© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715333

Enhanced Fraud Detection in Financial Transactions using Machine Learning

PRIYANKA M[1], M MONIKA[2], NAGALAKSHMI RG[3], M LAHARI REDDY[4], MOHITH TN[5]
[1, 2, 3, 4, 5]Computer Science and Engineering, CMR University

***Abstract-** As digital banking, online payments, and e-commerce rapidly evolve, financial transactions are increasingly vulnerable to fraud. Traditional fraud detection systems often produce high false positives and fail to adapt to changing tactics. Our solution employs advanced supervised learning algorithms like Random Forest, XGBoost, and Logistic Regression to accurately analyse transactional data and quickly identify fraudulent activity. By integrating sophisticated data preprocessing, feature engineering, and techniques to manage class imbalances, we significantly improve prediction accuracy. Our results show that these ensemble-based models outperform conventional classifiers, delivering enhanced fraud detection with fewer false alarms, ultimately strengthening transaction security and reducing financial losses in modern digital banking.*

***Keywords:** Financial Fraud Detection, Machine Learning, XGBoost, Random Forest, Class Imbalance, Digital Transactions*

I. INTRODUCTION

Online and card-based transactions are increasingly becoming the norm as the world embraces digital financial services. This shift brings with it undeniable benefits, such as convenience and efficiency for users. However, it also opens the door to significant security challenges, with financial fraud emerging as a substantial threat. Issues like unauthorised transactions, identity theft, account takeovers, and manipulation of online payments can lead to severe financial losses and can deeply undermine user trust in digital platforms.

Historically, fraud detection systems have relied on established rules and manual checks to identify suspicious activities. While these systems were somewhat effective in the past, they struggle to keep pace with the evolving nature of fraud. They often generate numerous false positives, inconveniencing legitimate users, and are vulnerable to the adaptive tactics employed by fraudsters. As a result, traditional methods become less effective over time, unable to accurately capture the complexity of modern fraud schemes.

In response to these challenges, machine learning (ML) emerges as a powerful and adaptive solution. By continuously analysing past transaction data, ML can detect anomalies in real time, learning and evolving as new data comes in. This capability enables ML-based fraud detection systems to handle extensive datasets and intricate relationships between features, ultimately improving the accuracy of fraud detection efforts. This research aims to develop an ML framework that can robustly and accurately identify fraudulent financial transactions, establishing a new standard for security in digitalfinance.

Moreover, the quick analysis of vast transaction datasets is a significant advantage of machine learning over manual or rule- based approaches. By enabling automated decisions, ML reduces the reliance on human intervention, allowing financial institutions to keep up with rapidly changing fraud tactics. This flexibility is particularly vital in fast-paced financial environments where fraud schemes evolve continuously.

Furthermore, ML-based systems are designed for real-time and scalable deployment within modern digital payment infrastructures. By incorporating predictive analytics alongside transaction monitoring systems, financial institutions can strengthen customer confidence, proactively thwart fraud attempts, and ensure that digital transactions remain secure. Given these significant advantages, machine learning stands out as a beacon of hope for the future of fraud detection systems, promising enhanced security and trust in digital financial interactions.

II. LITERATURE REVIEW

Wang et al. (2020) [1] proposed a Learning Automatic Windows (LAW) approach for online payment fraud detection. Their method dynamically adjusts time windows based on transaction behavior to capture evolving fraud patterns. The study showed improved detection of temporal fraud activities; however, it required continuous model updating and higher computational cost for real-time processing.

Albashrawi et al. (2016) [2] explored several classification algorithms, including Random Forest, Decision Trees, and Logistic Regression, specifically utilising labelled transaction datasets for supervised learning. They found that ensemble models like Random Forest provided better fraud detection accuracy compared to single classifiers. A major challenge they noted was class imbalance, which negatively affected the recall of minority fraud cases.

Liu et al. (2025) [3] conducted a comparative study of CatBoost, XGBoost, and LightGBM for credit card fraud detection. Their results showed that XGBoost achieved superior performance in handling large-scale and imbalanced datasets. However, the study highlighted the need for careful hyperparameter tuning to achieve optimal results.

