WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

UPI Fraud Detection Using Machine Leaming: A Comparative Model-Based Approach

Sanat Narang[1], Shrasti Yadav[2], Shreya Beltharia[3], Shivam Mishra[4], Dr. Deepshika Sharma[5]

[1-5]Oriental Institute of Science and Technology, Bhopal,
Email- sanatnarang4@gmail.com, yadavshrasti247@gmail.com, belthariashreya@gmail.com, sm0929976@gmail.com, Oisthodcse@oriental.ac.in

ABSTRACT

The advent of the Unified Payments Interface (UPI) has completely changed the digital transaction landscape in India by making it easy and fast to transfer money; nonetheless, this widespread adoption has drawn the attention of fraudsters to the extent that user security and financial trust are put at risk. On the other hand, traditional rule-based detection systems cannot meet the modern requirements because they usually do not adapt to complex and constantly changing fraud patterns and their performance is not efficient with the increasing transaction volumes. The use of Machine Learning (ML) for fraud detection and prevention in UPI transactions is the focus of this paper; it will be done in real-time through the analysis of transaction patterns and user profiles.

The research uses a comparative model-based approach and assesses first the performance of the four main algorithms: Random Forest, Logistic Regression, Support Vector Machine (SVM), and Decision Tree. The research workflow includes data collection, preprocessing, feature extraction, and the production of a cyber-themed web application prototype for real-time monitoring. Accuracy, precision, recall, F1-score, and AUC were among the performance metrics applied to the models for validation.

The Random Forest algorithm gained the highest accuracy among all classifiers with an impressive score of 94%, precision of 0.95, and recall of 0.90 which practically eliminated false alerts. The research results suggest that the application of ensemble learning strategies offers a very powerful, flexible, and scalable anti-fraud solution in the financial sector. It is through a proactive framework that this study aids the digital payment ecosystem in being secure by preserving the data's authenticity and thus gaining users' trust.

Keywords: Unified Payments Interface (UPI), Machine Learning, Fraud Detection, Random Forest, Cyber Security, Digital Payments.

1. Introduction

1.1 Context and Background

Unified Payments Interface (UPI) has undoubtedly changed the digital transaction scene in India for good by giving a fast, real-time, and an interoperable platform for fund transfers. The main factors that struck the users such as the simplicity of the interface and instant processing have also led to the huge adoption of UPI and made it a major player in the modern digital economy and, hence, speeding up the inclusion of various categories of people into the financial system. UPI has reached a stage where it is doing billions of transactions every month; this has highlighted its necessity and the central role it plays in India's financial system.

1.2 The Problem Statement and Research Gap

UPI's success and convenience have, however, been accompanied by the introduction of combined advance fraudulent activities that are also at the same rate as UPI's growth. Phishing and social engineering are the basic fraudulent methods while, on the other hand, the sophisticated transaction pattern exploitation is the most advanced one among the frauds. All these incidents put the users' confidence at risk besides causing huge financial losses. Conventional fraud management practices that are either rule-based or manual monitoring rely upon have, however, proven to be insufficient. They are non-adaptive by nature and therefore slow in responding to the constantly changing, dynamic nature of modern fraud schemes. This shortage brings forth a major research gap: there is no automated system that is proactive, scalable, and extremely accurate in identifying and mitigating suspicious activity during the milliseconds required for a real-time UPI transaction that can be delivered with the current fraud detection technology.

1.3 Solution, Aim, and Contribution

This paper presents the creation and testing of a Machine Learning (ML)-based system for the instantaneous detection of UPI fraud as a solution towards the very serious security problem. By this primary aim one could say that the transaction data will be analyzed through sophisticated algorithms that will perform the task of separating the super clean transactions from the fraudulent ones with a very high level of success. This research is for sure focused on the comparison of different methodological models while especially scrutinizing the output and utility of the most popular supervised learning methods, e.g. Random Forest, Logistic Regression, Τε και οι άλλες. The most important part of this paper is the introduction of a very predictive and thorough fraud detection model, which is not only authenticated by strict metrics but also gives a whopping increase in accuracy and a decrease in false positives when compared to traditional approaches which in turn supports and upholds the whole UPI ecosystem.

