**An Intelligent System for Detecting UPI Frauds Using Machine Learning**

**Ms. Manisha Patil, Dr. Tejashree T. Moharekar**

1 MCA Student, Yashwantrao Chavan School of Rural Development, Shivaji University Kolhapur, Maharashtra, India

2 Assistant Professor, Yashwantrao Chavan School of Rural Development, Shivaji University Kolhapur, Maharashtra, India

**ABSTRACT**

In recent times, the Unified Payments Interface (UPI) has come one of the most popular styles for digital deals in India. As further people shift towards cashless payments, the use of UPI for everyday deals similar as shopping, bill payments, and plutocrat transfers has grown significantly. Cybercriminals are constantly developing new styles to trick stoners through fake apps, swindle dispatches, phishing links, and unauthorized access. These frauds constantly affect in major financial losses for individualities and fiscal institutions, and they can damage public trust in digital payment systems. To reduce analogous risks, there is a growing need for smart systems that can automatically descry unusual or suspicious conditions during UPI deals. One analogous system is the Machine knowledge model, a statistical tool that can be trained to descry normal patterns of user gestures. This project proposes an intelligent system for detecting UPI frauds using machine learning techniques. The system leverages a dataset containing both valid and fraudulent transaction records, extracting key features such as transaction amount, balances, transaction type, and time patterns. Multiple machine learning models including Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Naive Bayes, Decision Tree, and Random Forest were applied and evaluated for their accuracy in fraud detection. Among them, the Random Forest model achieved the highest accuracy of 99.99%, demonstrating its effectiveness in identifying suspicious transactions. The trained model is deployed as a web-based application with a user-friendly interface developed using HTML and CSS, allowing users to input transaction details and receive instant fraud predictions. This intelligent system offers a practical solution for real-time fraud detection, enhancing the security of UPI transactions and contributing to safer digital payment ecosystems.

**Keywords:** Unified Payments Interface (UPI), UPI fraud discovery, fiscal trade monitoring, Machine literacy in fraud discovery.

1. **INTRODUCTION**

Unified Payments Interface (UPI) has completely transformed the way people in India manage financial transactions. Developed by the National Payments Corporation of India (NPCI), UPI allows users to send and receive money instantly using their smartphones. By linking bank accounts directly to a UPI-enabled app, users can make secure, real-time transactions without needing to visit a bank or use cash.

The system is available 24/7, including weekends and holidays, which makes it highly convenient for users across all age groups and regions. Its simplicity, speed, and ease of access have contributed to the massive growth in digital payments throughout the country. As UPI has become more widely accepted for everything from paying utility bills and shopping online to transferring money between friends and family, the volume of transactions has skyrocketed. Unfortunately, this widespread use has also attracted the attention of cybercriminals.

The increase in UPI transactions has been accompanied by a worrying rise in fraud cases. Scammers often trick users into sharing sensitive information such as their UPI PIN or OTP (one-time password) through fake phone calls, fraudulent mobile apps, phishing messages, and social engineering tactics. In some cases, users are directed to click on suspicious links or scan QR codes that lead to unauthorized money transfers. These fraudulent actions not only cause significant financial loss to individuals but also negatively impact the reputation of the entire digital payment system.

Because of the evolving nature of these fraud methods, traditional detection systems that rely on fixed rules or manual checks are no longer sufficient. The need for advanced, real-time fraud detection has become critical. To tackle this growing challenge, modern technologies such as machine learning, artificial intelligence, and behavioural analysis are being explored. These systems can learn from transaction data, identify patterns of normal user behaviour, and quickly detect suspicious or unusual activities that could indicate fraud. This project (or research paper/report) focuses on the use of intelligent techniques, particularly machine learning models like the Random Forest, to detect potential fraud in UPI transactions. By analysing various steps involved in a UPI transaction, such as the timing, frequency, transaction amount, and user behaviour, these models can differentiate between genuine and suspicious transactions. If a transaction does not match the trained behaviour patterns, it can be flagged for review.

In conclusion, while UPI has made digital payments more accessible and efficient, it has also opened the door to new risks. Therefore, combining the convenience of UPI with robust fraud detection systems is essential to maintain user trust and ensure the continued success of digital financial services in India.

