ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

UPI Fraud Detection using Machine Learning

K.DivyaTeja[1], K.V.Vaishnavi[2], M.Srija[3], M.RamSaiTejaswini[4].
Chaitanya Bharathi Institute of Technology, Proddatur, A.P, India.

vaishnavi6746@gmail.com[2], msrijahoney@gmail.com[3], tejaswinimanyam4@gmail.com[4]

ABSTRACT

With the rapid adoption of Unified Payments Interface (UPI) for digital transactions, the risk off fraudulent activities has also increased significantly. To address this challenge, we propose a novel approach to detect UPI fraud by analyzing transaction details such as the bank book name, transaction ID, and transaction amount. Our method employs three machine learning algorithms: Random Forest, K-Nearest Neighbors (KNN), and Decision Tree. The Random Forest classifier is known for its accuracy and resilience against overfitting, making it a robust choice for this application. It processes the provided transaction details to classify the transaction outcome as either "Transaction Failed: Incorrect Details Entered" or "Transaction Successful: Details Verified and Processed. “In addition to Random Forest, the KNN algorithm provides a straightforward approach to classification based on the similarity of transaction details with previously labeled data. The Decision Tree algorithm, with its intuitive structure, offers clear decision-making pathways, enhancing interpretability. This multi-algorithm approach not only helps in preventing fraudulent transactions but also ensures that legitimate transactions are processed smoothly. The proposed model aims to enhance the security of UPI transactions, providing users with an additional layer of protection against unauthorized activities. Initial evaluations suggest that these algorithms effectively distinguish between genuine and fraudulent transactions, demonstrating their potential for integration into real-world financial systems.

Keywords: UPI Digital Payments, Random Forest Algorithm, K-Nearest Neighbors, Decision Tree, Machine Learning

I. INTRODUCTION

With the rapid adoption of the Unified Payments Interface (UPI) for digital transactions, the associated risk of fraudulent activities has surged significantly. To combat this challenge, we propose a robust approach to detect UPI fraud by analyzing critical transaction details, including the bank book name, transaction ID, and transaction amount. Our method employs three machine learning algorithms: Random Forest, K-Nearest Neighbors (KNN), and Decision Tree. The Random Forest classifier is renowned for its accuracy and resistance to overfitting, making it a strong candidate for distinguishing between legitimate and fraudulent transactions. KNN offers an intuitive distance-based classification, while the Decision Tree provides a clear, interpretable model that facilitates decision-making. This ensemble of techniques classifies transaction outcomes as either "Transaction Failed: Incorrect Details Entered" or "Transaction Successful: Details Verified and Processed.”

II. LITERATURE SURVEY

Table 1. Comparative Literature Review of Fraud Detection Techniques and Their Research Outcomes (2000–2017)

S.No	Year	Authors	Title	Outcomes

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

1.	2000	ARTÍS M, Ay USO M, GUILLÉN M	Modeling different types of automobile insurance fraud behavior in the Spanish market	This study analyzes repeated choices in expensive automobile insurance policies using a four- year dataset, revealing significant factors like engine capacity, vehicle make, and consumer inertia in policy selection.
2.	2008	Hand C, Whit row DJ, Adams C, Juszczak NM, Weston P D	Performance criteria for plastic card fraud detection	The paper discusses optimizing predictive data mining algorithms for fraud detection in plastic card transactions. It defines two performance measures related to cost minimization and fraud reduction, and introduces a Plot for algorithm performance evaluation.
3.	2010	Becker RA, Volinsky C, Wilks AR	Becker RA, Volinsky C, Wilks AR	This paper explores the evolution of fraud detection at AT&T, highlighting key fraud schemes and techniques, advocating for simple models, visualization, flexible environments, and the importance of data management and human involvement.
4.	2016	Abdallah A, Maa rof MA, Zainal A	Fraud detection system: A survey. J Netw Computer Application	This survey paper examines issues in fraud detection systems (FDSs) within e- commerce, exploring challenges like concept drift and real time detection, and discusses trends and approaches across five e- commerce systems.
5.	2017	Sahin M	Understanding Telephony Fraud as an essential Step to Better Fight it	SIM Box fraud in tele communication causes $3-$7 billion in annual losses by redirecting illegitimate VoIP traffic. This study investigates its impact, detection approaches, and flawsfrom1994to 2021.

III. EXISTINGSYSTEM

K-Nearest Neighbors(KNN)is a simple, non-parametric algorithm used for both classification and regression tasks. It operates by finding the 'k' closest data points (neighbors) to a query point and making predictions based on the majority class(for classification)or average value(for regression)of the se neighbors. The algorithm is intuitive and easy to implement but can be computationally expensive, especially with large datasets, because It requires calculating the distance between the query point and all other points in the dataset. Additionally, KNN's performance can be heavily influenced by the choice of' k' and the distance metric used, and it is sensitive to irrelevant or noisy features, which can mislead the distance calculations and reduce its accuracy. Simplicity the quality of being easy to understand or use, with minimal complexity or clutter. Ersatility the ability to adjust to new or changing situations, environments, or requirements. Effectiveness optimizing the use of time, money, materials, or effort to achieve the desired outcome.

