International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Fraud Detection in UPI Transactions

Priyansh Agrawal[1], Sahil Garg[2], Sapna Gupta[3]

[1,2,3]Dept. Information Technology And Engineering, Maharaja Agrasen Institute Of Technology, Delhi, India

ABSTRACT

The introduction of Unified Payments Interface (UPI) has changed the digital transaction landscape in India, notwithstanding the possibility of several fraudulent operations. The use of machine learning (ML) techniques for UPI fraud detection is covered in brief here. In order to identify unusual patterns that might point to fraudulent activity, machine learning (ML) models examine transaction data using a combination of supervised, unsupervised, and semi-supervised learning techniques. As essential to optimizing the mod- els' performance, effective feature selection and engineering techniques further improve the process. Additionally, combining anomaly detection algorithms with cooperative techniques improves the accuracy of fraud identification. The resilience of UPI security is increased, and quicker responses to new fraud techniques are made possible by the implementation of real- time monitoring mechanisms and adaptive learning strategies. Financial institutions can enhance the security of UPI transactions and preserve their integritywhile fostering user trust by utilizing machinelearning capabilities.

Keywords: IMPS, BHIM, NPCI, UPI andMPIN

1. INTRODUCTION

The Unified Payments Interface (UPI) has become a disruptive force in India's digital pay- ments sector, reshaping the way financial transactions are conducted throughout the nation's varied economic environment [1]. The UPI system greatly increases the speed and ease of financial transactions by enabling quick and easy fund transfers between people, companies, and financial institutions. But in addition to being widely used and successful, UPIs have also brought about several new difficulties, mainlyin the form of fraudulent activity that seriouslyjeopardizes user and financial institutionsecurity and trust [10].

Detecting and mitigating UPI fraud cases effectively has become possible in this dynamic environmentthrough the use of machine learning algorithms. Of these algorithms, Random Forest has garnered a lot of interest because ofits strong and precise fraud detection capabilities [13]. Using the combined predictive strength of several decision trees, Random Forest is an ensemble learning technique that improves regression and fraud detection accuracy [5]. With a high degree of accuracy, Random Forest can analyze complex transaction data, intricate patterns, and anomalies that point to fraudulent activity by harnessing the power of ensemble learning [9]. UPI enables Virtual Payment Address (VPA), as opposed to tradi- tional online payment methods that require users to provide sensitive information such as IMPS, NEFT, account number, and IFSC code: A VPA is an ID. Users link their mobile devices to the app to create their accounts as a payment signal for the Money transaction (from the bank) [11]. Two-factor authentication (2-FA) with a

singleclick Fundamentals of UPI Operation Utilize
or scan QR codes VPA numbers for in-person transactions Easy-to- use payment options that shield us from financial difficulties when the order is delivered. Accepted payment methods include P2P (Peer-to-Peer) transactions, e- commerce payments, and scheduled transaction-requests [4,15].

2. LITERATURE REVIEW

Sagar Suresh et al. every nation's economy is built on regulations. The fintech sector is expanding worldwide, including in India. The government has launched a number of sig-nificant initiatives to support the growth of this still-developing industry since 2010. The National Payments Corporation of India (NPCI), Aadhaar, Jan Dhan yojana, and the Goods and Services Tax (GST) are a few of these initiatives [6]. In a variety of industries, including lending (more than 100), personal finance management (more than 40), and investment management (more than 90), FinTech startups have proliferated [14]. The per- centage of Indians who used the Internet increased from 7.5% in 2010 to 34.4% in 2017, and the number is expected to rise from 437.4 million in 2017 to 666.4 million in 2023.The impact of regulatory changes on FinTech is examined in this research paper. Industry from the standpoint of development [7].

S. Mohapatra et al. India's payment systems have changed dramatically in the last several years. The Reserve Bank of India produced a vision plan for payment systems that outlines a "less cash" society. Two key components of this initiative are the acceptance of card and mobile payment technologies [12]. This article provides an overview of the newly announced restructuring initiative of the National Payments Corporation of India (NPCI). The goal of this article is to teach banking customers how to use business-to-consumer (B2C) and person-to-person (P2P) payments to "send" and "collect" money in real time combination [3]. There is also a comparison analysis of different UPI apps in this article. India's payment systems have changed dramatically in the last several years. Numerous steps are outlined to establish a "less cash" society in a vision document on payment systems published by the Reserve Bank of India [8]. Acceptance of card and mobile pay- ment systems are two essential elements of this project. An overview of the National Payments Corporation of India's (NPCI) recently announced transformation program is given in this article [13]. The goal of this article is to teach banking customers how to use business-to-consumer (B2C) and person-to-person (P2P) payments to "send" and "collect" money in real time. Combination. There is also a comparison analysis of different UPI apps in this article [2].