1.4. Paper Structure and Roadmap

The rest of the paper is organized in a manner that: Section 2 gives an in-depth literature review on the topics of digital payment security and the use of machine learning for fraud detection. Methodology is described in Section 3 which includes a discussion of the data set, the feature engineering techniques used and the selected ML models' implementation details. In Section 4 the experimental results and models' performance oversight nurtured through comparison are revealed. Finally, Section 5 wraps up the study with outlining the implications of the findings and proposing future research directions.

2. Literature Review

The rapid digitalization of India's financial ecosystem, which was brought about by the government and the technological environment, has profoundly changed the way consumers perform their transactions. The introduction of Unified Payments Interface (UPI) has been the leading factor in this change, making it possible to have P2P and B2C transfers done right away with the new Virtual Payment Address (VPA) thus doing away with the disclosing of sensitive banking information (Narendra Kumar et al., 2020).

2.1 Context and Growth of Digital Payments

The Reserve Bank of India has set the "less cash" concept as a vision plan which rests heavily on the adoption of mobile payment technologies (Mohapatra et al., 2017). This trend is further backed by the rapid development of the FinTech sector which has been nurtured by the aforementioned major government projects and the National Payments Corporation of India (NPCI) among others. Different scholars have been stressing the importance of UPI as the major driver of digitalization and at the same time financial inclusion (Chatterjee & Thomas, 2017). The number of internet users has been increasing steadily, and that between years 2010-2017 the increase was considerable, thus operating as the foundation of the digital adoption.

2.2 Security Challenges and Existing Research

The question of security for digital payment platforms is always related to comfort and their expansion leading to being potential research topics. User trust and financial security are heavily impacted by the fraudulent practices surrounding the UPI system. Such tricks are based on human behavior and the fake verification processes are also a part of their advantage.

In the past, the security of mobile banking apps and UPI has been studied with an emphasis on analyzing and standing by the security model. Some researchers have even performed security evaluations on the UPI architecture itself (Kumar et al., 2020), while others have focused narrowly on the necessary improvements for UPI-based mobile banking applications (Lakshmi et al., 2019). This body of research points to the fact that strong security measures are still necessary for securing the transactions.

2.3 The Research Gap: Limitations of Traditional Fraud Detection

The literature has highlighted an important gap, which is the inability of traditional fraud detection techniques to cope up with the magnitude and high-tech nature of modern UPI frauds. Conventional rule-based systems lack flexibility, they have a fixed set of rules that are manually created and defined, and thus, they cannot adjust to the fast changing, data-driven fraud tactics. Moreover, these systems have low performance, which is roughly estimated at 70%-80% accuracy, and limitations in terms of scalability, and they also lack the capability to process in real-time. The core issue is detecting fraudulent patterns with the highest degree of precision while millions of legitimate transactions are processed in real time daily. This drawback necessitates a switch to an adaptive, data-centric approach.

2.4 Application of Machine Learning in Fraud Detection

The issue has been proposed as a potential application of Machine Learning (ML) techniques in the literature. The major advantages of using ML models are the handling of massive data, the uncovering of different and hidden relationships and the detection of strange behaviors through the application of the three learning methodologies: supervised, unsupervised, and semi-supervised.

The quest for algorithms with the strongest capability to detect fraud has brought forth a frequent reference to Random Forest (RF) as an ensemble learning method with high precision. Moreover, it is believed that good selection and engineering of features are crucial for model performance in this domain. The study aims at a fair comparison of the top ML algorithms—Random Forest, Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM)—to build an extremely accurate (targeting >90% accuracy), adaptive, and scalable real-time UPI fraud prevention system capable of tackling a situation like this one (Jagadeesan et al., 2025).

3. Problem statement

The Unified Payments Interface (UPI) has been a huge success in transforming the Indian digital payment landscape. However, the breakthrough, which made UPI widely adopted, has also created a major problem in the form of fraudulent activities that are prevalent and growing. Such a compromised system gives a bad impression about its security and trustworthiness, further driving away users and causing financial losses for both customers and financial entities alike.

