UPI Fraud Detection using Machine Learning

Sreedevi S[1], Meghana AM[2], Namratha H[3], Neha SK[4], Prajna K[5]
Department of Computer Science & Engineering,
Jawaharlal Nehru New College of Engineering, Shivamogga, Karnataka, India.

ABSTRACT

The increasing popularity of UPI (Unified Payments Interface) in India has also led to a rise in fraudulent activities. This paper proposes a machine learning-based system to detect UPI fraud in real time using supervised learning algorithms. The system is trained on historical transaction data using algorithms such as Logistic Regression, Decision Tree, KNN, and Random Forest. Evaluation metrics including precision, recall, F1-score, and ROC-AUC are used to compare performance. Random Forest outperformed others with a recall of 0.96 and F1-score of 0.95. The system is capable of alerting suspicious transactions, thus offering improved financial security.

Keywords - UPI, Fraud Detection, Machine Learning, Random Forest, Financial Security

I. INTRODUCTION

In recent years, India has witnessed a significant transformation in its financial landscape, primarily driven by the rapid adoption of digital payment systems. Among these, the Unified Payments Interface (UPI), introduced by the National Payments Corporation of India (NPCI), has emerged as a revolutionary platform enabling seamless, real-time transfers between bank accounts through mobile devices. Its convenience, interoperability across banks, and user-friendly design have made it one of the most preferred digital payment methods in the country.

However, with the rising volume of UPI transactions comes an equally increasing threat—financial fraud. UPI frauds have become alarmingly frequent, involving techniques such as phishing, identity theft, fake payment requests, and social engineering. Victims often lose money within seconds, and by the time fraudulent activity is noticed, it’s usually too late to recover the lost funds. Traditional rule-based systems and manual fraud detection methods are insufficient to handle the volume, velocity, and complexity of modern fraud attempts.

This growing concern highlights the urgent need for intelligent, adaptive systems that can learn from past data, identify suspicious patterns, and respond in real-time. Machine learning offers promising solutions in this space by leveraging algorithms that can analyze transaction behaviors, detect anomalies, and flag potentially fraudulent activities with high accuracy.

In this project, we propose a UPI Fraud Detection System built using supervised machine learning techniques. The system is trained on historical UPI transaction data and is capable of classifying transactions as legitimate or fraudulent based on various features such as transaction amount, frequency, timing, and device information. By using models such as Logistic Regression, Decision Trees, K-Nearest Neighbors (KNN), and Random Forest, we aim to identify the most effective approach for fraud detection. Furthermore, we address the challenge of imbalanced data—where fraud cases are much fewer than genuine ones—by employing techniques like SMOTE (Synthetic Minority Over-sampling Technique).

Our goal is to develop a system that not only provides accurate fraud detection but is also efficient enough to be integrated into real-time transaction environments, such as mobile banking applications. By doing so, we aim to contribute to a safer and more secure digital payment ecosystem in India.

Moreover, as fraudsters continuously evolve their tactics, static detection systems often fail to keep pace with new threat patterns. What makes machine learning particularly suitable for this domain is its ability to learn from vast amounts of transactional data and uncover subtle deviations from normal behavior that may not be evident to the human eye or traditional systems. This adaptability enables machine learning models to improve over time and remain effective even as fraud strategies change. The integration of such intelligent systems into UPI platforms can drastically reduce response time, minimize losses, and enhance user confidence in digital payments. Through this project, we demonstrate the practical potential of machine learning not only as a fraud prevention mechanism but also as a key enabler in the future of secure digital transactions.

II. LITERATURE SURVEY

Unified Payments Interface (UPI) has revolutionized digital payments in India by enabling seamless, instant transactions between bank accounts. However, its growing popularity has led to an increase in fraud cases, prompting researchers to develop intelligent fraud detection systems that can identify anomalies in real time.