Jayakumar et al. (2013) [4] showcased the prevalent use of outlier detection techniques in handling extremely unbalanced datasets for financial fraud detection, while Jans et al. (2011) categorised financial fraud into four main types, with a focus on transaction fraud related to mobile payments.

Faraji (2022) [5] reviewed various machine learning techniques such as Logistic Regression, Random Forest, and XGBoost for credit card fraud detection. The study highlighted that ensemble methods outperform traditional models in terms of accuracy and robustness. However, challenges such as data imbalance and feature selection remain significant limitations.

Ali et al. (2022) [6] explored machine learning-based fraud detection systems focusing on anomaly detection and classification techniques. Their findings emphasized the importance of hybrid approaches combining supervised and unsupervised learning. While the approach improved detection rates, it increased system complexity and computational requirements.

Brown et al. (2024) [7] focused on anomaly detection techniques for financial transactions using advanced machine learning models. Their research highlighted the effectiveness of unsupervised learning in identifying unknown fraud patterns. However, the lack of labelled data limited the interpretability and accuracy of the models.

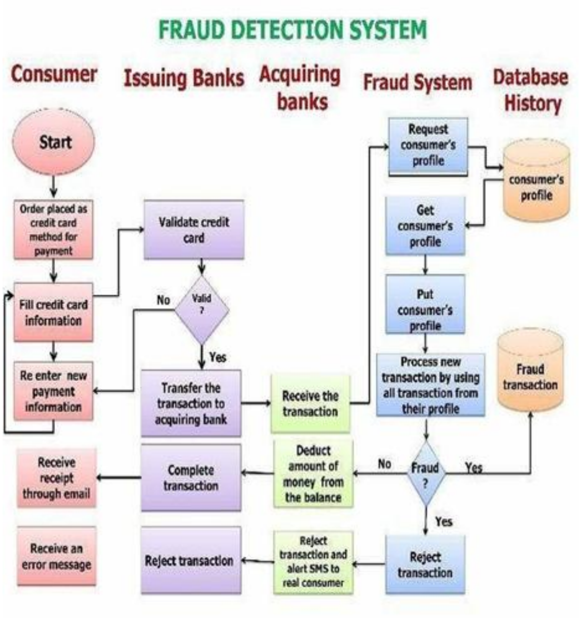
Shin et al. (2013) [8] developed a cost-sensitive decision tree model aimed at reducing financial losses due to fraud. By incorporating misclassification costs into the learning process, their approach prioritised detecting high-risk fraudulent transactions. This method improved cost efficiency but increased model complexity and required domain-specific cost estimation.

Sorournejad et al. (2016) [9] presented a comprehensive survey on credit card fraud detection techniques, focusing on data- driven approaches. Their work analysed various machine learning models and highlighted challenges such as data imbalance, feature selection, and evolving fraud patterns. The study provided a strong foundation for future research but lacked implementation details for real-time systems.

Wedge et al. (2018) [10] explored automated feature engineering to improve fraud prediction accuracy. Their approach reduced false positives by generating meaningful features from raw transaction data. The study demonstrated significant improvements in model performance; however, it required high computational power and sophisticated feature generation techniques.

III. METHODOLOGY

3.1 SYSTEM ARCHITECTURE


FRAUD DETECTION SYSTEM
Consumer
Issuing Banks
Acquiring banks
Fraud System
Database History
Start
Order placed as credit card method for payment
Fill credit card information
Re enter new payment information
Validate credit card
Valid ?
No
Yes
Transfer the transaction to acquiring bank
Receive the transaction
Request consumer's profile
consumer's profile
Get consumer's profile
Put consumer's profile
Process new transaction by using all transaction from their profile
Fraud transaction
Fraud ?
No
Yes
Yes
Deduct amount of money from the balance
Complete transaction
Receive receipt through email
Reject transaction
Reject transaction and alert SMS to real consumer
Reject transaction
Receive an error message

The proposed fraud detection system is designed to integrate machine learning techniques within a structured financial transaction architecture to enable accurate and real-time fraud detection. The system combines traditional banking workflow with advanced models such as XGBoost and Logistic Regression to enhance security in card-based transactions. This architecture ensures that every transaction is validated, processed, and analyzed before reaching a final decision.