1. **LITERATURE REVIEW**

As UPI (Unified Payments Interface) continues to grow in India, more people are using it for fast and easy money transfers. But with this growth, cases of online fraud have also gone up. Many researchers are now working on ways to detect and stop UPI fraud using smart technologies like machine learning.

[1] compared several machines learning models—like Decision Trees, Support Vector Machines (SVM), and Random Forest—to see which one works best in spotting fake UPI transactions. Her study helps choose the right tool for building fraud detection systems.

A study by [2] used a mix of methods like Hidden Markov Models (HMM), clustering, and anomaly detection to check if a UPI transaction is unusual. If a transaction doesn’t follow the normal pattern, the system marks it as possibly fake.

Another team, [3], created a model that combines different machine learning techniques to spot fraud in real time. Their system can adapt to new types of fraud better than older, rule-based systems, which often fail when fraudsters change their tactics.

One recent study by [4] looked at using machine learning to catch fraud in UPI payments. Their method uses data such as how users behave during transactions and how their devices work, to find patterns that look suspicious.

Some researchers are also thinking about privacy. [5] Awosika et al. (2023) used something called Federated Learning and Explainable AI. This method lets banks share information to train a model together without exposing any private user data. It also makes sure the system is easy to understand and not just a "black box."

In another study, Xu et al. (2023) introduced a powerful model called Deep Boosting Decision Trees (DBDT). It gives better results, especially when there is much more data from normal transactions than from fraud cases.

Overall, recent studies show that the best results come from combining smart models, real-time monitoring, and privacy-aware systems. These tools are helping make UPI transactions safer while keeping users’ personal information protected.

1. **METHODOLOGY**

This study focuses on developing a system that can detect fraudulent UPI (Unified Payments Interface) transactions by applying machine learning methods. The steps involved in the development of this system are as follows:

***1. Data Collection***

The first step involves gathering transaction data related to UPI payments. The dataset should include:

**TABLE 1: DATASET FEATURES**

|  |  |
| --- | --- |
| Feature | Description |
| Transaction ID | An exclusive reference number assigned to each transaction to distinguish it from others. |
| Sender and receiver information (anonymized) | Concealed identity details of both parties involved to ensure privacy. |
| Transaction amount | The monetary value exchanged during the transaction process. |
| Time and date | The logged moment, including both the day and time, when the transaction was made. |
| Balance of Sender | The amount of money in the sender's account around the time of the transfer. |
| Balance of Receiver | The receiver's account total after being updated with the incoming funds. |
| Label indicating if the transaction is genuine or fraudulent | A marker used to classify whether the transaction is legitimate or potentially dishonest. |

If access to real-world UPI transaction data is limited due to privacy concerns, a synthetic dataset can be generated based on known transaction behavior patterns to simulate fraud and normal transactions.

***2. Data Preprocessing***

After gathering the data, it must be organized and processed to make it ready for analysis. This involves:

Removing duplicate or missing records: Clean the dataset by eliminating repeated entries and handling null values to ensure data integrity. Converting categorical variables into numeric form: Transform text-based categories into numerical codes so that machine learning models can process them. Normalizing features to bring values to a standard scale – Scale features to a common range to improve model performance and convergence speed. Handling imbalanced datasets using oversampling or undersampling techniques (like SMOTE): Adjust class distribution using techniques like SMOTE to prevent bias in model predictions. Labeling the transactions appropriately for supervised learning: Assign correct class labels to data points to train models for accurate classification. This step helps make sure the data is clean, reliable, and ready to be used for training the model.

***3. Feature Selection***

To improve performance and reduce complexity, only relevant features are selected. Important indicators of fraud may include: Type of transaction: The nature of the transaction (e.g., payment, transfer), which can signal typical or suspicious behavior. Unusual transaction time: Transactions occurring at odd hours may serve as warning signs of fraudulent activity.

Change in balance: Sudden or unexpected shifts in account balance can be strong indicators of potential fraud.

Selecting the right features helps the model focus on patterns that are most likely linked to fraud.

***4. Model Selection***

Various machine learning models are chosen to detect and classify fraudulent UPI transactions. The models selected for comparison include:

Logistic Regression: This technique calculates the chance of an event happening, like detecting fraud, using a formula based on data. Decision Tree: A model that makes decisions step-by-step, asking questions at each branch based on input values.