3. PROPOSED METHODOLOGY

The suggested solution aims to increase UPI fraud detection capabilities by optimizing the Random Forest algorithm's performance as shown in Figure 1. Through ensemble learn- ing, the predictive potential of numerous decision trees is harnessed, enabling this approach to detect fraudulent transactions with high certainty and accuracy. The Random Forest method looks for odd patterns in transaction data that can point to fraud by using supervised learning techniques. Because of its clustering skills, it can handle complex data structures and correlations with efficiency, which is its main strength. In order to provide a dynamic and responsive environment, the system also incorporates real-time monitoring and adaptive learning algorithms. This flexibility guarantees quick reaction to newly developed deception tactics as well as prompt identification and suppression of fraudulent activity.This all-encompassing strategy seeks to improve fraud detection accuracy while

fortifying the system's resistance to new attacks in the quickly evolving field of electronic transactions.

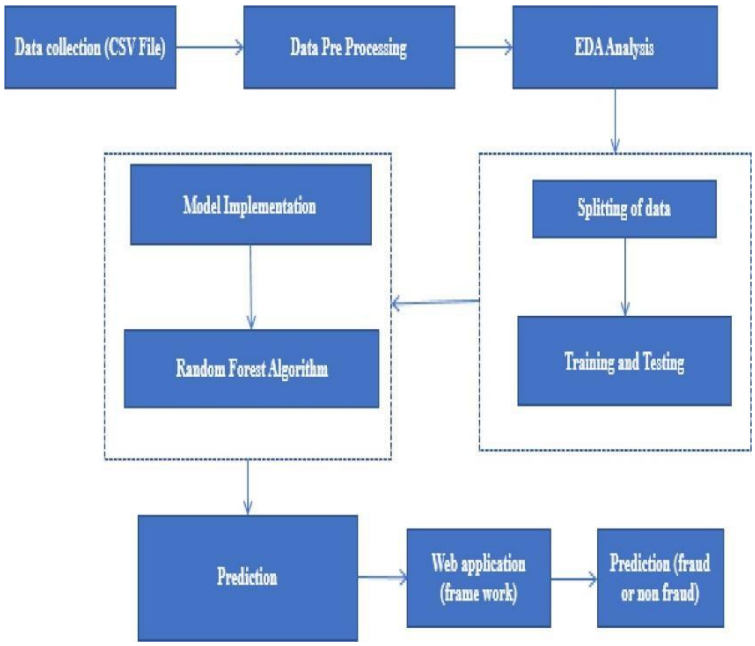

Figure 1. Proposed methodology.

4. FEATURE EXTRACTION

Future research on machine learning-basedUPI fraud detection will employ a thoroughapproach in order to further enhance systemperformance.To increase accuracy anddecrease false positives, feature engineering and algorithm optimisation will be the main areas of emphasis for machine learning models' ongoing development. Ensuring openness in the decision-making process through the investigation of comprehensible AI techniques and offering insights into the models' ways of fraud prediction generation would be a crucial component of this strategy. Additionally, by addressing security andprivacy concerns, integrating cutting-edgetechnology like blockchain into UPI frauddetection systems can increase overallresilience. The integrity of transaction data canbe further guaranteed by taking advantage of the decentralised and intrinsic immutability of blockchain technology. Industry-wide cooperation will be necessary to provide reliable and flexible defence against new fraud schemes. It's crucial that UPI transactions are dynamic. Resolving resource optimisation and scaling concerns to guarantee the smooth implementation of machine learning models throughout the extensive and dynamic digital transaction landscape Integration of cutting- edge technologies, cooperation amongst industries, and resolution of logistical issues. In addition to increasing fraud detection accuracy, this all-encompassing approach seeks to build a more resilient, secure, and flexible framework for protecting UPI transactions in the dynamic digital environment.

5. MODULE DESCRIPTION

5.1.Dataset selection

The dataset that served as the basis for this study was carefully chosen from among the enormous and varied collection of datasets available on the Kaggle platform. This processnecessitated paying

close attention to a varietyof factors, including the dataset's complete- ness, relevance, and use for addressing thecomplexity of the study's objectives. Bymaking use of Kaggle's extensive resources, which give access to a wide and diverse pool of data, it may be investigated further. Byusing the CSV file format, the project uses a standardized and comprehensible approach to data management that increases the efficacy of the research study overall and facilitates integration with analytical tools.

