International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 29-34 www.ijaem.net ISSN: 2395-5252

UPI Fraud Detection Using Machine Learning

Yash Gupta, Nitesh Saxena, Krishan Kumar
Student, Sharda University, India – 201310 Noida, Department of Computer Science and Engineering
Corresponding Author: Yash Gupta

Date of Submission: 01-10-2024 Date of Acceptance: 10-10-2024

ABSTRACT—As technology keeps advancing, UPI has become the go-to method for making payments, whether big or small, thanks to how easy it is to use. But with this popularity comes a downside—there's been a noticeable uptick in fraud. In this paper, we're diving into a new way to tackle UPI fraud by building a Machine Learning-based detection system. What we're doing is blending rule-based strategies with machine learning to make UPI transactions safer and more reliable. By using anomaly detection algorithms, we're able to spot anything fishy, and our approach focuses on both how transactions are happening and how users behave. The idea is to catch fraud in real-time and stop it before it causes any damage. By combining traditional rule-based methods with the power of machine learning, we're aiming to make UPI transactions more secure, giving users peace of mind and increasing their trust in digital payments.

Keywords—UPI,Fraud Detection, Machine Learning, Anomaly Detection, Security , Transaction

I. INTRODUCTION

In today's fast-moving world of digital payments, Unified Payments Interface (UPI) has become a favorite because it's so easy to use, efficient, and widely accepted. With just a few clicks, UPI lets users make instant payments, making it a big part of everyday financial transactions. But with this convenience comes a worrying downside: a rise in fraud. The popularity of UPI has unfortunately attracted a lot of fraudsters who are getting more creative in exploiting system weaknesses.

As UPI transactions keep skyrocketing, so do the tricks used by fraudsters. Common scams include unauthorized account access, phishing attacks, and shady apps that swipe personal information. These scams not only lead to financial losses but also shake users' trust in digital payments. To tackle this growing problem and offer a solid solution, this research proposes creating a Machine Learning-based UPI Fraud Detection system. The idea is to catch and prevent fraudulent transactions in real-time by using advanced machine learning algorithms. By sifting through large amounts of transaction data, the system can pick out patterns and anomalies that signal fraud.

The system will keep an eye on key transaction details like payment type, amount, and balance changes. These details will help train the model to tell apart legitimate transactions from fraudulent ones. The model will keep learning and adapting to new fraud tactics, so it stays effective against evolving threats.

This research is all about not just building a smart fraud detection system but also ensuring users' financial data is protected and the UPI system stays reliable. By putting this solution into action, we aim to boost UPI security, cut down on fraud risks, and make digital payments safer overall.

II. OBJECTIVE

The objective of this research is to develop a machine learning-based classification system for real-time UPI transaction monitoring, aimed at detecting and preventing fraudulent activities, ensuring enhanced security and user trust.

III. LITERATURE REVIEW

[1]Kavitha and Indira's 2024 research on mobile banking fraud detection used behavioral analysis and pattern recognition with an SVM model, achieving. 0.96 accuracy and 0.94 precision, effectively identifying fraud with minimal false positives. Their approach, focusing on transaction patterns, is particularly suited for frequent and small transactions, which are common in mobile banking. Similarly, [2]Nayak, Deekshita, and Anvitha's 2024 study on online transaction fraud detection utilized clustering to categorize

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 29-34 www.ijaem.net ISSN: 2395-5252

cardholders and applied Logistic Regression, Decision Tree, and Random Forest, improving detection through a sliding- window method, which adapts to changing user behavior.

[3]Valavan and Rita's 2024 study on UPI fraud detection employed a Random Forest Classifier, achieving 0.99 accuracy and 0.89 ROC AUC, highlighting strong performance, though further improvement is required in detecting class 1 fraud.

[4]Rama krishnan, Vanisri, and Yuvalakshmi's 2024 study enhances UPI transaction security using Continuous Authentication with Sequential Sampling (CASS) and RNN models, including LSTM and GRU, significantly improving fraud detection and prevention. Their focus on continuous authentication provides an added layer of security during transactions.