Random Forest: Builds many decision trees and merges their outcomes to give a more reliable result.

Support Vector Machine (SVM): Classifies data by drawing the widest possible line that separates different groups.

K-Nearest Neighbors (KNN) – Predicts the outcome of a data point by checking what outcomes its closest neighbors have.

Naive Bayes: A quick method that guesses the result by calculating probabilities, assuming features are unrelated

Each of these models has unique strengths. For example, Random Forest is a powerful model that uses multiple decision trees to improve accuracy and reduce errors, making it effective for detecting fraud.

***5. Model Training and Testing***

The dataset is divided into two parts:

Training Set (typically 70–80%): Used to train the machine learning model. Testing Set (20–30%): This part of the data is used to check how well the model works on new, unseen data and helps measure its real-world accuracy. Cross-validation methods can also help prevent overfitting and make the model perform better on new data.

***6. Model Evaluation***

To measure the performance of each model, the following evaluation metrics are used:

Accuracy: It shows how many times the model gives the correct result out of all predictions made. Precision: It tells how many of the transactions flagged as fraud by the model were actually fraudulent. Recall (Sensitivity): The proportion of actual fraudulent transactions the model successfully detects. F1-Score: It represents the harmonic mean of precision and recall, offering a unified measure that balances the trade-off between false positives and false negatives.

These metrics help in identifying the model that gives the most reliable results in fraud detection.

***7. Implementation of Fraud Detection System***

After identifying the best-performing model, it is implemented as a fraud detection system. This system monitors ongoing UPI transactions and flags those that appear suspicious based on learned patterns.

These alerts can help banks and users act fast to stop possible fraud and avoid losing money.

***8. Continuous Improvement***

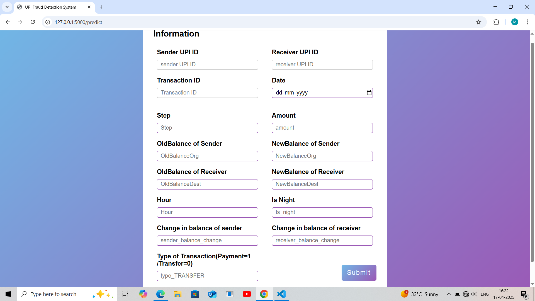
Since fraud patterns can change over time, the model must be updated regularly. The system can be updated with new data, helping it adjust to changing fraud patterns and keep its detection performance accurate.

1. **SYSTEM DESIGN**

The UPI Fraud Detection System is a web-based application designed to help detect potentially fraudulent Unified Payments Interface (UPI) transactions using machine learning. The front end of the application is developed using HTML and CSS, providing a clean and user-friendly interface where users can input transaction details such as amount, transaction time, and other relevant features. The interface includes a simple “Check” button that triggers the fraud detection process. On the backend, a machine learning model trained on historical UPI transaction data is deployed using frameworks like Flask or Django. This model analyzes the input features in real time and predicts whether the transaction is likely fraudulent or genuine. The system is designed to be fast, responsive, and accurate, providing users with instant feedback while maintaining strong security and data validation practices. By integrating the trained model into the web application, this system offers a practical tool for early detection and prevention of UPI-based financial frauds.

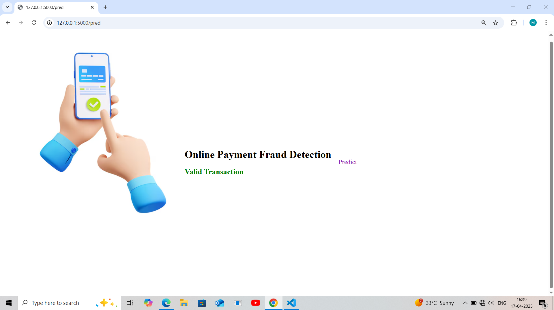
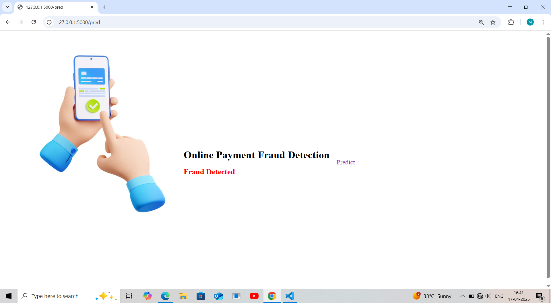