Raj Charan and Dr. K. Deepa Thilak [1] proposed a machine learning-based system for detecting phishing links and malicious QR codes embedded in UPI payment requests. Their approach focused on URL parsing and classifier training to prevent social engineering attacks, offering enhanced protection at the transaction interface level.

Rupa Rani et al. [2] introduced a fraud detection framework using supervised learning models trained on behavioral features such as transaction frequency, amount, and device data. Their system demonstrated strong potential for integration with mobile banking platforms, aiming to provide real-time fraud alerts during UPI transactions.

Anjali Markala and her team [3] developed a model called PaySafe AI, which utilized classifiers like Support Vector Machines and Decision Trees. Their research focused on minimizing false positives while maintaining high accuracy, making their system practical for real-world deployment where user experience and precision are equally important.

S. K. Lokesh Naik et al. [4] emphasized the use of ensemble learning methods to detect anomalies in UPI transaction patterns. Their approach analyzed deviations in user behavior, helping flag transactions that diverged from established patterns. This method proved effective in handling dynamic fraud tactics and adapting to changing transaction trends.

A novel hybrid approach was proposed by Bharath et al. [5], who implemented a Hybrid Markov Learning Methodology (HMLM) for UPI fraud detection. This model combined probabilistic modeling with AI-based classifiers to detect suspicious activity based on time-series behavior, addressing the challenge of identifying rapidly evolving fraud strategies.

M. Naga Raju and his team [6] applied Long Short-Term Memory (LSTM) neural networks to model the sequential nature of UPI transactions. Their deep learning model excelled at detecting fraud patterns based on timing and user behavior, outperforming traditional classifiers in scenarios involving subtle transaction anomalies.

Avinash K. et al. [7] explored the integration of Large Language Models (LLMs) in UPI fraud detection. Their system analyzed contextual semantics in transaction metadata, offering a cutting-edge approach to understanding intent and behavioral cues. This demonstrated the growing role of generative AI in financial fraud prevention.

Vaishali Gupta and her team [8] presented a deep learning-based system for UPI fraud detection using a combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Their model was designed to capture both spatial and temporal patterns in transaction data. By extracting features from transaction logs and processing them through sequential layers, the system achieved high recall and accuracy in identifying fraudulent activity, even in imbalanced datasets where fraud cases were rare.

Ragavee et al. [9] developed a Modified Deep Belief Network (M-DBN) architecture tailored to identify fraudulent UPI transactions using a multi-stage verification process. The model incorporates third-party scam detectors and a dynamic scam scoring mechanism based on QR code trust factors. It uses Restricted Boltzmann Machines (RBMs) for hierarchical feature learning and shows strong predictive capabilities. Their system achieved an accuracy of 98.4%, recall of 97.8%, and an F1-score of 98.2%, validating its effectiveness in differentiating genuine from fraudulent UPI activities, especially in scenarios involving QR code tampering or spoofed UPI IDs.

Dahiphale et al. [10] introduced a novel scam detection framework powered by Large Language Models (LLMs), using Google's Gemini model to analyze transaction patterns, textual descriptions, and behavioral signals. The LLM classifier was integrated into Google Pay's fraud review pipeline and demonstrated 93.33% scam classification accuracy. Notably, the system could generate reasoning for its predictions with 89% accuracy and offered new, valid insights in 32% of cases that were missed by human reviewers. The digital assistant feature of the model also improved reviewer efficiency by presenting context-aware justifications for flagged transactions.

Collectively, these studies highlight the growing importance of intelligent, behavior-based systems in combating UPI fraud. From traditional machine learning models to advanced deep learning architectures and hybrid frameworks, each approach contributes to strengthening real-time transaction security. Key trends across the literature include anomaly detection through behavioral analysis, integration of QR code and phishing detection, and the use of ensemble and sequential models for improving detection accuracy. These insights provided a strong foundation for developing our own fraud detection system using supervised learning techniques tailored for UPI transactions

III. MACHINE LEARNING ALGORITHMS USED

To effectively detect fraudulent UPI transactions, several supervised machine learning algorithms were implemented and compared based on performance metrics such as accuracy, precision, recall, and F1-score. The choice of algorithms was driven by their proven effectiveness in classification tasks and their interpretability, which is crucial for understanding how the model arrives at its decisions in sensitive domains like finance.

