

UPI Fraud Detection Using Machine Learning with AI-Based Risk Score and Transaction Blocking

Poomathi M.¹; Mirunalini A.²

^{1,2}Department of CSE, K. L. N. College of Engineering, Pottapalayam, 630612

Publication Date: 2026/10/22

Abstract: The rapid growth of Unified Payments Interface (UPI) transactions has increased the need for effective methods to detect and prevent digital payment fraud. This project proposes a machine learning-based UPI fraud detection system that analyzes transaction details and identifies patterns that may indicate suspicious activity. Features such as transaction amount, time, frequency, recipient information, device details, and transaction history are used to assess each payment. The system applies machine learning to classify transactions and generate an AI-based risk score. It provides explanations for flagged transactions, such as unusually large payments, repeated transfers within a short period, or payments to unfamiliar recipients. Based on configurable risk thresholds, the system can approve low-risk transactions, send medium-risk transactions for review, and block high-risk transactions. A web-based dashboard displays transaction records, risk scores, alerts, explanations, and decision status for users and administrators. The proposed system aims to support faster fraud monitoring and provide a structured process for handling suspicious payments. Its performance depends on the quality and representativeness of the transaction data used for training and evaluation. The system is intended as a fraud prevention support tool, with final deployment and transaction blocking dependent on integration with authorized payment services

Keywords: UPI, Machine Learning, Fraud Detection, AI-Based Risk Scoring, Transaction Blocking, Digital Payments, Anomaly Detection.

How to Cite: Poomathi M.; Mirunalini A. (2026) UPI Fraud Detection Using Machine Learning with AI-Based Risk Score and Transaction Blocking. *International Journal of Innovative Science and Research Technology*, 11(10), 1175-1180. <https://doi.org/10.38124/ijisrt/26oct663>

I. INTRODUCTION

Digital payment systems have changed how people transfer money and pay for goods and services. In India, the Unified Payments Interface (UPI) enables users to make instant payments through mobile applications using details such as a UPI ID, QR code, or linked bank account. Its convenience has encouraged widespread use by individuals, merchants, and businesses. At the same time, the speed and scale of digital payments make it important to identify suspicious activity as early as possible.

UPI users may face different types of fraud, including phishing, fake payment requests, impersonation, QR code scams, account compromise, and transfers to fraudulent recipients. Fraudulent activity can resemble normal payment behavior, making it difficult to identify using a single transaction detail. For example, a payment may be suspicious because of its amount, timing, recipient, device, or relationship to several recent transactions.

Traditional fraud detection methods often rely on fixed rules and manual review. Rules can be useful for detecting known patterns, such as repeated payments within a short

period, but fraud methods can change over time. A rule that is too strict may flag legitimate payments, while a rule that is too broad may fail to identify suspicious activity. Reviewing every transaction manually is also challenging when payment volumes are high.

Machine learning provides a way to analyze multiple transaction features together. A model can learn patterns from previously labeled transaction records and use those patterns to estimate whether a new transaction may be fraudulent. Relevant features may include transaction amount, time, frequency, recipient information, device details, location where available, and the user's transaction history. Data preparation and model evaluation are important because inaccurate or unbalanced training data can affect the results.

This project proposes a machine learning-based UPI fraud detection system that evaluates transaction information and generates an AI-based risk score. The score represents the system's estimate of transaction risk and helps distinguish transactions that appear low, medium, or high risk. The system can also display contributing factors—such as an unusually high amount, multiple rapid payments, or a first-

time recipient—to help users and administrators understand why a transaction was flagged.

The proposed system uses configured risk thresholds to support transaction handling. Low-risk transactions may be approved, medium-risk transactions may be sent for additional review, and high-risk transactions may be blocked within the application workflow. A web-based dashboard displays transaction records, risk scores, explanations, alerts, and decision statuses. This provides a central place for monitoring transactions and reviewing suspicious cases.

The system is intended to support fraud monitoring by organizing transaction analysis and risk-based actions. Its output is not proof that a transaction is fraudulent; model predictions may include false alerts or fail to detect some fraud. Reliable results depend on the quality, relevance, and representativeness of the data used for training and evaluation. Integrating blocking into a live UPI payment service would also require connection with an authorized payment provider.