5.2. Pre-processing of data

Classifying this data is necessary after the Twitter stream has been extracted into datasets. Before analyzing the input, the classifier eliminates unnecessary information from it, such as stop words and emoticons. It promises to identify and remove non-textual information. Raw feature vectors are the mostappropriate representations for downstream estimators when using sk-learn. Train-test segmentation is a technique used to assess how well a machine learning algorithm performs. This also holds true for difficulties posed by supervised learn- ing methodologies. It can be applied to issues involving regression or classification and used with any supervised learning technique. The process separates the data set into two sections. Taking a random sample of a portion of the data is the simplest and most widely used technique for segmenting such a data collection. Using a random selection method, one alternativestrategy is to use 80 percent of the data set's sequences for training and the remaining 20 percent for testing.

5.3. Model implementation

For this section, a machine learning algorithmakin to a random forest for this project tocategorize the samples which gives the abilityto handle complex datasets effectively.

5.3.1. Random forest algorithm

This algorithm is a well-known method thatcombines multiple decision trees to make pre- dictions for classification or regressiontasks. During training, it generates a lot of decision trees, from which the meanprediction for regression analysis or theclass mode for classifi- cation are extracted.

5.3.2. Gini impurity

Gini impurity quantifies the degree of ambiguity or impurity in a collection of labels, cal- culated as

$$Gini(D) = 1 - \sum_{i=1} p_i^2 \quad (1)$$

5.3.3. Entropy

Entropy quantifies the impurity or uncertaintyof a node

$$H(D) = -\sum_{i=1} p_i \log_2 (p_i); \quad (2)$$

5.3.4. Information gain

The reduction in entropy obtained by dividingthe data according to a certain attribute is measured as information gain which is computed as

$$IG(D; A) = H(D) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} \times H(S_v); \quad (3)$$

Entropy calculates the entropy. Finally, forclassification tasks, the final prediction is determined by the mode (most common) classlabel among the decision trees.

5.4.Prediction

At the forefront of fraud detection, the moduleunder review effortlessly integrates data from many mechanisms. It effectively finds anomalies that point to possible fraudulent conduct within the context of UPI transactionsby combining the strength of severalalgorithms. With this cutting-edge method, anomalies can be detected and dangers can be promptly reduced by taking preventative action. The module guarantees a proactive approach against emer- ging risks and enhances the overall security of UPI transactions through constant monitor- ing and analysis. Through the integration ofdiverse algorithms, a resilient and adaptablesecurity system is produced. The inclusion of these results highlights the module's dedicationto enhancing the security environment and giving users and stakeholder's faith in the reliability of UPI transactions.

6. RESULTS

This section highlights the usefulness of random forest for detecting upi fraud andtheir potential to improve security in digital payment systems.

Machine learning has shown promising results in detecting UPI fraud, with a notable decrease in false positives and accurate identification of questionable transactions. This section evaluates the efficacy of several machine learning techniques for detecting UPI fraud. It compares the performance of many methods, including Random Forest, Logistic Regression, SVM, and Decision Tree. The evaluation was done using important measure line accuracy, precision, recall, F1-score, and AUC as shown in Figure 2.

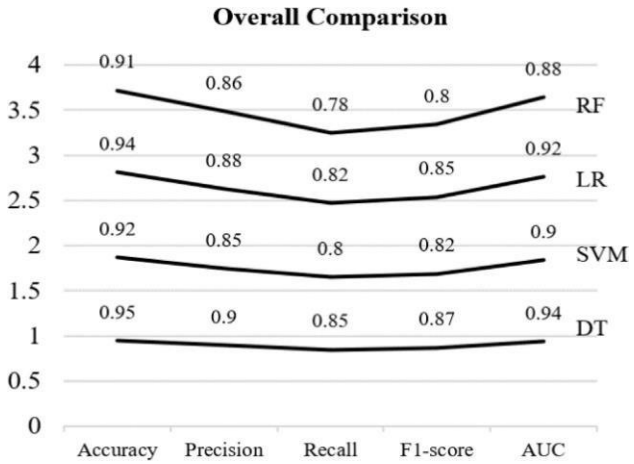

Figure 2. Performance comparison.

Table 1. Performance of testing.