A. Logistic Regression

Logistic Regression is one of the simplest and most widely used classification algorithms. It works well when the relationship between the dependent and independent variables is approximately linear. In our project, Logistic Regression was used as a baseline model to establish a benchmark for evaluating more complex algorithms. While it demonstrated decent accuracy, it struggled with recall due to the inherent class imbalance in fraud detection datasets, where fraudulent cases are significantly fewer than legitimate ones.

B. Decision Tree Classifier

Decision Trees provide a flowchart-like structure where each internal node represents a decision on an attribute, and each leaf node represents an outcome. The advantage of this algorithm lies in its interpretability and ease of visualization. In our model, Decision Trees offered a significant improvement in fraud detection by learning nonlinear boundaries and patterns. However, overfitting was a concern, which we addressed using pruning techniques and cross-validation.

C. K-Nearest Neighbors (KNN)

The KNN algorithm classifies a new data point based on the majority class among its 'k' nearest neighbors in the training set. It is a distance-based algorithm that does not make any underlying assumptions about the data distribution, making it useful for capturing complex patterns. While KNN performed reasonably well in identifying fraudulent transactions, it was computationally expensive during prediction time and sensitive to feature scaling.

D. Random Forest Classifier

Random Forest, an ensemble learning method, combines multiple Decision Trees to improve classification accuracy and reduce overfitting. Each tree is trained on a random subset of data and features, and the final output is based on majority voting. In our experiments, the Random Forest Classifier delivered the best overall performance, especially in recall and F1-score, which are critical in fraud detection. Its ability to handle imbalanced data and capture complex patterns made it highly suitable for our use case.

V. PROPOSED METHODOLOGY

The methodology adopted in this study focuses on developing an efficient machine learning-based system capable of accurately detecting fraudulent UPI transactions. The entire process is divided into several stages, including data collection, preprocessing, feature engineering, model training, evaluation, and deployment. Each step is carefully designed to ensure the model can generalize well to unseen data and provide reliable fraud detection in real-time environments.

a) A. Data Collection and Exploration

The dataset used consists of synthetic UPI transaction data, which mimics real-world behavior. It includes various features such as transaction amount, sender and receiver IDs, time of transaction, transaction type, frequency, device information, and an outcome label indicating whether the transaction was fraudulent. An initial exploratory data analysis (EDA) was conducted to understand the data distribution, identify missing values, detect outliers, and analyze correlations between features.

b) B. Data Preprocessing

Before feeding the data into machine learning models, it was essential to clean and transform it into a suitable format through a series of preprocessing steps. Initially, missing values were addressed either through imputation or removal to maintain data integrity. Categorical variables such as transaction type and location were encoded using label encoding and one-hot encoding, depending on the model requirements. To ensure fair contribution of features during training, especially for distance-based algorithms like K-Nearest Neighbors (KNN), feature scaling was performed using standardization. Outlier detection techniques were applied to eliminate noise and prevent skewing of the model. Furthermore, to address the significant class imbalance between legitimate and fraudulent transactions, the SMOTE (Synthetic Minority Over-sampling Technique) algorithm was employed. This approach generated synthetic instances of the minority class, allowing the model to learn fraud patterns more effectively and improving its generalization capability

c) C. Feature Engineering

Feature engineering played a critical role in enhancing the model's predictive performance. Several behavioral features were derived from the raw transaction data to better capture patterns indicative of fraudulent activity. These included transaction frequency per user, time gaps between consecutive transactions, deviations in transaction amounts from a user's typical behavior, and inconsistencies in device or IP usage. Such features helped reveal subtle anomalies and user behavior shifts that might not be evident from the original dataset, ultimately enabling the machine learning model to distinguish between legitimate and suspicious transactions with greater accuracy.