Overall, the project applies machine learning to a practical digital payment security problem. By combining transaction analysis, AI-based risk scoring, reason explanations, and approval, review, or blocking actions, the system aims to help identify suspicious UPI transactions and support timely responses.

II. LITERATURE SURVEY

- “A Novice Approach for UPI Fraud Detection by Using Machine Learning Techniques” compares classifiers such as Decision Tree, SVM, Random Forest, and Logistic Regression, including ensemble combinations. The study highlights the potential of combining models to classify suspicious UPI transactions. Its experiments use a synthetic, imbalanced dataset, so validation using representative transaction data remains important.
- “Enhanced Detection of UPI Frauds Using Advanced Machine Learning Techniques for Secure Digital Transactions” proposes a framework combining transformer models, causal inference, and federated learning. It focuses on transaction behavior, privacy, and interpretable risk scoring. The advanced approach may require substantial data and computing resources to implement.
- “FedUPI: Federated Learning Empowered Detection of Fraudulent UPI Transactions” explores federated learning for UPI fraud detection. Its approach is designed to train models across separate data sources while limiting the sharing of raw transaction data. This can support privacy, though federated model development and coordination can add implementation complexity.
- “Adaptive UPI Fraud Detection: A Hybrid Machine Learning Framework” combines machine learning methods, including CNN and Hidden Markov Model

techniques, to detect fraud patterns in UPI payments. The study emphasizes adaptive and real-time detection. Its hybrid design may be more complex than a lightweight model intended for a small application.

- “Machine Learning-Based UPI Fraud Detection: A Comprehensive Approach Using Random Forest” applies Random Forest to transaction data that includes amount, timing, sender and receiver activity, and device information. It describes preprocessing and feature selection as important steps in fraud classification. The model’s results depend on the dataset used and require evaluation on data representative of the intended system.
- “Secure UPI: Machine Learning-Driven Fraud Detection System for UPI Transactions” applies XGBoost to labeled UPI transaction data and uses feature importance to identify indicators of potential fraud. The study’s focus on transaction features and real-time monitoring is relevant to this project. However, an operational system must also define how predictions lead to review or blocking actions.
- “Experimental Evaluation of Unified Payment Interface (UPI) Fraud Detection System Using Elevated Deep Learning Methodology” combines ShuffleNet with SVM for fraud classification. It explores deep learning for extracting transaction patterns and classifying suspicious activity. The approach focuses on detection performance; risk-based transaction decisions and explanations can be developed as additional system functions.
- “Fraud Fighters: How AI and ML are Revolutionizing UPI Security” discusses using AI and machine learning to identify potentially fraudulent UPI activity, including challenges such as imbalanced data and changing fraud patterns. It supports the use of adaptive data-driven detection. A practical system must also manage false alerts so legitimate payments are not unnecessarily interrupted.
- “UPI Fraud Detection Using Machine Learning Algorithms” compares techniques including Autoencoder, Local Outlier Factor, and K-means clustering for detecting suspicious UPI transactions. Its use of multiple methods reflects the value of testing both supervised and anomaly-based approaches. The study also points to the challenge of minimizing false positives in transaction monitoring.

➤ Existing System

Existing UPI fraud detection systems commonly use transaction rules, predefined limits, and basic monitoring to flag suspicious payments. Some systems apply machine learning to identify unusual transaction patterns.

However, approaches based mainly on fixed rules may not adapt quickly when fraud tactics change, and can flag genuine activity or miss suspicious transactions.

In many basic implementations, the system reports a transaction as suspicious without providing a clear risk score or explaining the factors behind the alert. Transaction analysis, alert review, and blocking may also be handled as separate processes.

This can make it difficult for users or administrators to decide how to respond consistently. These limitations motivate a system that combines machine learning analysis, explainable risk scoring, and approval, review, or blocking actions in one workflow.

➤ *Proposed System*

The proposed system detects potentially fraudulent UPI transactions using machine learning and an AI-based risk score.

It analyzes transaction features such as amount, time, frequency, recipient details, device information, and user transaction history to identify unusual patterns.

The model classifies each transaction and generates a risk score that indicates its estimated fraud risk.

The system explains the factors contributing to a suspicious alert, such as an unusually large payment, rapid repeated transactions, or a new recipient.

Based on configured risk thresholds, low-risk transactions can be approved, medium-risk transactions sent for review, and high-risk transactions blocked.

