Grenze International Journal of Engineering and Technology, January Issue

Online Payment Fraud Detection using ML and its Interpretation

Roma kudale[1], Bharath T[2] and Aman Abhishek[3]
[1-3]Dept. of Artificial Intelligence and Data Science, CMR Institute of Technology, Bengaluru, India
Email: roma.k@cmrit.ac.in, bht21ainds@cmrit.ac.in, amab21ainds@cmrit.ac.in

***Abstract*— Online payment fraud poses a critical threat to the digital financial ecosystem, demanding advanced detection mechanisms to minimize losses. Traditional rule-based approaches often fail to adapt to rapidly evolving fraud strategies, highlighting the potential of machine-learning models. This study evaluates the Random Forest classifier for identifying fraudulent transactions using a dataset of 71,759 records, which are carefully preprocessed to manage class imbalance and remove irrelevant features. The performance of the model was measured using metrics such as precision, recall, and AUC-ROC, with the results showing 98.75% accuracy, 91.07% precision, and 98.72% recall. To improve model transparency, SHAP values were utilized to explain feature contributions. Furthermore, a rule-based chatbot is integrated to assist users. The results affirm the model's effectiveness and reliability in fraud detection while minimizing false positives.**

***Index Terms*— Fraud Detection, Random Forest, SHAP, Online Payments, Machine Learning, Financial Security.**

I. INTRODUCTION

As digital payment systems grow in popularity, the threats posed by online payment fraud significantly undermine financial security. Fraudsters exploit the vulnerabilities in these systems, leading to financial losses and decreased user trust. Traditional detection frameworks that rely on fixed rules often struggle to adapt to the evolving fraudulent tactics. By contrast, machine learning provides a more dynamic and scalable approach for identifying fraud. Among the available models, Random Forest stands out for its capability to manage high-dimensional and imbalanced datasets effectively. This paper presents the design and implementation of a fraud detection system using a Random Forest classifier enhanced with SHAP value-based explanations for interpretability and a chatbot inter-face for user support and engagement.

II. LITERATURE SURVEY

Fraud detection in online payment systems has been an area of extensive research, owing to the increasing sophistication of fraudulent activities. Various Machine learning algorithms have been explored to enhance fraud-detection accuracy, optimize false-positive rates, and improve real-time detection capabilities. Several studies have highlighted the importance of ensemble learning methods, feature selection techniques, and data pre-processing strategies for fraud detection. Recent research has demonstrated the effectiveness of machine learning models, such as Random Forest, Gradient Boosting, and Logistic Regression, in fraud detection.

Grenze ID: 01.GIJET.12.1.127

A comparative study evaluating these algorithms on credit card fraud datasets found that ensemble methods perform better than individual classifiers by effectively capturing nonlinear fraud patterns [1]. Similarly, a systematic review of fraud-detection techniques indicated that ensemble learning models, including Random Forest and Gradient Boosting, have shown high accuracy in detecting fraudulent transactions while minimizing false positives [2]. Another study emphasized that class imbalance remains a significant challenge, where fraudulent transactions constitute a small fraction of the dataset, requiring advanced techniques, such as oversampling, undersampling, and cost-sensitive learning [3]. One of the key factors in improving fraud detection accuracy is feature selection and engineering. Research has shown that domain-specific feature engineering plays a crucial role in identifying fraudulent transactions [4]. In cases where fraudulent transactions are rare, imbalanced handling techniques such as SMOTE, Random Undersampling, and cost-sensitive learning improve the classifier's ability to differentiate between fraudulent and genuine transactions [5]. A study comparing various imbalanced classification techniques concluded that handling data imbalance effectively enhanc-es fraud detection performance, as models trained on imbalanced data often misclassify fraudulent transactions ow-ing to their rarity [6]. Although machine learning models demonstrate high predictive power, their lack of transparen-cy poses challenges in real-world adoption. Explainable AI (XAI) techniques such as SHAP values and LIME have been employed to make fraud-detection models more interpretable. A study on explainable models for Medicare fraud detection highlighted that integrating SHAP values helps financial institutions understand why a transaction is classified as fraudulent [7]. Moreover, real-time fraud detection has been an area of focus where machine learning models must provide instant decisions with justifications for end-users and analysts [8]. Fraud detection models should be capable of processing transactions in real-time to prevent unauthorized activities. Research has indicated that integrating machine learning models with real-time monitoring systems significantly improves fraud detection efficiency [9]. Additionally, an integrated multistage ensemble learning approach was proposed to combine the strengths of multiple classifiers, resulting in enhanced fraud-detection rates [10]. Based on the reviewed research, it is evident that Random Forest and ensemble-based techniques remain among the most effective approaches for fraud detection because of their robustness against noisy data and adaptability to complex fraud patterns. However, the challenges of class imbalance, explainability, and real-time detection require further investigation. The current study aims to address these gaps by implementing a Random Forest-based fraud-detection system with SHAP value-based explanations and a chatbot for user support.