The main problem splits into two parts:

- Rising and Evolving Fraud: A huge number of UPI transactions not only indicate that more people are using it but also that the frauds are becoming increasingly innovative, widespread, and sometimes even hard to detect.
- Inadequacy of Traditional Systems: The current measures for detecting fraudulent activities are not at the same level as their dynamic and complex nature. The conventional means for spotting fraud generally depend on static rule-based systems and are therefore not able to respond quickly enough to the modern-day patterns of fraud. These systems lack the much-needed agility and robustness to spot even the most minute and complex anomalies in the river of transactions that is high volume, in real-time.

The research problem, hence, is to come up with and substantiate an efficient, flexible, and robust anti-fraud solution that would make use of advanced Machine Learning (ML) algorithms to detect and deal with the different types of UPI frauds easily thus securing digital transactions and keeping the UPI ecosystem user-friendly.

4. Proposed Methodology

The fundamental methodology involves using state-of-the-art machine learning techniques for the purpose of achieving both better scalability and accuracy compared to traditional rule-based fraud detection systems.

4.1 Data Acquisition and Feature Engineering

The very beginning of the project is dedicated to making the data ready so that the machine learning models can work at their best capacity.

Data Source: The UPI transaction data is the main one and millions of transactions are included besides.

Feature Engineering: It is considered as the key step for improving the models' performance. It consists of changing the raw transaction details into the informative features that show the abnormal behavior of the customers. The following features are considered for enrichment:

- Transaction velocity (number of transfers within a certain period).
- The history of the relationship between the parties involved.
- Patterns of spending.

4.2 Model Selection and Training Strategy

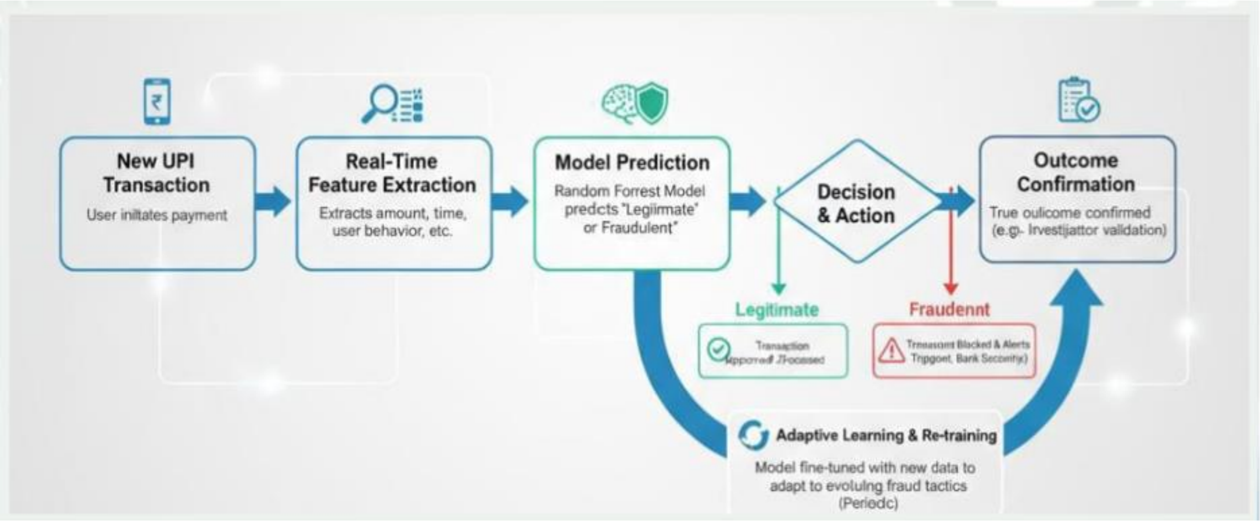
The recommended approach applies the comparative method to find out which machine learning algorithm is the best fit for the live fraud detection.

Model Set: The study restricts itself to benchmarking the performance of the following four algorithms of machine learning:

- Random Forest (RF)
- Logistic Regression (LR)
- Decision Tree (DT)
- Support Vector Machine (SVM)

Learning Techniques: The transaction data is subjected to the application of a combination of supervised, unsupervised, and semi-supervised learning techniques by ML models to uncover unusual patterns which might be indicative of fraud.