A web dashboard displays transaction details, risk scores, explanations, alerts, and decision status to help users and administrators monitor payments and review flagged activity.

This integrated approach aims to support timely fraud detection and consistent transaction handling.

III. METHODOLOGY

➤ *Transaction Data Collection:*

The first step is to collect transaction records for training, testing, and evaluating the fraud detection model. Each record may contain the transaction amount, date and time, sender and recipient details, transaction frequency, device information, and available transaction history. Records should be labeled as genuine or fraudulent where labels are available. Sensitive information should be removed or masked before it is used for analysis.

➤ *Data Preprocessing:*

The collected data is checked for missing values, duplicate records, invalid entries, and inconsistent formats. Categorical values, such as recipient or device categories, are converted into a format the model can process. Numerical features are checked and scaled when required by the selected algorithm. The dataset is then divided into training and testing sets to evaluate how well the model handles transactions that were not used during training.

➤ *Feature Engineering and Class Imbalance Handling:*

Useful features are prepared from the transaction records. These may include the amount compared with the user's usual payment range, the number of transactions within a recent time window, whether the recipient is new, and whether the device has been used before. Fraudulent transactions may be much less common than genuine ones, creating an imbalanced dataset. Class weighting or other suitable sampling methods can be considered to help the model learn from the minority fraud cases.

➤ *Machine Learning Model Training:*

Machine learning algorithms such as Random Forest, Logistic Regression, or Gradient Boosting can be trained using the prepared transaction features. The models learn patterns associated with genuine and fraudulent transactions from the labeled training data. More than one model can be compared to select an approach that balances fraud detection, false alerts, and processing time.

➤ *Model Evaluation and Selection:*

The trained model is evaluated using separate test data. Accuracy alone may not be sufficient because fraudulent cases can be relatively rare. Measures such as precision, recall, F1-score, confusion matrix, and false positive rate can provide a clearer view of performance. The selected model should be based on evaluation results relevant to the project's objectives and dataset.

➤ *AI-Based Risk Scoring:*

For each submitted transaction, the model estimates the probability that the transaction is fraudulent. This probability is converted into a risk score, for example from 0 to 100, where higher scores indicate greater estimated risk. The score represents a model-based estimate and is used to support transaction handling.