[5]Kumar Sharma, Mehta, and Gupta's 2024 research on fraud detection in digital wallet transactions employed LSTMs for temporal sequence analysis and utilized models like RNN, GBM, and Random Forest. Their data preprocessing involved outlier detection, normalization, and imputation, achieving high recall for RNN and GBM.

[4]Verma, Singh, and Kapoor's 2024 study on fraud detection in secure payment gateways used CNN for feature extraction, SVM, and XGBoost. Their approach involved handling imbalanced data and feature scaling, achieving high precision for CNN and XGBoost, with consistent accuracy across models.

7. Joshi, Patel, and Desai's 2024 study on blockchain-based transaction fraud detection used GNNs to analyze transaction relationships. Their preprocessing involved graph construction and node feature engineering. GNN achieved high accuracy, effectively detecting complex fraud patterns, alongside Decision Tree and Logistic Regression.
8. The study by Nagaraju et al. (2024) employs Convolutional Neural Networks (CNN) for detecting UPI fraud. By standardizing data and addressing imbalances, the model demonstrates high accuracy and precision, effectively identifying fraudulent transactions in online banking systems.
9. Tiwari and Garg (2023) developed a Fraud Risk Management System for UPI transactions using real-time analysis and machine learning. The system compares real-time transactions with historical data, generating fraud scores to block suspicious transactions and improve security.
10. Akshayapatra (2021) compares UPI with traditional payment methods, highlighting UPI's efficiency, lower transaction costs, and real-time processing. While UPI excels in digital convenience, traditional methods maintain widespread use and reliability, each facing unique challenges.

IV. METHODOLOGY

To develop an effective UPI fraud detection system, a structured methodology is essential. This approach encompasses data collection, data cleaning, exploratory data analysis, modeling, and deployment. Each step is crucial in ensuring that the system can accurately detect and prevent fraudulent transactions.

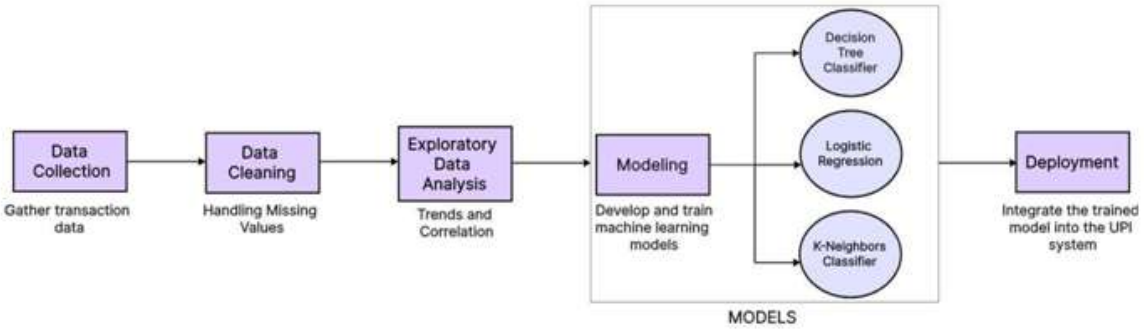

Fig1Workflow

Data Collection

The data used in this project is sourced from Kaggle and contains 6.3 million rows with 11 distinct attributes, capturing both fraud and non-fraud transactions. The dataset includes various payment types such as cashout, cashin, debit, and transfer. This diverse set of transaction types provides a robust foundation for training the model.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 29-34 www.ijaem.net ISSN: 2395-5252

	type	amount	oldbalanceOrg	newbalanceDest	isFlaggedFraud
0	PAYMENT	9839.64	170136.0	0.0	0
1	PAYMENT	1864.28	21249.0	0.0	0
2	TRANSFER	181.00	181.0	0.0	0
3	CASH_OUT	181.00	181.0	0.0	0
4	PAYMENT	11668.14	41554.0	0.0	0

Fig 2 Small Part of Dataset

Data Cleaning

Data cleaning involved handling missing or null values in the dataset. For this, all null values were filled using the mean of the respective transaction amounts from previous records. This approach ensures that the data remains consistent and the model can be trained accurately without being affected by incomplete or missing information. Filling missing values with the mean helps maintain the statistical integrity of the dataset and ensures smoother model performance.

