on anomaly detection techniques across various domains, including banking and financial systems. Their study explored a wide range of methods such as ML-based classification, optimization-driven neural networks, and clustering techniques. They addressed critical challenges such as imbalanced datasets, where fraudulent transactions are significantly fewer than legitimate ones. The study highlighted the use of K-means clustering for grouping similar transaction behaviors and discussed optimization-based backpropagation techniques for improving neural network performance. Their work emphasized that hybrid and adaptive models are essential for handling dynamic fraud patterns in real-world applications.

I. Gonzalez-Carrasco, et al. [5] presented a large-scale fraud detection framework that utilized multiple ML algorithms, including KNN, Logistic Regression, and Naïve Bayes, to analyze banking transaction data. Their study

evaluated model performance using comprehensive metrics such as precision, recall, Matthews correlation coefficient, and balanced classification rate, providing a well-rounded assessment of model effectiveness. They further proposed a scalable real-time fraud detection system (SCARFF) that leveraged big data technologies such as Apache Spark, Cassandra, and Kafka to process massive transaction streams. Their approach demonstrated high scalability, fault tolerance, and improved accuracy, making it suitable for real-world deployment in financial systems. M. Pohoretskyi, et al. [6] focused on improving fraud detection performance by applying feature engineering techniques to reduce false positive rates. Their study emphasized the importance of selecting relevant and informative features to enhance the effectiveness of anomaly detection algorithms. By reducing redundancy and eliminating irrelevant attributes, they improved both computational efficiency and classification accuracy. Their work highlighted that feature engineering plays a crucial role in financial fraud detection, as poorly selected features can lead to misclassification and increased operational costs.

K. Noor, et al. [7] developed a surveillance-based system that utilized Haar cascade classifiers and optical flow methods for detecting and tracking vehicles. Although primarily focused on visual surveillance, their approach contributed to fraud detection by enabling monitoring of suspicious movements and activities in real-world environments. The study demonstrated how computer vision techniques can be integrated with intelligent systems to enhance detection capabilities, especially in scenarios where fraud involves physical movement or behavioral tracking. G. Baader, et al. [8] proposed a method to reduce false positives in fraud detection systems by combining the traditional red flag approach with process mining techniques. Their work focused on analyzing business process flows and identifying deviations that may indicate fraudulent behavior. By integrating process mining with rule-based detection, they improved the accuracy of fraud identification while minimizing incorrect alerts. Their study highlighted that understanding process-level behavior is crucial for enhancing fraud detection systems in financial environments. P. Ravisankar, et al. [9] conducted a study on detecting financial statement fraud using data mining techniques and feature selection methods. Their approach employed ML algorithms such as SVM to classify fraudulent and non-fraudulent financial records. They emphasized the role of feature selection in improving model performance by identifying the most relevant financial indicators. Their results demonstrated that selecting optimal features significantly enhances classification accuracy and reduces model complexity. K. Seeja, et al. [10] proposed a fraud detection model named FraudMiner based on frequent itemset mining techniques. Their study focused on identifying recurring patterns in transaction data that are associated with fraudulent behavior. By combining frequent pattern mining with classification methods such as KNN and SVM, they improved the detection of hidden relationships within the data. Their approach demonstrated that mining frequent patterns can effectively reveal fraud trends that are not easily captured by traditional ML models. C. Tyagi, et al. [11] analyzed various credit card fraud detection techniques and provided a comparative evaluation of different ML-based approaches. Their study focused on understanding the strengths and limitations of different algorithms in detecting fraudulent transactions. They highlighted that no single model is universally optimal and emphasized the importance of hybrid approaches that combine multiple techniques to improve detection accuracy and robustness.