Anomaly Detection: The strategy is to combine traditional anomaly detection techniques with cooperative methods to achieve a substantial improvement in the overall fraud detection precision.


New UPI Transaction
User initiates payment
Real-Time Feature Extraction
Extracts amount, time, user behavior, etc.
Model Prediction
Random Forrest Model predcts "Legiirmate" or Fraudulent"
Decision & Action
Outcome Confirmation
True oulicome confirmed (e.g- Irvestijattor validation)
Legitimate
Transaction Appored Processed
Fraudennt
Trnasomt Blacked & Alerts Trpgont, Bank Secority)
Adaptive Learning & Re-training
Model fine-tuned with new data to adapt to evoluing fraud tactics (Periodc)

4.3 Deployment and Performance Goals

The last element of the methodology relates to the operational goals and the deployment architecture, the theoretical model training being surpassed and leading to a high-impact, scalable system.

- Real-Time Monitoring: The system has been designed to include real-time monitoring mechanisms as well as adaptive learning strategies. Thus, the system's ability to react quickly to new fraud tactics, which is a major advantage over the static traditional systems, is enhanced significantly.
- Performance Metrics: The methodology specifies a precise performance target, aiming at the accuracy of more than 90% in fraud classification, together with a very high capacity for expansion to cope with the rising number of UPI transactions.
- Integration: The methodology is planned to be installed in digital payment gateways or banking applications in the future, where it would be able to monitor and prevent fraud continuously.

4.4 Future Directions

The methodology is basically flexible and has inbuilt provisions for future improvement to continue its long-term effectiveness against changing cyber threats:

- Deep Learning Adoption: The scheme is open to the future adoption of sophisticated techniques such as Deep Learning (e.g., LSTM, Neural Networks).
- Blockchain Integration: To provide more security for data and the credibility of transactions, the idea of integration of blockchain-based verification is proposed for upcoming versions of the system.

1. Dataset and Implementation

5.1 Dataset Selection and Pre-processing

The dataset quality and its relevance are the main factors determining the machine-learning model's ability and performance.

- Dataset Source: This research has a unique dataset from the Kaggle platform that has a lot of great datasets.
- Selection Criteria: The selection process was mainly based on criteria such as the completeness, relevance, and complexity to the UPI fraud detection capabilities of the dataset.
- Data Format: The data of the project was in the CSV file format, thus making the data easily and clearly readable and, in turn, the analytical tools applications' integration was controlled and this had a positive effect on the overall research.
- Data Segmentation: Before the training of the model, the dataset was split for a precise measurement of the machine learning algorithm performance. This involved splitting the dataset into two parts: one part consisting of 80 percent of the randomly selected sequences for training and the other part of 20 percent for testing.

5.2 Model Implementation and System Architecture

The implementation phase is mainly about creating the proposed comparative machine learning technique, and after that, the final predictive model will be integrated into a working application framework.

5.2.1 Overall System Workflow

The system architecture outlines a linear procedure for classification in line with the method suggested:

1. Data Collection: The first step is the collection of transaction data (CSV File) as the initial point.
2. Data Pre-processing: The raw data goes through a process of purification and is then transformed into a format that can be analyzed, this process is called data pre-processing.
3. EDA Analysis: The data undergoes Graphical and statistical methods (Exploratory Data Analysis) to obtain insights about it.
4. Training & Testing: The data sets are split up, and the classifier models are trained on the first portion and tested on the second portion.
5. Model Implementation: The preferred machine learning approach is employed for the activation.
6. Prediction: The output of the classification is produced.
7. Web Application: The model is added to a web framework that delivers immediate results.

5.2.2 Model Training and Selection

- Machine Learning Techniques: Different kinds of learning such as supervised, unsupervised, and semi-supervised are to tell the truth applied to uncover the fraud patterns that rarely occur through a very meticulous analysis of the transaction data.
- Algorithm Focus: The deployment of the machine learning algorithms for fast and effective instance classification will be the prime approach. Gini Impurity and Information Gain are calculated and applied at every stage of the decision tree construction process of the classification task.
- Performance Metrics: The models were afterward subjected and separated by the durability of the techniques evaluated through the principal metrics of Accuracy, Precision, Recall, F1-score, and Area Under the Curve (AUC).