IV. RESULT

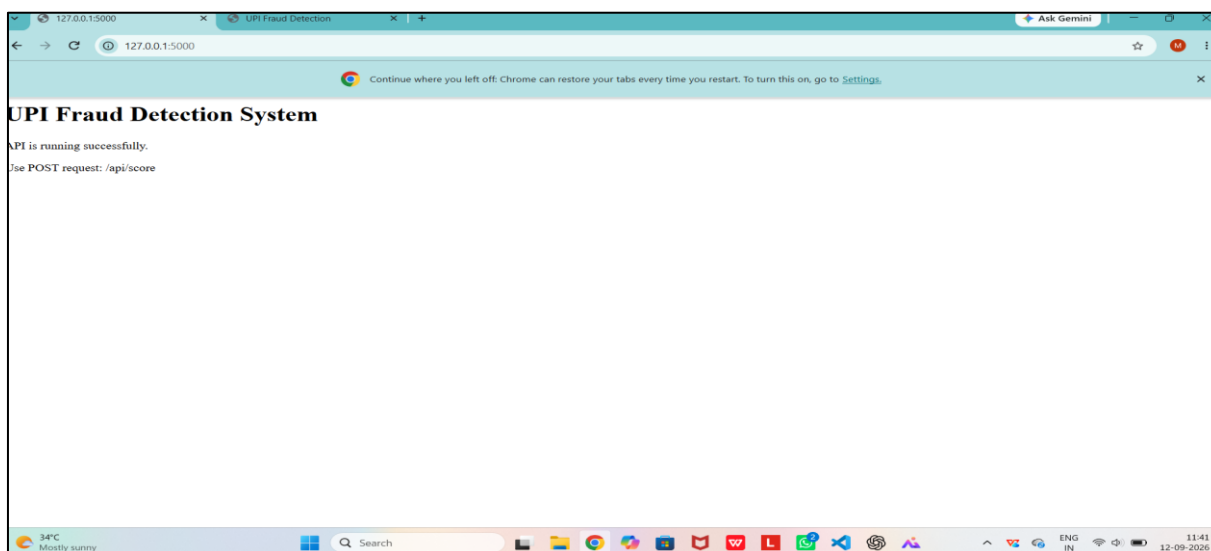


Fig 1 Web Page to Detect

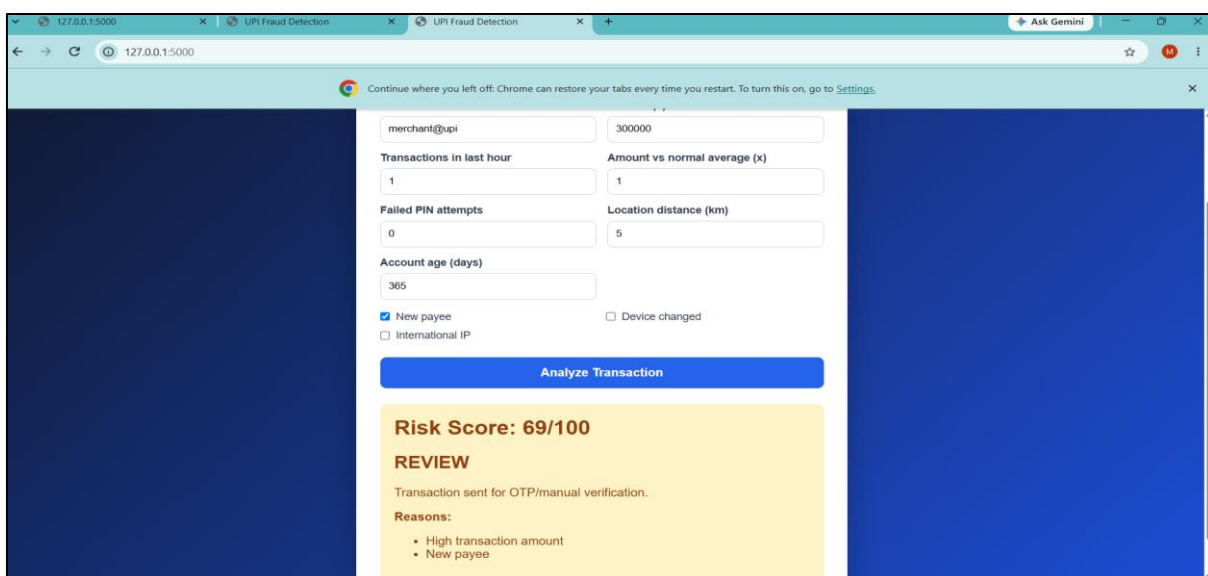
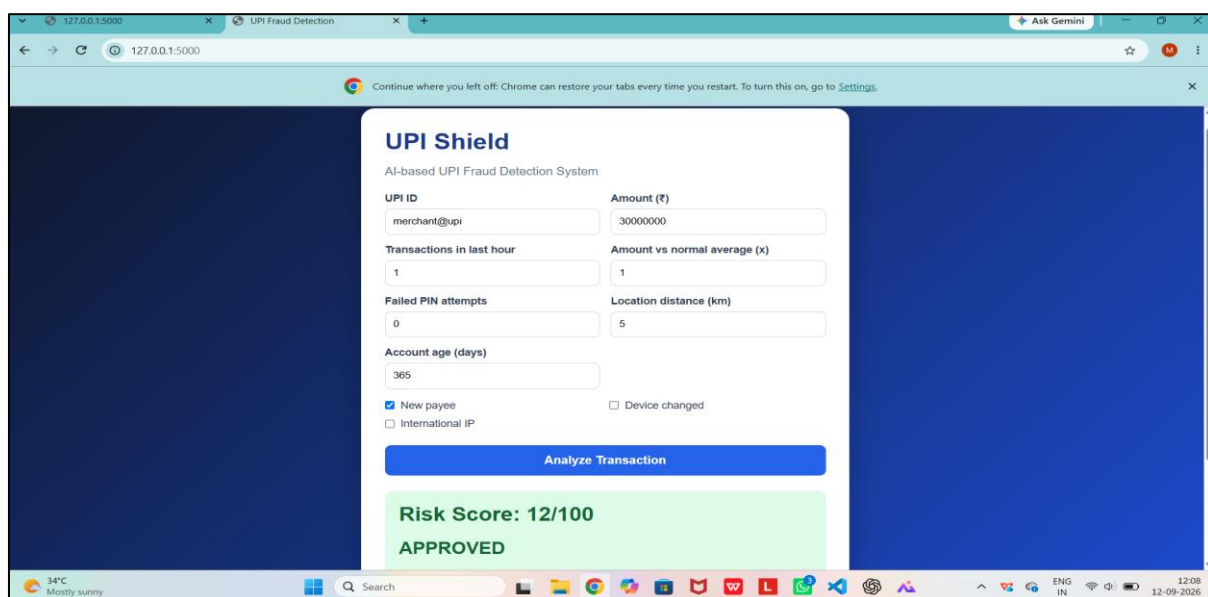