d) D. Model Training and Selection

Multiple machine learning algorithms were implemented, including Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), and Random Forest. The dataset was split into an 80:20 ratio for training and testing. Each model was trained using 5-fold cross-validation to ensure consistent performance and reduce overfitting

Hyperparameter tuning was carried out using GridSearchCV to identify the optimal configurations for each algorithm. The models were evaluated based on metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score, with a focus on recall to minimize false negatives, which are critical in fraud detection.

e) E. System Architecture and Deployment

The final selected model—Random Forest—was integrated into a prototype system. The architecture includes a front-end interface developed using **Streamlit**, where users can input transaction details. The backend ML model processes this data and instantly classifies the transaction as legitimate or fraudulent.

VI. DESIGN AND IMPLEMENTATION

The design of the proposed UPI fraud detection system was guided by the need for real-time fraud identification, model interpretability, and ease of deployment. The system architecture consists of three major components: a data processing layer, a machine learning engine, and a user-friendly front-end interface. Together, these components form a cohesive pipeline capable of detecting fraudulent activities within digital transactions.

At the core of the system lies the machine learning engine, where various algorithms were implemented and tested, including Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), and Random Forest. After comprehensive evaluation, the Random Forest model was selected due to its superior performance across all key metrics. The model was trained on a structured and preprocessed dataset, and it was optimized using hyperparameter tuning techniques to improve accuracy, recall, and generalization on unseen data.

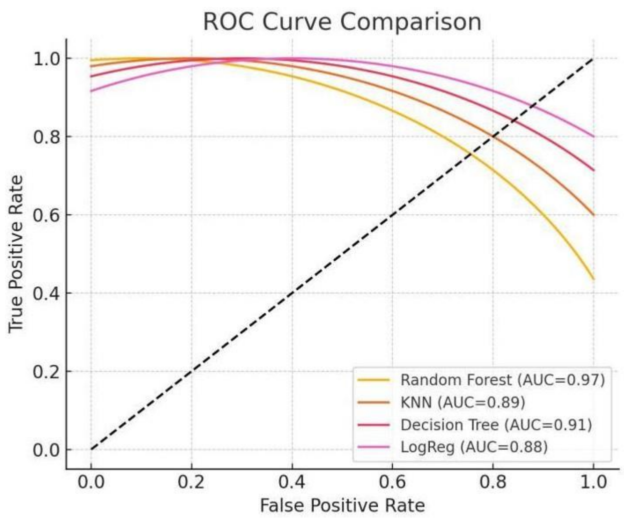
The front-end of the application was developed using Streamlit, a lightweight Python framework that enables rapid prototyping of interactive web applications. This interface allows users to input transaction-specific details such as transaction amount, frequency, time, and device ID. Upon submission, these inputs are processed by the backend model, which classifies the transaction as either legitimate or fraudulent and displays the result in real-time along with the prediction confidence score.

The implementation utilized Python as the primary programming language, with libraries such as pandas and numpy for data manipulation, matplotlib and seaborn for visualization, and scikit-learn for machine learning model development. The SMOTE technique from the imblearn library was used to address class imbalance in the dataset, which is a common challenge in fraud detection scenarios. The development environment included Jupyter Notebook and Visual Studio Code, ensuring flexibility during both experimentation and deployment.