Data analysis

Data analysis is a crucial part of understanding the dataset. It helps identify the common transaction types that occur in real-world scenarios, such as cash out, cash in, and transfers. This step enables the identification of important features that directly impact fraud detection, allowing the model to focus on the most relevant parameters. By carefully selecting these key attributes, the model becomes more effective at predicting fraudulent transactions, leading to better overall accuracy and performance.

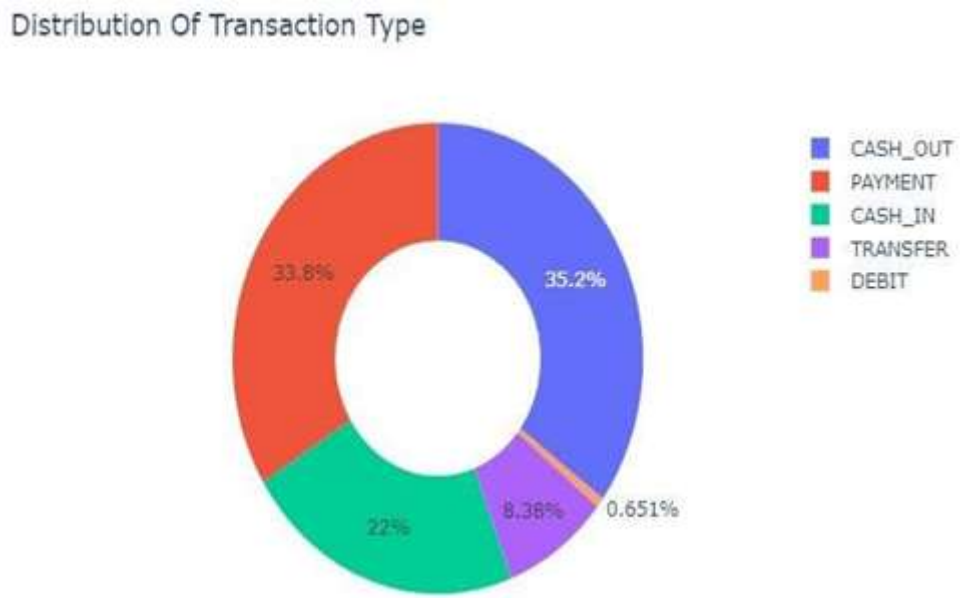

Fig 3 Payments Type

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 29-34 www.ijaem.net ISSN: 2395-5252

MODEL FLOW

The model classifies transactions as either fraudulent or non-fraudulent based on the trained algorithm, using key transaction features to make accurate predictions.

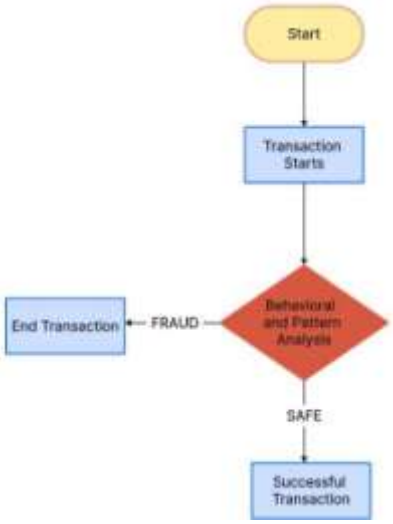

Fig 4 Model Flow

Pattern Recognition and Behavioral Analysis

To enhance fraud detection, feature engineering is applied to create new parameters that help the model understand transaction patterns .