The process begins at the consumer layer, where the user initiates a transaction through the developed web application by entering credit or debit card details. The system captures important transaction attributes such as transaction amount, time, and user-related information. These inputs serve as the primary data for both validation and fraud detection processes.

Once the transaction is initiated, the issuing bank verifies the authenticity of the card details. If the entered information is invalid, the transaction is immediately rejected, ensuring security at an early stage. If the card details are valid, the transaction request is forwarded to the acquiring bank for further processing, maintaining a secure and structured flow of operations.

At the acquiring bank level, the transaction is processed and the amount is temporarily deducted or reserved. The transaction details are then forwarded to the fraud detection module, which acts as the core component of the proposed system. This stage ensures that every transaction undergoes an additional layer of security before final approval.

The fraud detection module applies machine learning algorithms, where Logistic Regression is used as a baseline model and XGBoost is used as the primary model for prediction. The system analyzes various features such as transaction patterns, frequency, and historical user behavior to classify transactions as fraudulent or legitimate. XGBoost significantly improves performance by handling imbalanced data and capturing complex patterns more effectively than traditional methods.

The system also maintains a database that stores transaction history, user profiles, and fraud labels. This data is continuously used to train and improve the machine learning models, making the system adaptive to evolving fraud techniques. The integration of historical data enhances the model's ability to detect anomalies with higher accuracy.

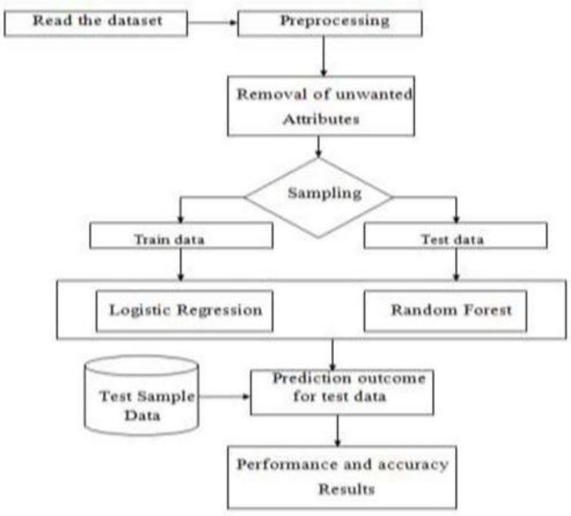
Finally, based on the prediction results, the system makes a decision to either approve or reject the transaction. If fraud is detected, the transaction is blocked and an alert is generated. If the transaction is legitimate, it is successfully completed and confirmation is provided to the user through the web application. This real-time decision- making capability ensures both security and user convenience

Furthermore, the system architecture is designed with a modular approach, where each component operates independently while remaining well-integrated with the overall system. This modularity allows easy scalability and flexibility for future enhancements, such as incorporating Explainable AI (XAI) techniques to improve transparency in fraud detection decisions. The architecture also ensures efficient data handling and fast processing, which is essential for real-time transaction analysis.

In addition, the system incorporates secure data transmission and storage mechanisms to protect sensitive financial information. The interaction between the frontend user interface and backend machine learning model is optimized to minimize latency and provide quick responses. This design ensures that the system can handle large volumes of transactions without performance degradation, making it suitable for real-world deployment in banking and financial applications.

Overall, the system architecture provides a strong and scalable foundation that supports accurate fraud detection, efficient processing, and seamless integration of machine learning models into a real-time application environment.

3.2 DATA FLOW



The data flow of the proposed fraud detection system begins with reading the transaction dataset, which contains historical records of financial transactions. This dataset includes both legitimate and fraudulent transaction labels, which are essential for training supervised machine learning models. The quality and structure of this dataset play a crucial role in determining the effectiveness of the system.

Once the dataset is loaded, preprocessing is performed to clean and prepare the data for analysis. This step includes handling missing values, removing duplicates, and normalizing the data to ensure consistency. Proper preprocessing improves the efficiency and accuracy of the machine learning models by eliminating noise and irrelevant variations in the data.