3. PROPOSED SYSTEM

The architecture of a machine learning–based fraud detection system as illustrated in Fig. 1, in banking is designed to process large volumes of transactional data efficiently while

identifying suspicious activities in real time. The system begins with a data acquisition layer that gathers information from multiple sources such as transaction records, customer profiles, device information, and historical fraud logs. This raw data is then passed through a preprocessing module, where operations such as data cleaning, missing value handling, normalization, and feature engineering are performed to improve data quality and make it suitable for model training. Feature selection techniques may also be applied to reduce dimensionality and enhance model performance. The processed data is then used to train machine learning models, including supervised algorithms such as Decision Trees, Random Forest, Support Vector Machines, and Neural Networks, which learn patterns that distinguish legitimate transactions from fraudulent ones. In some systems, ensemble or hybrid models are employed to improve accuracy and robustness. Once trained, these models are deployed within a real-time inference layer that continuously evaluates incoming transactions. This layer performs rapid prediction and anomaly detection, ensuring minimal latency while maintaining high detection accuracy.

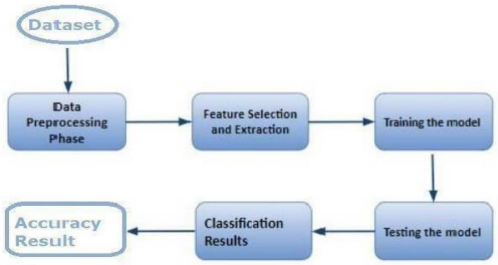

Fig. 1: System Architecture

A decision-making module is integrated to interpret model outputs and take appropriate actions, such as flagging suspicious transactions, generating alerts for security teams, or automatically blocking high-risk activities. Additionally, the system may include a feedback loop where newly identified fraud cases are fed back into the model to improve future predictions. To support real-world deployment, the architecture is designed with scalability and reliability in mind, often leveraging distributed computing and cloud-based infrastructure to handle high transaction throughput while ensuring consistent performance and timely fraud detection.

3.1 RANDOM FOREST

Online fraud transaction detection using the Random Forest model involves leveraging machine learning to identify fraudulent transactions based on patterns in historical transaction data. The model is trained on features like transaction amount, user behavior, time, location, and device used, with the target variable being whether the transaction is fraudulent or not. Random Forest works by creating multiple decision trees that analyze different subsets of the data, and the final prediction is made by aggregating the results from these trees. This ensemble approach helps reduce over fitting and improves accuracy. Key challenges include handling class imbalance, as fraudulent transactions are rare, and ensuring the model can adapt to evolving fraudulent patterns.

Analysing online frud data using a Random Forest model can provide insights into player performance, match outcomes, and team strategies. Random Forest is an ensemble learning method that uses multiple decision trees to make predictions, which can be particularly useful in sports analytics where there are many variables and relationships between them. For each bootstrapped dataset, a decision tree is built. This decision tree is a simple model that makes decisions based on feature splits.

Node splits: At each decision point, the tree splits the data based on which feature (e.g., batting average, team composition, match venue) results in the best division of the data. This is typically done by looking for the feature that minimizes

Classification (Match Outcome Prediction): If the task is to predict match outcomes (win or loss), each decision tree in the Random Forest will predict a result. The final prediction is made by taking a **majority vote** among all the trees. For example, if 60 trees predict a win and

Received: 02-04-2025 | Accepted: 12-05-2025 | Published: 19-05-2025 | www.ajaccm.com

40 trees predict a loss, the final prediction will be a "win."

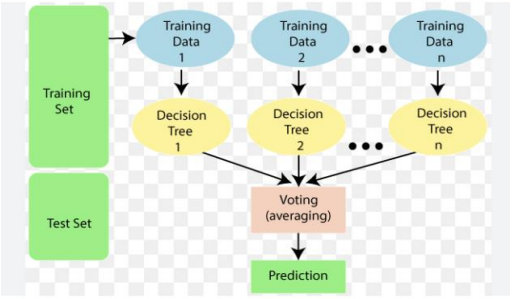

Fig. 2: Random Forest Structure

Regression (Player Performance Prediction): If the task is predicting continuous variables like player scores or wickets, the final prediction will be the average of all tree predictions. For example, if individual trees predict different scores, their average will be the predicted score.