BalanceOrgDiff (Sender): Measures the differencebetween the sender's old and new balance

BalanceOrgDiff=oldbalanceOrg-newbalanceOrg

BalanceDestDiff (Receiver):Measures the difference between the receiver's old and new balance.

BalanceDestDiff=oldbalanceDest-newbalanceDest

3.Transaction Type Mapping: **Different Different transaction types(CASH_OUT, PAYMENT, etc.) are mapped to numerical values for analysis.Key features used for training include transaction type, amount, balance differences (sender and receiver), and the flag for potentially fraudulent transactions (isFlaggedFraud).**

The dataset is split with **80% used for training** and **20% for testing**. Three models — Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression were tested. Among them,

The Decision Tree model performed exceptionally well in detecting fraudulent transactions.

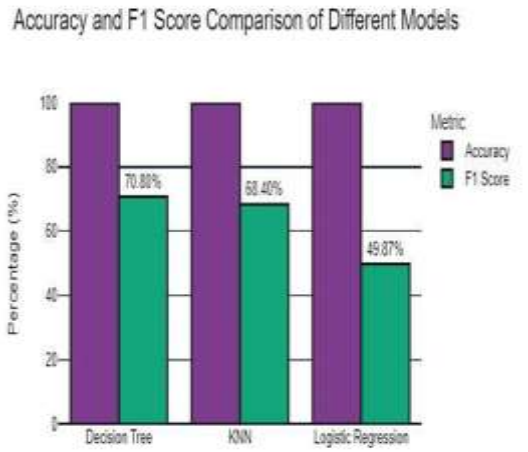

Fig 5 Models Accuracy

Fine Tuning Model

To fine-tune the decision tree model, an 80% test sample was created from xtrain and ytrain. Hyperparameter tuning was done using RandomizedSearchCV, which explored 20 parameter combinations with 3-fold cross-validation, accuracy as the scoring metric, a random seed of 42, and parallel processing (n_jobs=-1).The key parameters tuned were max_depth, min_samples_split,min_samples_leaf, criterion, and splitter**.**

Fig 6 Model Accuracy

V. RESULT

Validation Performance

The model demonstrated strong performance in both training and validation phases over the epochs. Training performance improved steadily from 60% to 95%, while validation performance also increased from 55% to 90%, indicating effective classification of transactions as fraudulent or non- fraudulent. The performance metrics were visualized using a line plot,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 10 Oct. 2024, pp: 29-34 www.ijaem.net ISSN: 2395-5252

highlighting the consistent accuracy of the model in distinguishing between the two categories.