Following preprocessing, unwanted or irrelevant attributes are removed from the dataset. This feature selection step helps in reducing dimensionality and improving model performance. By focusing only on important features such as transaction amount, time, and behavioral patterns, the system becomes more efficient and avoids overfitting.

The processed dataset is then subjected to sampling techniques to handle class imbalance, which is a common issue in fraud detection. Since fraudulent transactions are much fewer compared to legitimate ones, techniques such as resampling are applied to balance the dataset. The data is then divided into training and testing sets to evaluate the model's performance effectively.

In the training phase, machine learning models are applied. In this project, Logistic Regression is used as a baseline model, while XGBoost is used as the primary model for improved performance. These models learn patterns from the training data and build a predictive model capable of distinguishing between fraudulent and genuine transactions.

After training, the models are tested using unseen test data. The system generates prediction outcomes for each transaction, classifying them as fraud or non-fraud. This step simulates real- world scenarios where the model must make decisions on new incoming transactions.

Furthermore, the data flow within the system is structured to ensure smooth and efficient movement of data across all processing stages. Each stage of the data pipeline is responsible for validating, transforming, and preparing the data before passing it to the next stage. This ensures data consistency and reduces the chances of errors during processing.

The system is designed to support continuous and real-time data flow, where incoming transactions are processed instantly without delays. The seamless communication between the frontend interface, preprocessing module, and machine learning model enables quick classification of transactions. This efficient flow of data enhances the system's ability to detect fraud in real time and respond immediately.

Additionally, the structured data flow improves system reliability and maintainability by clearly defining how data moves through different components. It ensures that each module performs its function effectively while contributing to the overall goal of accurate and timely fraud detection.

Finally, the performance of the models is evaluated using metrics such as accuracy, precision, recall, and F1-score. The results show that XGBoost provides better performance compared to Logistic Regression due to its ability to handle complex patterns and imbalanced data. This data flow ensures an efficient and accurate fraud detection process suitable for real-time applications.

IV. RESULT AND ANALYSIS

The performance of the proposed fraud detection system was evaluated using multiple supervised machine learning algorithms, including XGBoost, Random Forest, and Logistic Regression. The models were trained on a highly imbalanced financial transaction dataset and tested to classify transactions as either legitimate or fraudulent. The evaluation focuses on model accuracy, fraud detection capability, and real-time applicability.

Exploratory Data Analysis: To better understand the dataset and identify important features influencing fraud detection, exploratory data analysis (EDA) was performed. The analysis focused on identifying features that show positive and negative correlations with fraudulent transactions.

Positive Correlation Analysis:
The figure below illustrates features that exhibit a positive correlation with fraudulent transactions. It can be observed that attributes such as transaction amount and balance differences tend to have higher values in fraudulent cases compared to legitimate ones. The box plots highlight a wider spread and presence of outliers in fraudulent transactions, indicating irregular and suspicious activity patterns. These variations help the machine learning model to effectively distinguish between genuine and fraudulent transactions.

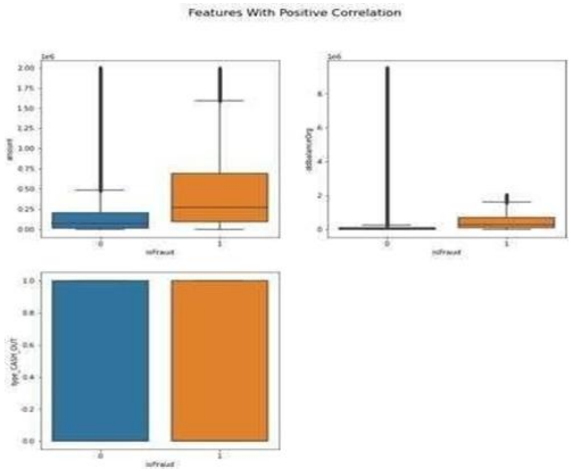


Fig. 1: Features showing Positive correlation

Negative Correlation Analysis:
The figure below presents features that demonstrate a negative correlation with fraud occurrence. It is evident that certain attributes maintain lower or more stable values in fraudulent transactions compared to non-fraudulent ones. This inverse relationship provides additional insights for the model, enabling it to recognize patterns where legitimate transactions behave consistently, while fraudulent ones deviate significantly. Such features play a crucial role in improving the robustness of the classification model.