4. CONCLUSION

The study presented an integrated machine learning framework for detecting fraudulent activities in banking systems by leveraging a publicly available dataset from the UCI repository. The dataset exhibited a strong class imbalance, with a significantly higher proportion of legitimate transactions compared to fraudulent ones, which was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) to balance the data distribution. The researchers implemented multiple models, including Logistic Regression and Random Forest, and observed certain challenges related to model performance and handling imbalanced data, particularly with the RF approach. Despite these challenges, the proposed model achieved a high-performance rate of 97.74%, demonstrating its effectiveness in identifying fraudulent transactions. Additionally, the analysis revealed that individuals in the age group of 19–25 showed a relatively higher tendency toward fraudulent activity, indicating the importance of incorporating demographic features into fraud detection systems. The combination of Logistic Regression and RF in an ensemble framework proved to be a robust solution for improving detection accuracy while effectively addressing data imbalance issues in banking applications.

REFERENCES

[1] S. M. Darwish, "An intelligent credit card fraud detection approach based on semantic fusion of two classifiers," (in English), Soft Computing, Article vol. 24, no. 2, pp. 1243-1253, Jan 2020.

[2] Eshghi and M. Kargari, "Introducing a new method for the fusion of fraud evidence in banking transactions with regards to uncertainty," (in English), Expert Systems with Applications, Article vol. 121, pp. 382-392, May 2019.

[3] S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime Prediction Using Spatio-Temporal Data," in Computing Science, Communication and Security, Singapore, 2020, pp. 277-289: Springer Singapore.

[4] M. Zamini and S. M. H. Hasheminejad, "A comprehensive survey of anomaly detection in banking, wireless sensor networks, social networks, and healthcare," (in English), Intelligent Decision Technologies-Netherlands, Article vol. 13, no. 2, pp. 229-270, 2019.

[5] Gonzalez-Carrasco, J. L. Jimenez-Marquez, J. L. Lopez-Cuadrado, and B. Ruiz-Mezcua, "Automatic detection of relationships between banking operations using machine learning," (in English), Information Sciences, Article vol. 485, pp. 319-346, Jun 2019.

[6] M. Pohoretskyi, D. Serhieieva, and Z. Toporetska, "The proof of the event of a financial resources fraud in the banking sector: problematic issues," (in English), Financial and Credit Activity-Problems of Theory and Practice, Article vol. 1, no. 28, pp. 36-45, 2019.

[7] Kalae, U. K. (2021). Creating tailored

Power Apps to optimize data collection and reporting across multiple platforms. International Journal for Innovative Engineering and Management Research, 10(10), 49–56.

[8] Noor et al., "Performance analysis of a surveillance system to detect and track vehicles using Haar cascaded classifiers and optical flow method," 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, 2017, pp. 258-263.

[9] Galina Baader and Helmut Krcmar. Reducing false positives in fraud detection: Combining the red flag approach with process mining. International Journal of Accounting Information Systems, 2018.

[10] Saai Reddy Purmani, S. (2023). The Transformation of IT Leadership in Business Organizations: Shifting from Technical Supervision to Strategic Empowerment. JOURNAL OF ADVANCE AND FUTURE RESEARCH, 1(5). https://doi.org/10.56975/jaafr.v1i5.503885

[11] Ravisankar P, Ravi V, Raghava Rao G, and Bose, Detection of financial statement fraud and feature selection using data mining techniques, Elsevier, Decision Support Systems Volume 50, Issue 2, p491-500 (2011) SVM

[12] Seeja, and M. Zareapoor, “FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining,” The Scientific World Journal,2014, pp. 1-10. KNN, SVM

[13] C. Tyagi, P. Parwekar, P. Singh, and K. Natla, “Analysis of Credit Card FraudDetection Techniques,” Solid State Technology, vol. 63, no. 6, 2020, pp. 18057- 18069. Credit card faud