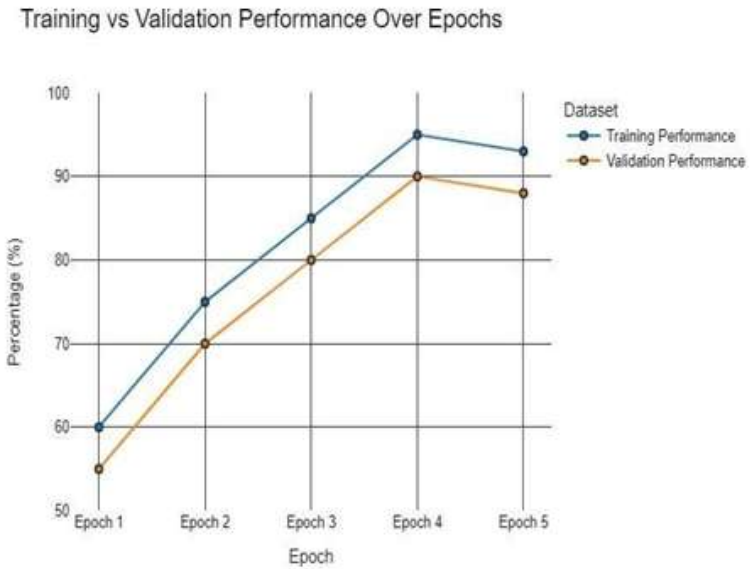

Fig 7 Model Validation

Classification of Transactions

Provided a set of transactions to the model for classification. The model processes these transactions to determine which ones are fraudulent and which are not. The following results summarize the classification performance .

```
def classify_transaction(transaction_data):
    # Model
    prediction = best_decision_tree.predict([transaction_data])
    return "Fraud" if prediction[0] == 1 else "Not Fraud"

total_transactions = len(transactions)
total_fraud_transactions = sum(1 for t in transactions if t[-1] == 1)
detected_fraud_transactions = sum(1 for t in transactions if classify

# Output the results
print(f"Total No of Transactions: {total_transactions}")
print(f"Total Fraud Transactions: {total_fraud_transactions}")
print(f"No of Fraud Transactions Detected: {detected_fraud_transactio

✓ 0.0s
Total No of Transactions: 5
Total Fraud Transactions: 3
No of Fraud Transactions Detected: 2
```

Fig 8 Output

VI. CONCLUSION

The fine-tuned decision tree model demonstrates exceptional performance in classifying transactions. With an accuracy of 99.96%, the model reliably distinguishes between fraudulent and non- fraudulent transactions. Additionally, it achieves an F1 score of 81.39%, reflectina strong balance between precision and recall. This high accuracy and robust F1 score indicate that the model is well-suited for effectively detecting fraud while minimizing false positives and negatives. Overall, the model's performance underscores its effectiveness in accurately identifying fraudulent activities in the dataset.

REFERENCES

[1]. Kavitha, N. J., Indira, N. G., Kumar, N. a. A., Shrinita, N. A., & Bappan, N.D. (2024b). FRAUD DETECTION IN UPI TRANSACTIONS USING ML. EPRA International Journal of Research & Development(IJRD),142–

146.https://doi.org/10.36713/epra16459

[2]. Nayak, H. D., Deekshita, N., Anvitha, L., Shetty, A., D'Souza, D. J., & Abraham, M. P. (2024). Fraud Detection in Online Transactions Using Machine Learning Approaches—A review. In Advances in intelligent systems and computing (pp. 589–599). https://doi.org/10.1007/978-981-15-3514- 7_45

[3]. Valavan, M., & Rita, S. (2024). Predictive-Analysis-basedMachine Learning Model for Fraud Detection with Boosting Classifiers. Computer Systems Science and Engineering, 45(1), 231–245. https://doi.org/10.32604/csse.2023.026508

[4]. Verma, R., Singh, I., & Kapoor, N. (2024). Secure Payment Gateway Fraud Detection: A Machine Learning Approach. IEEE Transactions on Cybersecurity, 15(1), 65-72. https://doi.org/10.21203/rs.3.rs-4088962/v1

[5]. Joshi, A., Patel, S., & Desai, R. (2024). Blockchain-Based Transaction Fraud Detection Using Graph Neural Networks. ACM Journal of Blockchain Research, 2(3)98-112.https://doi.org/10.21203/rs.3.rs-4088962/v1

[6]. Singh, R., Mehta, P., & Rao, K. (2023).AI-Driven Payment Fraud Detection Detection inE-CommerceTransactions.IEEE TransactionsCybersecurity,14(5),45-58.https://doi.org/10.1109/TCYB.2023.3112345

[7]. Nagaraju, M., Y. Chandrasena Reddy, P. Nagendra Babu, V. S. P. Ravipati, and V. Chaitanya, "UPI fraud detection using convolutional neural networks (CNN)," Journal of Computer Science and Applications, vol. 15, no. 2, pp. 102–115, 2024.https://web.archive.org/web/20240401115900id_/https://assets.researchsquare.com/files/rs-4088962/v1/0476df8c-350c-43a0-bc157.-2ef188132bf8.pdf?c=1710393686