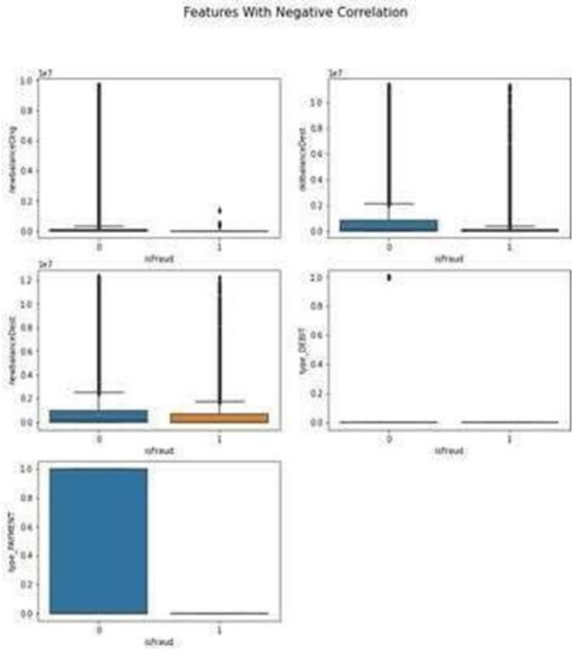


Fig. 2: Features showing Negative correlation Model Performance Analysis:

The classification performance of the models was analysed based on their ability to correctly identify fraudulent transactions while minimising false positives.

Among the evaluated models, the XGBoost algorithm demonstrated superior performance due to its ability to handle large datasets and capture complex non-linear relationships between features. It achieved higher accuracy and recall compared to other models, ensuring that most fraudulent transactions were successfully detected. This is particularly important in financial systems, where missing a fraudulent transaction can lead to significant losses
.

Random Forest also provided strong performance by leveraging ensemble learning, but it was slightly less efficient compared to XGBoost in terms of computational speed and detection capability. Logistic Regression, while simple and computationally efficient, showed lower performance due to its limitation in modelling complex patterns in the data.

Real-Time System Evaluation:
The trained model was successfully integrated into a web-based transaction system designed to simulate real-world banking operations. The system processes transaction details in real time and classifies them as either legitimate or fraudulent. Upon detecting suspicious activity, the system generates instant alerts, thereby enhancing the responsiveness and security of the system.

Heatmap Analysis:
The heatmap of all features in the training dataset was generated to visualize the correlation between different variables and their influence on fraud detection. This graphical representation helps in identifying the strength and direction of relationships among features such as transaction amount, account balances, transaction type, and time step. From the heatmap, it is observed that certain features show strong positive or negative correlations with the target variable, indicating their significance in distinguishing fraudulent transactions from legitimate ones.

In particular, features related to transaction amount and balance differences demonstrate noticeable correlation patterns, suggesting that abnormal changes in account balances are strong indicators of fraudulent behaviour. On the other hand, some features exhibit weak or negligible correlation, implying that they contribute less to the prediction process. This analysis helps in effective feature selection by prioritising the most relevant attributes and eliminating redundant ones, thereby improving model performance.

Furthermore, the heatmap provides insights into multicollinearity among features, where highly correlated independent variables may affect model stability. By identifying such relationships, appropriate preprocessing steps can be applied to reduce redundancy and enhance model efficiency. Overall, the heatmap analysis plays a crucial role in understanding the dataset structure and supports the development of a more accurate and robust fraud detection model.