***Figure 1: UPI Fraud Detection System Web App***



***Figure 2: Input Form***

This screen shows the input form of the UPI Fraud Detection System Web App, where users enter transaction details for fraud analysis. It collects key fields like Sender UPI ID, Receiver UPI ID, Transaction ID, Amount, Balances before and after the transaction, Time, and Transaction type. After filling the form, the user clicks Submit, which sends the data to the backend for fraud prediction using a machine learning model. The design is simple, with a clear layout and a gradient background, making it easy to use.



***Figure 2: Predicted Result***

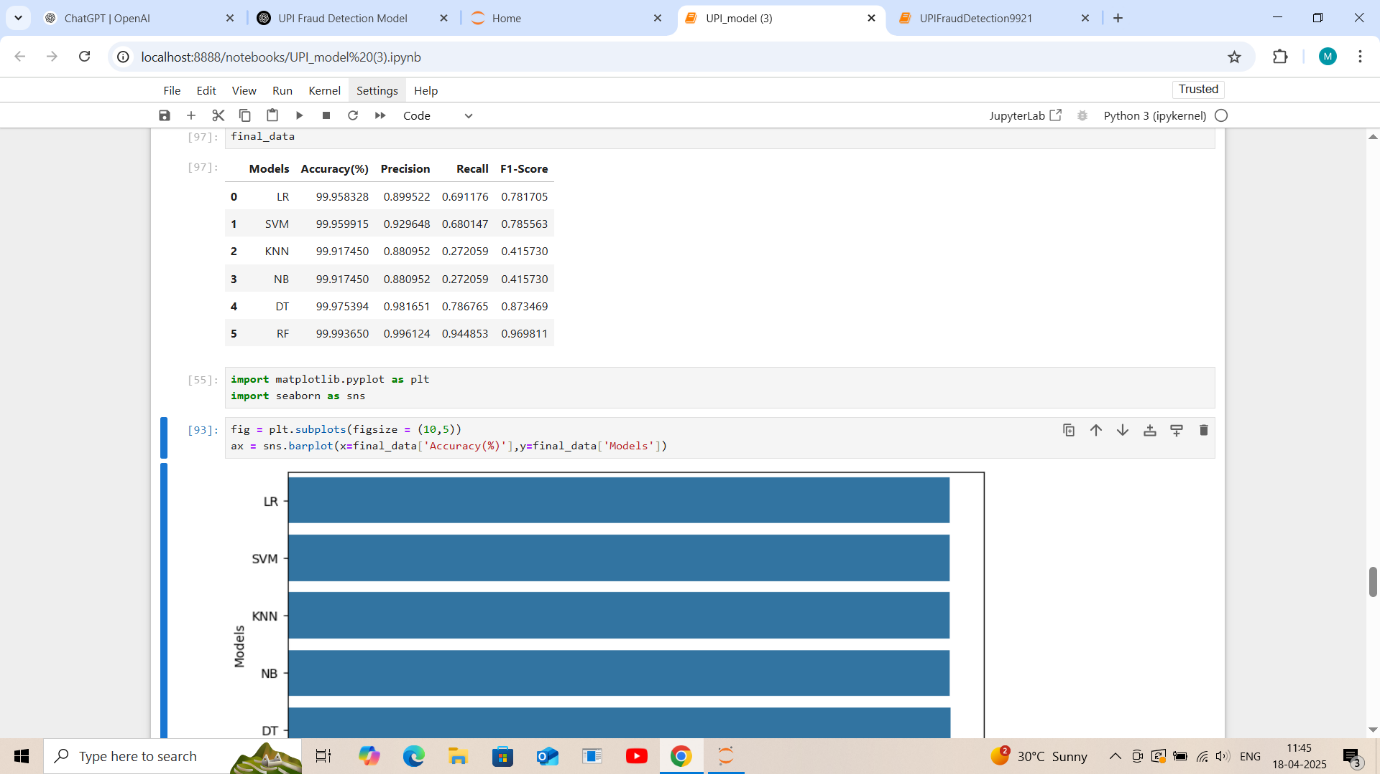
This page confirms the prediction outcome to the user. If fraud is detected, the text would likely display "Fraud Detected" in red and green text if the transaction in valid transaction.