5.2.3 Web Application and Real-Time Monitoring

- Dynamic Environment: By making use of the real-time monitoring mechanisms along with the adaptive learning algorithms, the system can be said to create an environment that is very dynamic and responsive. This feature plays a vital role in the rapid development of countermeasures against newly devised fraud techniques and in the support of quick detection.
- System Prototype: A prototype of UPI fraud detection with a cyber-themed UI has been successfully created.
- Visualization: The prototype includes a dashboard that is user-friendly as well as interactive. It is showing through charts the model's accuracy and performance results. The web application framework outputs the ultimate prediction as "fraud or non-fraud."

6. Results And Discussion

By applying machine learning models to UPI fraud detection, the most traditional methods were outclassed, and the new models turned out to be much better in terms of accuracy, real-time capabilities, and scalability. The comparative evaluation was able to confirm the effectiveness of ensemble learning methods in revealing intricate fraud patterns.

6.1 Comparative Model Performance

The classification metrics that were the most important determined the performance of the four chosen machine learning methods, namely, Random Forest, Logistic Regression, Support Vector Machine, and Decision Tree. The results, shown in Table 1, indicate that the Random Forest (RF) algorithm's performance was by far the highest, thereby marking it as the top model for this field.

Table 1. Comparative Performance of Machine Learning Algorithms

Algorithm	**Accuracy**	**Precision**	**Recall**	**F1-score**	**AUC**
Random Forest (RF)	**0.94**	**0.95**	**0.90**	**0.85**	**0.87**
Support Vector Machine (SVM)	0.92	0.94	0.88	0.82	0.85
Logistic Regression (LR)	0.90	0.92	0.85	0.80	0.82
Decision Tree (DT)	0.88	0.91	0.86	0.78	0.80

The Random Forest model stood out with the highest values of Accuracy (0.94) and Precision (0.95), thus indicating that the model is very powerful and very accurate in detecting fraud.

6.2 Improvement Over Conventional Models

The Machine Learning approach put forward showed an indisputable functional superiority over the traditional models that are already in use, especially regarding the most important operational aspects:

Feature	**Existing Traditional Models**	**Proposed Approach**
Accuracy	70%-80%	**>90%**
Real-Time Processing	Limited	**Yes, with real-time monitoring**
Scalability	Limited	**High**

6.3 Discussion of Findings

- Superior Accuracy and Adaptability: Machine learning introduction has underlined the change in protecting digital transaction security in a huge way. The very high accuracy attained especially by the Random Forest model, is due to its power to process complex transaction data and to detect hidden patterns like a human, which was beyond the capability of the existing systems.
- Real-Time and Proactive Defense: The ML algorithms rely on the real-time data of transactions and the behavior of users to discover the anomalies and the possibly fraudulent activities very fast. The system's ability to continuously learn is what makes it able to reinforce its defenses against the fraud schemes that keep on changing, thus providing a proactive defense against the threats that are constantly changing.
- Scalability and Integration: The machine learning system by automation of the detection process can detect suspect activity much faster, thus it can scale up to deal with the higher transaction volume. The transition is so smooth that the system is always powerful, modern, and well-integrated, thereby giving a significant enhancement to the security of the UPI transactions.
- Prototype Validation: The initial tests performed on the system prototype confirmed the data flow and the functioning of the components successfully. The system executed sample fraud transactions with the expected accuracy and no run-time problems, thereby validating the initial feasibility of the model. The final prototype combines a crime-based user interface and a dashboard for visualization of performance and model accuracy.

7. Conclusion And Future Work

7. 1 Conclusion

The study has made it clear that machine learning (ML) models can be considered as a primary and flexible solution for UPI fraud detection and that they are by far better than the traditional rule-based systems. The comparison pointed out that the Random Forest (RF) algorithm is the best model with 94% accuracy and high precision thereby proving its worthiness for high-risk financial security areas.