Disadvantages:

KNN can be slow with large data because it compares the query with every data point.
KNN can be in accurate if the data set has irrelevant or noisy data, affecting distance calculations and predictions.
KNN struggles with many features because it finds it hard to identify close point accurately.

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

IV. PROPOSEDSYSTEM

Random Forest is a powerful machine learning algorithm that can be highly effective in detecting UPI fraud. It operates by building multiple decision trees during training and combining their predictions to make a final decision. This ensemble method improves accuracy and reduces the risk of overfitting, making it well suited for complex tasks like fraud detection where the data can be noisy, imbalanced, and involve intricate relationships between features. In the context of UPI fraud detection, Random Forest can analyze various transaction attributes such as the bank book name, transaction ID, and amount, identifying patterns that distinguish between legitimate and fraudulent transactions. By leveraging the combined power of multiple trees, Random Forest can capture the subtle indicators of fraud that might be missed by simpler models.

Advantages:

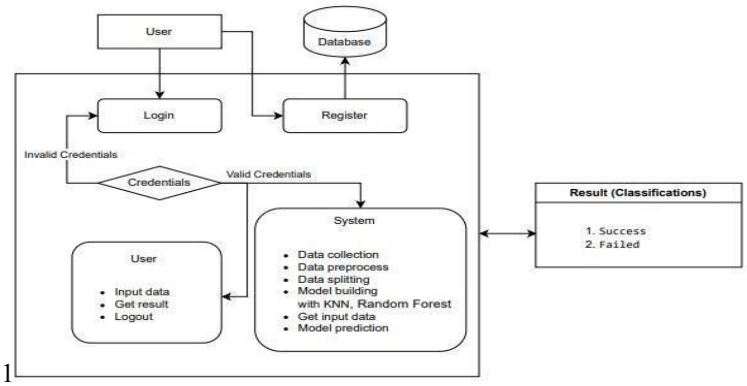
Robustness Against Overfitting: Random Forest's ensemble approach mitigates overfitting, making the system resilient to noisy and imbalanced transaction data.

Interpretability: Decision Trees provide clear decision-making pathways, allowing users to understand the rationale behind classification outcomes, which is crucial for trust in financial systems.

Real-time Detection: The system can analyze transaction details in real time, enabling swift identification and prevention of fraudulent activities.

Flexibility: The multi-algorithm approach can adapt to evolving fraud patterns by retraining with new data, ensuring continued effectiveness.

ARCHITETURE:



1

Figure 1 : Shows Architectural flow

MODULES:

Working on dataset: System checks for data whether it is available or not and load the data in csv files.

Pre-processing: Data need to be pre-processed according the models it helps to increase the accuracy of the model and better information about the data.

Training the data: After pre-processing the data will split in to two-parts as train and test data before training with the given algorithms.

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

Model Building: To create a model that predicts the personality with better accuracy, this module will help user.

Generated Score: Here user view the score in%

V. CONCLUSION

The increasing popularity of Unified Payments Interface (UPI) has made digital payments more convenient but has also exposed users and financial institutions to a higher risk of fraudulent activities. To mitigate these risks, this project utilizes a multi-algorithm approach, employing Random Forest, K-Nearest Neighbors (KNN), and Decision Trees to analyze transaction data and detect fraudulent behavior effectively. Each algorithm brings unique strengths to the table: Random Forest enhances accuracy by preventing overfitting, KNN offers straightforward classification through similarity-based analysis, and Decision Trees provide interpretable decision-making paths. By leveraging these techniques, the system can effectively identify suspicious transactions, thereby ensuring the security of UPI transactions. The ultimate goal of this project is to create a robust fraud detection system that enhances user trust and confidence in digital payment systems. With the growing reliance on UPI for everyday transactions, having an efficient and reliable fraud detection mechanism is crucial in safeguarding both users and financial institutions from potential losses. By implementing this multi-algorithm approach, the project contributes to making UPI a safer and more secure platform for digital transactions, paving the way for continued growth in the adoption of digital payments.

REFERENCES:

[1] ALESKEROVE, FREISLEBEN, B., and, RAOB(1997)CARDWATCH: A neural network-based data base mining system for credit card fraud detection. In Conference(pp. 220–226). IEEE, Piscataway, NJ
[2] Sahin M(2017) Understanding Telephony Fraud as an Essential Step to Better Fight it [Thesis]. École Doctor ale Informatique, Télé communication et Électronique, Paris Abdallah A, Maaro fMA, ZainalA(2016) Frauddetection system: A survey. J Netw Compute Appl 68:90–113
[3] ANDREW SPP, PETER SON MB (eds)(1990) Criminal Intelligence Analysis. Palmer Enterprises, Loomis, CA
[4] ARTÍSM, AyUSOM, GUILLÉNM (1999) Modeling different types of automobile insurance fraud behavior in the Spanish market. Insurance Math Econ 24:67–81
[5] BARAO MI, TAWN JA (1999) Extremal analysis of short series with outliers: Sea-levels and athletics records. Appl Stat 48:469–487 Page 14/16
[6] BLUNT G, HAND DJ (2000) The UK credit card market. Technical report, Department of Mathematics, Imperial College, London
[7] BOLTON RJ, HAND DJ (2001) Unsupervised profiling methods for fraud detection. In Conference on Credit Scoring and Credit Control 7, Edinburgh, UK, 5–7 Sept