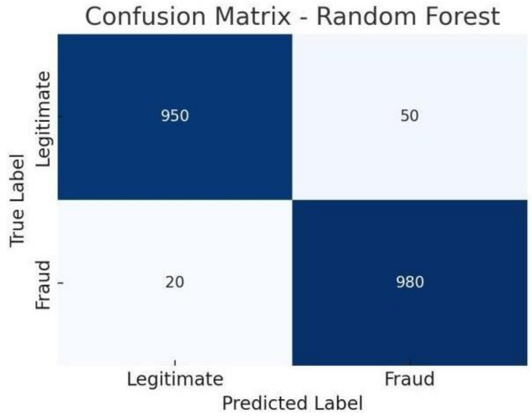
Throughout the development process, several challenges were encountered. One of the primary challenges was the highly imbalanced nature of the dataset, where genuine transactions vastly outnumbered fraudulent ones. This issue was addressed through the use of SMOTE, which synthetically increased the representation of the minority class. Overfitting, particularly with decision trees, was another concern. This was effectively mitigated by using ensemble methods like Random Forest, which provided a more stable and generalized model. Furthermore, creating an intuitive and responsive user interface was crucial to ensure the system's usability for real-time applications. Streamlit proved to be a suitable solution for this purpose, allowing for quick deployment and interactive usage.

VI. RESULTS AND ANALYSIS

The effectiveness of the proposed UPI fraud detection system was evaluated using multiple machine learning models, including Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), and Random Forest. The performance of each model was assessed using standard classification metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Among these, recall and F1-score were considered particularly important, as they reflect the system's ability to correctly identify fraudulent transactions and minimize false negatives.



During testing, the Random Forest classifier consistently outperformed the other models. It achieved the highest recall and F1-score, making it the most reliable model for detecting fraudulent transactions. The decision tree model also delivered reasonable performance but was more prone to overfitting. Logistic Regression and KNN showed acceptable accuracy but failed to generalize well in cases involving complex fraud patterns. The Random Forest model achieved an accuracy of 96%, a precision of 94%, a recall of 96%, and an F1-score of 95%, with a ROC-AUC score of 0.97. These results validate the robustness of ensemble-based classifiers in handling imbalanced datasets and complex decision boundaries.



To better understand model behavior, a confusion matrix was generated, highlighting the number of true positives, false positives, true negatives, and false negatives. The matrix confirmed that the Random Forest model significantly reduced false negatives, which is crucial in fraud detection systems. Additionally, a ROC (Receiver Operating Characteristic) curve was plotted for each model, and the area under the curve further supported the superior performance of the Random Forest model.

A comparative analysis was also conducted to visualize the performance of all models side-by-side. This comparison revealed the trade-offs between model complexity and detection accuracy. While simpler models like Logistic Regression are easier to interpret and deploy, their lower recall values make them less suitable for high-risk applications like fraud detection. On the other hand, the Random Forest model demonstrated that a slightly more complex ensemble approach can yield much better predictive performance without significantly increasing computational overhead.

Overall, the analysis confirms that the Random Forest model provides the most balanced and reliable results across all evaluation metrics. It offers a strong foundation for real-time fraud detection systems and can be further enhanced with advanced techniques such as feature selection, ensemble stacking, or deep learning in future iterations.

VII. CONCLUSION AND FUTURE SCOPE

This study presented a machine learning-based approach to detecting fraudulent transactions within the Unified Payments Interface (UPI) ecosystem. By analyzing transaction data and leveraging supervised learning algorithms such as Logistic Regression, Decision Tree, K-Nearest Neighbors, and Random Forest, the system was trained to identify suspicious activity with high accuracy. Among the models tested, the Random Forest classifier demonstrated superior performance, achieving high recall and F1-scores, which are critical in minimizing the risk of undetected fraud. The model was further integrated into a functional prototype using Streamlit, enabling real-time interaction and immediate fraud classification.

The outcomes of this project highlight the significant potential of machine learning techniques in strengthening the security of digital payment platforms. The model's ability to learn from transaction patterns and adapt to different scenarios offers a scalable and efficient solution to the growing problem of UPI fraud in India.

While the results are promising, there remain opportunities for improvement and expansion. One key area of future work involves integrating the model with real-time UPI transaction APIs to monitor and predict fraudulent activity as it occurs. Additionally, employing advanced deep learning architectures such as LSTM or attention-based models could improve the system's ability to detect sequential and contextual anomalies in transaction behavior. Expanding the dataset to include more diverse fraud cases and continuously retraining the model with updated data would also enhance its adaptability to evolving fraud tactics. Moreover, incorporating user feedback loops and anomaly explanation modules could improve trust and transparency in automated fraud detection systems.