Fig 2 Result Analysis

V. CONCLUSION

The UPI Fraud Detection System developed in this project demonstrates how machine learning can support the identification of potentially fraudulent digital payment transactions. By analyzing transaction details such as amount, time, frequency, recipient information, device status, and available transaction history, the system generates a fraud prediction and an AI-based risk score.

The system connects risk scoring with a transaction handling workflow. Based on configured thresholds, low-risk transactions can be approved, medium-risk transactions sent for review, and high-risk transactions blocked within the application. The dashboard presents transaction information, alerts, risk scores, explanations, and decision statuses to users and administrators.

The project brings together transaction processing, machine learning analysis, risk-based decisions, and record management in a web-based system. Its effectiveness depends on the quality and representativeness of the data used to train and evaluate the model. Overall, the system provides a foundation for UPI fraud monitoring and can be extended with additional data, improved models, and authorized integration with payment services.

REFERENCES

- [1]. Sethi, Bhramanand, Sarvednya Mhatre, Sachin Yadav, Siuli Das, and Vaishali Jadhav. "Machine Learning-Based UPI Fraud Detection: A Comprehensive Approach Using Random Forest." *Proceedings of ICAIEHS 2025*, 2025. https://doi.org/10.2991/978-94-6463-852-3_29
- [2]. Sree, Mangali Vijaya, Panyam Bhuvaneshwari, Karne Nandini, Koppula Ratna Phoebe Amulya, Palla Swetha, and B. Chandrakala. "Enhanced Detection of UPI Frauds Using Advanced Machine Learning Techniques for Secure Digital Transactions." *Proceedings of ICITSM 2025*, 2025. <https://doi.org/10.4108/eai.28-4-2025.2357830>
- [3]. "A Novice Approach for UPI Fraud Detection by Using Machine Learning Techniques." *2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA)*, 2025. <https://doi.org/10.1109/ICCCA66364.2025.11325722>
- [4]. "AI-Driven Detection Mechanism for UPI Fraud and QR Code Tampering." *2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2025. <https://doi.org/10.1109/ICICV64824.2025.11085656>
- [5]. "FedUPI: Federated Learning Empowered Detection of Fraudulent UPI Transactions." *2025 5th International Conference on Soft Computing for Security Applications (ICSCSA)*, 2025. <https://doi.org/10.1109/ICSCSA66339.2025.11170880>
- [6]. "Secure UPI: Machine Learning-Driven Fraud Detection System for UPI Transactions." *2024 2nd International Conference on Disruptive Technologies (ICDT)*, 2024. <https://doi.org/10.1109/ICDT61202.2024.10489682>
- [7]. "Experimental Evaluation of Unified Payment Interface (UPI) Fraud Detection System Using Elevated Deep Learning Methodology." *2024 2nd International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS)*, 2024. <https://doi.org/10.1109/ICETEMS64039.2024.10965083>
- [8]. "Fraud Fighters: How AI and ML are Revolutionizing UPI Security." *2024 International Conference on Science Technology Engineering and Management (ICSTEM)*, 2024.

<https://doi.org/10.1109/ICSTEM61137.2024.10560740>

- [9]. Sindhu, Jallapuram, and Vijaya Sree Swarupa. “UPI Fraud Detection Using Machine Learning Algorithms.” *International Journal of Engineering Research and Science & Technology*, Vol. 20, No. 4, pp. 57–67, 2024. Article link
- [10]. Anupriya, K., K. Avinash, S. Hariharan, R. Bhavesh, and R. R. Kumaran. “LLM-Powered UPI Transaction Monitoring and Fraud Detection.” *2024 International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2024. <https://doi.org/10.1109/ICSCAN62807.2024.10894012>