III. METHODOLOGY

The methodology adopted in this project follows a structured approach consisting of data pre-processing, model training, transaction processing, fraud explanation, and user interaction. The system was designed to efficiently de-tect fraudulent transactions while maintaining model transparency and user accessibility through explainability tech-niques and chatbot assistance.

A. Data Preprocessing

The dataset used for training and testing consists of 71,759 transaction records with features such as step (time), transaction type, amount, origin balance, and destination balance. Because real-world transaction data often contain inconsistencies, missing values are handled to ensure that the dataset is clean and reliable. Categorical variables, such as the transaction type, were converted into numerical representations using label encoding to make them suitable for machine learning models. Since financial transaction values can vary significantly, the numerical features were standardized using StandardScaler, ensuring that the model was not influenced by large numerical differences.

One of the primary challenges in fraud detection is a severe class imbalance, as fraudulent transactions make up only a small percentage of total transactions. If left unaddressed, machine-learning models may become biased to-ward the majority class, leading to poor fraud-detection performance. To overcome this, Random Undersampling is applied, reducing the size of the non-fraudulent class while maintaining essential fraud patterns. This balancing tech-nique prevents the model from being overwhelmed by non-fraudulent cases, thereby ensuring that it can effectively learn the patterns associated with fraudulent transactions.

B. Model Training

For fraud classification, the Random Forest Classifier was selected because of its ability to handle large datasets efficiently, its robustness against overfitting, and its capability to rank feature importance. The dataset was split into 80% training and 20% testing using stratified K-fold cross-validation, ensuring that both fraudulent and non-fraudulent cases were well distributed across the training and testing sets. To further optimize the model, hyperpa-rameter tuning was conducted using GridSearchCV. Key parameters such as the number of estimators (trees), maxi-mum tree depth, and minimum samples per split were fine-tuned to maximize classification accuracy. Additionally, because fraud detection requires minimizing false negatives (i.e., failing to detect fraudulent transactions), class weights were adjusted to a 1:8 ratio for non-fraud and fraud classes. This ensured that the model prioritized fraudu-lent cases without excessively increasing the number of false positives. The trained model was evaluated using mul-tiple performance metrics including accuracy, precision, recall, F1-score, and AUC-ROC. The final model achieved 98.75% accuracy, 91.07% precision, and 98.72% recall, demonstrating its effectiveness in detecting fraudulent transactions, while minimizing false alarms.

C. Transaction Processing

Once trained, the model was deployed to classify transactions in real-time. Whenever a user initiates a transac-tion, the system extracts the relevant transaction features and preprocesses them to match the input format of the model. The trained Random Forest model then predicts the likelihood of fraud, providing a binary classification in which transactions are labelled as either fraudulent or non-fraudulent. To enhance fraud detection accuracy, the system applies a threshold of 0.65 to fraud probability scores. Transactions with fraud probability exceeding this threshold are flagged as fraudulent. This threshold selection is based on optimizing the trade-off between false posi-tives (legitimate transactions flagged as fraud) and false negatives (missed fraudulent transactions).

D. Fraud Explanation Using SHAP Values

The A significant challenge in fraud detection systems is the lack of interpretability. Many machine learning models function as "black boxes," making it difficult for users to understand why a transaction is classified as fraudulent. To address this, SHAP (SHapley Additive exPlanations) values were integrated into the system to provide a detailed breakdown of how each feature influenced the fraud prediction. For each transaction, SHAP values de-termine which factors contributed most to the fraud classification. Features such as transaction amount, balance dif-ferences, and transaction type are analysed to generate an impact score. These values are dynamically displayed in a feature impact report, providing users with clear insights into why their transaction was flagged. The integration of SHAP values ensures model transparency, making it easier for financial analysts and customers to trust and under-stand fraud predictions.