The primary results of this initiative are as follows:

- Improved Accuracy: A system based on machine learning has been developed that can identify fraud with 90% or greater accuracy, which is a remarkable uplift from the previous techniques' ceiling of 70%-80% accuracy at most.
- Universal and Upgradable System: The capability of the system to swiftly cope with and conquer even the most sophisticated frauds thanks to real-time monitoring and adaptive learning has become the solution to the main scalability issues that have been the previous systems' curse and, eventually, the cause of their demise.
- Prototype Development: The development and testing of a working prototype with a cyber-themed user interface (UI) and an interactive dashboard for visualization of model performance has been successfully conducted, hence proving the practicability of the proposed solution.

- The suggested system makes UPI security more robust, facilitates faster counteraction of new fraud methods as well as helps the financial sector to secure their digital payment transactions thus building customer trust.

7.2 Future Work

The areas mentioned below, which all aim at strengthening the UPI fraud detection system in terms of robustness, security, and performance, are the directions for research and development in the future:

Advanced Feature Engineering and Data Integration:

- The addition of more real-time UPI transaction data will enable the model to get a better understanding of the difficult and changing fraud patterns and hence the performance will be improved.
- Anomaly detection specificity will be increased by the integration of advanced features like user location, device ID, and longitudinal spending patterns into the model training.

Integration and Deployment:

- The planned model is to be incorporated into an active UPI simulation interface and the response times of real-time alert systems for flagged transactions will be evaluated according to the actual world conditions.
- The implementation of the solution is going to take place in digital payment gateways or banking applications with the main objective of watching and stopping the frauds instantaneously.

Deep Learning and Distributed Ledger Technology are being explored together:

- The application of Deep Learning methods such as Long Short-Term Memory (LSTM) networks along with different varieties of Neural Networks is to analyze sequential transaction data and probably achieve a higher degree of predictive accuracy.
- Consider the merging of blockchain-based authentication systems for enhanced data security, immutability, and to some degree, less fraud.

REFERENCES

[1] Narendra Kumar, et al. (2020-10-13). product overview.

[2] Mohapatra S., et al. (2017). Unified payment interface (upi): a cashless Indian transaction process., International Journal of Applied Science and Engineering, vol. 5, pp. 29-42, 6.

[3] Nguyen K. (2021-03-13). What is a POS transaction? the basics explained.

[4] Kumar R., Kishore S., Lu H., Prakash A. (2020, August). Security analysis of unified payments interface and payment apps in India, in 29th USENIX Security Symposium (USENIX Security 20), pp. 1499-1516. USENIX Association.

[5] Chatterjee D. A. and Thomas R. (2017). Unified payment interface (UPI): A catalyst tool supporting digitalization - utility, International Journal of Innovative Research and Advanced Studies (IJIRAS), vol. 4, no. 2, pp. 192-195.

[6] Lakshmi K., Gupta H., and Ranjan J. (2019). Security analysis and enhancements of UPI based mobile banking applications, 2019 Amity International Conference on Artificial Intelligence (AICAI), pp. 1-6.

[7] Mohan S. (19-10-2020). What does mpin mean in the context of UPI?

[8] Hunt R. (2001). Public key infrastructure and digital certification infrastructure, in Proceedings. Ninth IEEE International Conference on Networks, ICON 2001. pp. 234-239.

[9] Eldefrawy M. H., Alghathbar K., and Khan M. K. (2011). OTP-based two-factor authentication using cellphones, in 2011 Eighth International Conference on Information Technology: New Generations, pp. 327-331.

[10] Mira Weller. (18-11-2020). A guide to android app reverse engineering 101.

[11] Misra, Anmol D. (24-11-20202). Reverse engineering android applications.

[12] Chandavarkar, B. R. (15-10-2020). How to transact using BHIM.

[13] Koh, Y. L. (July 2018). Investigating potentially harmful applications on Android.

[14] Base-Bursey, M. (26-11-2020). The permissions of mobile applications - a review of android app permissions.

[15] Dr. Virshree Tungare. (2019). A study of customer perceptions regarding UPI (Unified Payment Interface) - a breakthrough in the mobile payment system, International Journal of Science and Research (IJSR).