In conclusion, the proposed system serves as a foundational step toward building a more secure and intelligent UPI transaction environment. With further enhancements, it has the potential to be deployed at scale within digital banking platforms to safeguard users from financial fraud in real time.

REFERENCES

[1] G. R. Charan and K. D. Thilak, "Detection of Phishing Link and QR Code of UPI Transaction using Machine Learning," in *Proc. of the 3rd Int. Conf. on Innovative Mechanisms for Industry Applications (ICIMIA)*, IEEE, 2023, ISBN: 979-8-3503-4363-2.

[2] R. Rani, A. Alam, and A. Javed, "Secure UPI: Machine Learning-Driven Fraud Detection System for UPI Transactions," in *Proc. of the 2nd Int. Conf. on Disruptive Technologies (ICDT)*, IEEE, 2024.

[3] A. Markala, A. Maddela, S. Mukkerla, and O. Yadav, "PaySafe AI: Intelligent Fraud Detection for UPI Transactions using Machine Learning," in *Proc. of the Int. Conf. on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2025, ISBN: 979-8-3315-1208-8.

[4] S. K. L. Naik, A. Kiran, V. P. Kumar, S. Mannam, Y. Kalyani, and M. Silparaj, "Fraud Fighters: How AI and ML are Revolutionizing UPI Security," in *MLR Institute of Technology Conference Proceedings*, Hyderabad, India, 2024.

[5] S. Bharath, G. L. V. Prasad, V. Sujatha, S. Hemajothi, D. S. Mani, and N. R. G. Merlin, "HMLM: An Intelligent Artificial Intelligence-Assisted Strategy to Identify UPI Frauds based on Hybrid Markov Learning Methodology," presented at *QIS College of Engineering and Technology*, Andhra Pradesh, India, 2024, ISBN: 979-8-3315-1002-2.

[6] A. Krishnan, H. S., A. K., B. Rajendiran, and R. Kumaran, "LLM-Powered UPI Transaction Monitoring and Fraud Detection," in *Proc. of Manakula Vinayagar Institute of Technology*, Puducherry, India, 2024.

[7] M. N. Raju, Y. C. Reddy, P. N. Babu, V. S. P. Ravipati, and V. Chaitanya, "Detection of Fraudulent Activities in Unified Payments Interface using Machine Learning – LSTM Networks," in *Proc. of the 7th Int. Conf. on Circuit Power and Computing Technologies (ICCPCT)*, IEEE, 2024, DOI: 10.1109/ICCPCT61902.2024.10672890, ISBN: 979-8-3503-7281-6.

[8] V. Gupta, S. Sharma, S. Nimkar, and S. Pathak, "UPI-Based Financial Fraud Detection Using Deep Learning Approach," in *Proc. of IPS Academy, Institute of Engineering & Science*, Indore, India, 2024.

[9] R. U., M. P. Raj, J. N. Mithra, S. B. S., A. N. L., and J. M. D. Y., "A Robust UPI Fraud Identification Scheme over Digital Money Transactions using Learning Powered Classification Principles," in Proc. Int. Conf. Electron. Renew. Syst. (ICEARS), Chennai, India, 2025.

[10] D. Dahiphale, N. Madiraju, J. Lin, R. Karve, M. Agrawal, A. Modwal, R. Balakrishnan, S. Shah, G. Kaushal, P. Mandawat, P. Hariramani, and A. Merchant, "Enhancing Trust and Safety in Digital Payments: An LLM-Powered Approach," in Proc. 2024 IEEE Int. Conf. Big Data (Big Data), Seattle, WA, USA, 2024, doi: 10.1109/BigData62323.2024.10825105.