IV. SYSTEM ARCHITECTURE

As shown in Fig. 1, proposed system consists of five main components:

- Data Pre-processing: Cleanses and transforms raw transaction data.
- Model Training: Uses Random Forest to classify transactions as fraudulent or non- fraudulent.
- Transaction Processing: Processes real-time transactions and applies fraud detection.
- Fraud Explanation: Uses SHAP values to interpret fraudulent predictions.
- User Interaction: Provides transaction insights via a chatbot.

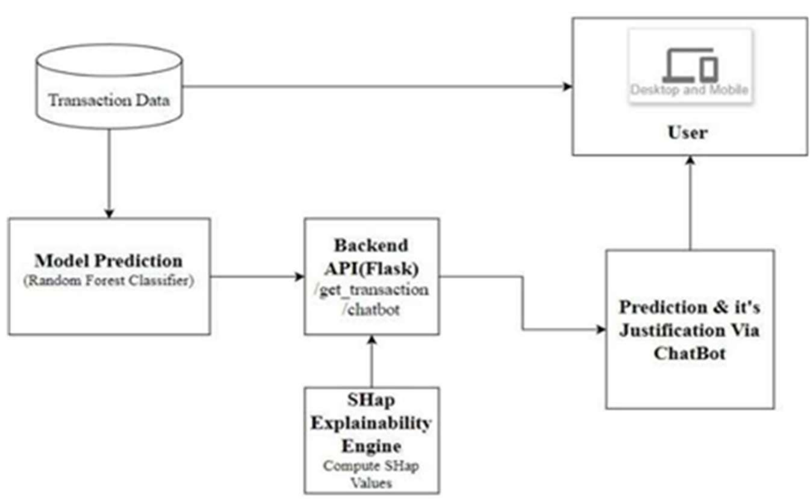


Fig. 1. System Architecture

V. SETUP AND PERFORMANCE

A. System Setup

The implementation of the fraud detection system was carried out in a Python-based environment using Flask for the backend and JavaScript with HTML/CSS for the frontend. The machine learning model was developed using the scikit-learn library, with additional support from SHAP for explainability and NumPy, Pandas, and Matplotlib for data handling and visualization. The entire system was hosted and tested on a local development server, ensuring smooth integration between the model, transaction processing, and chatbot functionality.

B. Performance Evaluation

The fraud detection model was evaluated based on multiple performance metrics to assess its accuracy, robustness, and effectiveness in identifying fraudulent transactions. Given the highly imbalanced dataset, where fraudulent transactions account for only a small percentage, special emphasis was placed on precision, recall, and AUC-ROC scores, rather than just accuracy.

C. Confusion Matrix Analysis

The confusion matrix for the trained Random Forest model was as follows:

TABLE I. CONFUSION MATRIX

	Predicted Non-Fraud	Predicted Fraud
Actual Non-Fraud	12550	159
Actual Fraud	21	1621

As shown in Table 1, the results indicate that the model successfully identifies fraudulent transactions with a low false-negative rate (21 cases of missed fraud out of 1,642 total fraud cases). However, there are 159 false positives, meaning some legitimate transactions were incorrectly flagged as fraud.

D. Performance Metrics

The key performance metrics of the model are as follows: Accuracy: 98.75% measures the overall correctness of fraud and non-fraud predictions. Precision: 91.07% – Ensures that flagged fraud cases are truly fraudulent, thereby reducing unnecessary alerts. Recall: 98.72% – Indicates how well the model detects actual fraud, which is crucial for minimizing financial losses. F1-Score: 94.75% – Provides a balanced measure of precision and recall. The AUC- ROC Score of 0.9985 shows excellent model performance in distinguishing fraud from non-fraud. These results demonstrate that the model is highly effective in detecting fraudulent transactions while maintaining high precision and recall values, making it a reliable tool for fraud prevention.

E. Real-Time Performance & Deployment Considerations

Transaction Processing Time: Few seconds per transaction, ensuring near-instant fraud detection. Scalability: The model can handle large transaction volumes efficiently with minimal computational overheads. User Experience: The integrated chatbot provides real- time explanations, improving user trust and interaction with the system.