Algorithm		Accuracy	Precision
RF		0.94	0.95
LR		0.90	0.92
SVM		0.92	0.94
DT		0.88	0.91

7. CONCLUSION

The way that digital transaction security isprotected has changed dramatically with the introduction of machine learning into UPI fraud detection. Machine learning algorithms leverage real-time transaction data and user behaviour analysis to promptly discover anomalies and potentially fraudulent activity. These models are unique in that they can learncontinuously, which allows them to increase their defences against ever-evolving fraud schemes. This flexibility guarantees that the system is strong and current, offering a proac- tive defence against changing threats. Because machine learning models are flexible, they may be seamlessly integrated, increasing the overall security of transactions made via a single payment interface. These algorithms detect suspicious activity more quickly and scale well to accommodate growing transaction volumes since they automate the detection pro- cess. In the end, machine learning's precision, flexibility, and real-time capabilities can greatly aid in protecting UPI transactions and fortifying the financial system against the always changing array of cyber threats.

A reliable and adaptable way to identify and stop fraudulent activity is to incorporate the Random Forest Algorithm into UPI fraud detection. With the help of its ensemble learning capabilities, this system effectively evaluates transaction data from the past to predict pos- sible fraud cases. Random Forest stands out due to its versatility, which enablesit to effec- tively manage asymmetric datasetsand a variety of data sources. Users' confidence in UPI security is increased as a result of the system's rapid detection of abnormalities that signify fraudulent behaviour, which greatly increases the dependability of digital transactions. To keep UPI security measures resilient and ahead of changing threats, Random Forest must be continuously evolved and integrated withfraud detection systems. Future research in thisfield should put an emphasis on enhancing thealgorithm's performance, making sure it can adjust to new fraud strategies, and investigating novel machine learningtechniques. Together, these initiatives will improve UPI fraud detection capabilities,hence enhancing digital transaction security and developing a more dependable and safer UPI ecosystem.

REFERENCES

1. Narendra Kumar,etal. (13-10-2020).Product overview
2. https://www.npci.org.in/what-we-do/upi/product-overview.
3. Mohapatra S., et al. (2017). Unified payment interface (upi):a cashless indian transaction process.,
4. International Journal of Applied Science andEngineering, vol. 5, pp. 29–42, 6.
5. Nguyen K. (13-3-2021). What is pos transaction? the basics explained. https://blog. magestore.com/pos-transaction/, 2021.
6. Kumar R., Kishore S., Lu H., and Prakash A.(Aug. 2020). Security analysis of unified payments interface and payment apps in india, in 29th USENIX Security Symposium (USENIX Security 20),
7. pp. 1499–1516, USENIX Association.
8. Chatterjee D. A. and Thomas R. (2017). Unified payment interface (upi): A catalyst tool supporting digitalization – utility, prospects and issues, International Journal of Innovative Research and Advanced Studies (IJIRAS), vol. 4, no. 2, pp. 192–195.
9. Lakshmi K., Gupta H., and Ranjan J. (2019).Upi based mobile banking applications – security ana- lysis and enhancements, in 2019 Amity International Conference on Artificial

Intelligence (AICAI), pp. 1–6.
10. Mohan S. (19-10-2020). What is the mpin inupi?. https://razorpay.com/learn/ generate-upi-pin/, 2019.
11. Hunt R. (2001). Pki and digital certification infrastructure, in Proceedings. Ninth IEEE International Conference on Networks,ICON 2001. pp. 234–239.
12. Eldefrawy M. H., Alghathbar K., and Khan
13. M.K.(2011).Otp-based twofactor authentication using mobile phones, in 2011 Eighth International ConferenceonInformation Technology: New Generations, pp. 327–331.
14. Mira Weller. (18-11-2020). android appreverseengineering 101
15. https://maddiestone.github.io/AndroidAppRE/.
16. Anmol Misra D. (24-11-20202). Reverse engineering android applications.https://www.oreilly.com/ library/view/android- security/9781439896464/.
17. Chandavarkar B. R. (15-10-2020). How to transact using bhim.https://www.bhimupi.org.in/.
18. Koh Y. L. (07 2018). Investigating potentially harmful applications on android.
19. Base-Bursey M. (26-11-2020). What are apppermissions - a look into android apppermissions. https://
20. www.wandera.com/mobile-security/app-and-data-leaks/app-permissions/..
21. Dr. Virshree Tungare. (2019). A study on customer insight towards upi (unified paymentinterface) - an advancement of mobile payment system, International Journal of Science and Research (IJSR).