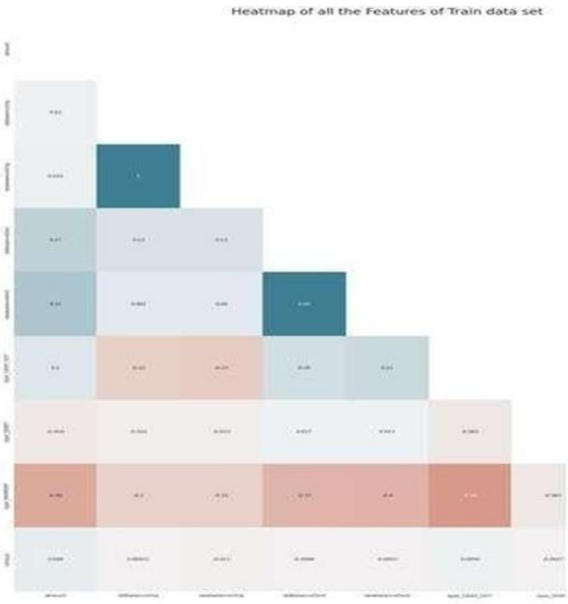


Fig. 3: Heatmap Features of Train dataset

The results clearly indicate that the proposed fraud detection system is effective in identifying fraudulent transactions with high accuracy. The combination of data preprocessing, feature analysis, and advanced machine learning techniques significantly improves detection performance. Additionally, the integration of the model into a real-time application demonstrates its practical usability in real- world financial environments. The system is scalable, efficient, and capable of adapting to evolving fraud patterns, making it a reliable solution for modern digital payment systems.

V. CONCLUSION

This project successfully developed a machine learning integrated web-based fraud detection system. Through a website, the system enables users to conduct financial transactions and instantly ascertains whether the transaction is legitimate or fraudulent. The system achieves dependable fraud detection performance by integrating supervised learning algorithms with appropriate data preprocessing and handling of class imbalances.

XGBoost outperformed Random Forest and Logistic Regression among the models that were used. The suggested system demonstrates how well web technologies and machine learning can be combined to improve transaction security. This strategy can be applied to actual banking applications to lower financial fraud and boost user confidence

V1. FUTURE SCOPE

There are colourful ways to ameliorate the suggested fraud discovery system, indeed more. To enhance the identification of intricate fraud patterns, sophisticated styles like deep learning models can be incorporated. Planting the system in real- time banking and payment surroundings allows for real- time perpetration, allowing for the instant discovery of fraud during deals. druggies can cover deals and get immediate fraud cautions by extending the system to mobile operations. Resolvable AI styles can also be used to give precise explanations for fraud prognostications, boosting openness and confidence. Pall- grounded deployment for lesser scalability and integration with colourful banking and payment platforms to accommodate a wider stoner base are implicit, unborn advancements.

REFERENCES

[1] E. Ngai et al., Decision Support Systems 50, 2011, 559 – 569, The operation of Data Mining ways in Financial Fraud Detection: A Bracket Framework and an Academic Review of Literature

[2] Albashrawi et al., in Journal of Data Science 14(2016), 553- 570, Detecting Financial Fraud Using Data Mining Ways: A Decade Review from 2004 to 2015.

[3] Synthetic fiscal Datasets for Fraud Detection, TESTIMON atNTNUwasattainedfrom https//www.kaggle.com/ntnutestimon/paysim1

[4] A Novel Approach to Clustering Utilising Multivariate Outlier Identification by Jayakumar et al. Data Science Journal 11(2013) 69 – 84

[5] Jans et al., Expert Systems with Applications 2011; 38 13351 – 13359, A Business Process Mining operation for Internal Transaction Fraud Mitigation

[6] Phua et al., Classifying Skewed Data in Fraud Detection: A Minority Report. 2004; 6 50 – 59 in ACM SIGKDD studies Newsletter.

[7] Dharwa et al., A Grounded-on Data Mining with Hybrid Approach, International Journal of Computer Operations, 2011; 16 18 – 25.

[8] A cost-sensitive decision tree system for fraud discovery was developed by Shin et al. and published in Expert Systems with Applications in 2013; 40 5916 – 5923.

[9] Sorourenejad et al., A Survey of Styles for Detecting Credit Card Fraud: An Approach concentrated on Data and ways, 2016

[10] Wedge et al.," Automated Feature Engineering Solves the False Cons Problem in Fraud Prediction," Machine Learning and Knowledge Discovery in Databases, pp. 372 – 388, 2018