F. Results and Discussion

The fraud detection system developed in this study demonstrates high accuracy and reliability in identifying fraudu-lent transactions. By leveraging the Random Forest Classifier, the model effectively detects fraudulent activities while ensuring interpretability through SHAP- based feature analysis. One of the most significant contributions of this sys-tem is its ability to explain model decisions, offering insights into why a particular transaction is classified as fraudu-lent. This level of transparency is especially vital in financial systems where accountability and clarity are essential for trust and compliance.

The system is designed to process transactions in real time and respond immediately with classification and analysis. As illustrated in Fig.4, every transaction is clearly labelled as either Fraudulent or Non-Fraudulent after processing. This immediate feedback enables faster decision- making and helps prevent financial loss. To further enhance user engagement, the system includes an assistive chatbot interface, shown in Fig.2, which allows users to ask predefined questions regarding fraud decisions, SHAP- based interpretability, and feature impact. This integration of natural language interaction offers a user-friendly explanation of model behaviour.

When a transaction is flagged as suspicious, the system takes prompt action by notifying the user through a clear alert mechanism. Fig.3 illustrates the ROC Curve of the proposed model, highlighting its excellent discriminatory power with an AUC close to 1.0, indicating high effectiveness in distinguishing between fraudulent and non-fraudulent transactions. Such real-time rejection of fraudulent payments increases user awareness and confidence, while reducing potential financial threats. The integration of SHAP-based explanations not only helps justify model predictions but also supports informed decisions by highlighting the top contributing features, such as transaction amount, balance disparities, and transaction type. The chatbot's ability to relay these SHAP insights in a simple lan-guage further enhances its interpretability. This combination of technical robustness and user-centred design makes the proposed solution both effective and accessible.

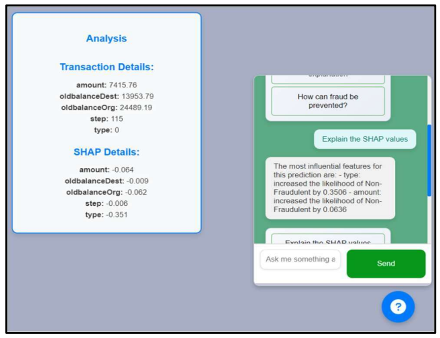
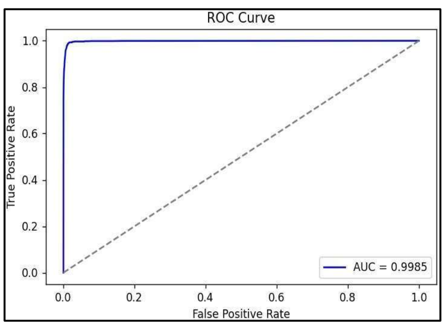


Fig. 2. Assistive chatbot explaining fraud detection



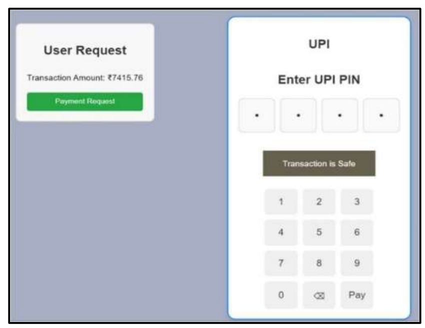


Fig. 3. ROC curve of Model   Fig. 4. Real-time classification of a transaction as fraudulent or non-fraudulent

Fig. 2 shows how assistive chatbot is helping to explain fraud detection. It gives transaction details and SHAP details. Fig. 3 shows Area Under the Curve as 0.99. ROC curve is a curve between True Positive rate Vs False Positive rate. Fig. 4. real-time classification of a transaction as fraudulent or non-fraudulent

VI. Conclusion

In this study, we presented a robust system for Online Payment Fraud Detection utilizing machine learning techniques, with a primary focus on a Random Forest classifier. The proposed solution successfully detects fraudulent transactions by analyzing key features, such as transaction amounts, user behavior patterns, and temporal factors. To enhance the interpretability of the model's predictions, we incorporated SHAP-based explanations, offering transpar-ency in the decision-making process. Additionally, the rule-based chatbot integrated into the system provides users with clear and contextual fraud alerts, further improving the user experience and understanding of the system's pre-dictions.

To optimize performance, we employed several techniques, including class balancing, hyperparameter tuning, and the application of Explainable AI (XAI). These improvements have ensured that the model is efficient and accessible in terms of understanding its reasoning. The system's ability to distinguish between legitimate and fraudulent transac-tions demonstrates its potential to effectively mitigate fraud risks in real-world online payment systems.

Future Work

While the current implementation of the fraud detection system is effective, there are several avenues for future de-velopment that could further enhance its capabilities and performance.

One key direction is the advancement of real-time fraud detection, which would allow the system to process transac-tions as they occur and provide immediate alerts, potentially preventing fraud before it affects users or merchants. This improvement significantly enhances the practical utility of the system in live environments.
Additionally, expanding the range of features incorporated into the model could provide a more comprehensive analysis of transaction data. Features such as geographical data, IP address analysis, and device fingerprinting can help identify fraudulent activities more accurately by considering broader user behavior patterns.
Exploring the integration of more sophisticated deep learning models could also be a promising avenue. These mod-els, particularly neural networks, are well-suited for detecting complex, non-linear relationships in large datasets, which might enhance the model's accuracy and robustness against evolving fraud tactics.
In terms of scalability, the system could be optimized to handle large volumes of data by adopting cloud-based de-ployment with horizontal scaling, allowing it to process millions of transactions per day without sacrificing perfor-mance. This would be crucial for systems supporting global or high-volume payment platforms.
Moreover, the incorporation of adaptive learning mechanisms would enable the model to continuously evolve and adapt to new fraud patterns as they emerge, thereby ensuring that the system remains resilient to novel threats over time.
Finally, the concept of collaborative fraud detection can be explored, wherein multiple platforms or organizations share anonymized fraud data. This collaboration could improve the detection of fraud across a broader network, making it more difficult for fraudsters to operate undetected.
In summary, the system demonstrated strong potential, and these future developments could further enhance its robustness and applicability in diverse real-world scenarios.

APPENDIX

Dataset Source: The dataset used in this study was obtained from an open-source repository on Kaggle titled "Fraudulent Transactions Prediction" by Vardhan Siramdasu. It contains transactional records with labeled fraud instances used for training and evaluating the fraud detection model.
Link: https://www.kaggle.com/datasets/vardhansiramd asu/fraudulent-transactions-prediction

REFERENCES

[1] Sopiyan, Muhamad & Kasyfi, Fauziah & Wijaya, Yunan. (2022). Fraud Detection Using Random Forest Classifier, Logistic Regression, and Gradient Boosting Classifier Algorithms on Credit Cards. JUITA: Jurnal Informatika. 10. 77. 10.30595/juita. v10i1.12050.
[2] Talari, Navaneetha & Geetha, M & Nuthana, Bellamkonda & Rohan, Remalli. (2024). REAL- TIME FRAUD DETECTION IN ONLINE PAYMENTS: A COMPREHENSIVE REVIEW OF MACHINE LEARNING TECHNIQUES. 1341-1346. 10.58257/IJPREMS37828.
[3] Bilgaiyan, Ayush & Patel, Vinod & Journal, Ijeasm. (2024). A Machine Learning-based Credit Card Transaction Fraud Detection System for Imbalanced Data. 5. 2582-6948.
[4] K. G. Dastidar, O. Caelen and M. Granitzer, "Machine Learning Methods for Credit Card Fraud Detection: A Survey," in IEEE Access, vol. 12,pp.158939158965,2024,doi:10.1109/ACCES S.2024.3487298.
[5] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. -S. Hacid and H. Zeineddine, "An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection," in IEEE Access, vol. 7, pp. 93010-93022,2019,doi:10.1109/ACCESS.2019.2927266.
[6] Akinbusola, O., & Joseph, M. (2024). Fraud Detection using Machine Learning. ScienceOpen Preprints.
[7] Real-Time Fraud Detection Using Machine Learning. (2024.
[8] An integrated multistage ensemble machine learning model for fraud detection. (2024). Journal of Big Data, 9(1), 1-19.
[9] Hancock, J. T., & Bauder, R. A. (2023). Explainable machine learning models for Medicare fraud detection. Journal of Big Data, 10(1), 1-17.
[10] Financial fraud detection through the application of machine learning techniques: A systematic literature review. (2022). Humanities and Social Sciences Communications, 9(1), 1-10.