1. **RESULTS AND DISCUSSION**

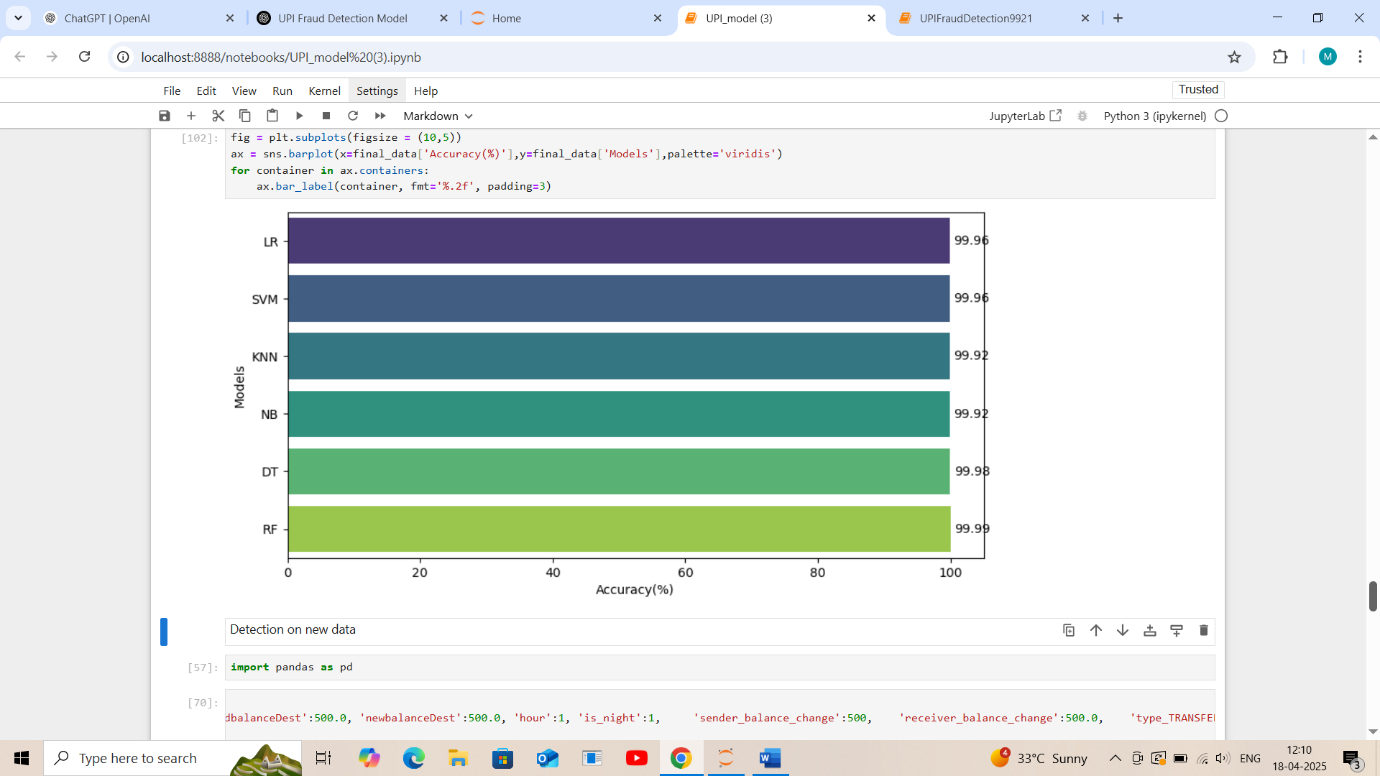
The UPI Fraud Detection System Web Application was successfully developed and tested to classify online payment transactions as either fraudulent or valid using machine learning. The system allows users to input detailed transaction features such as Sender UPI ID, Receiver UPI ID, Amount, Old and New Balances, Transaction Time, and Type of Transaction. After submission, the machine learning model processes these inputs and returns a prediction in real-time.

To build an accurate fraud detection system, six different machine learning models were applied and evaluated based on their accuracy in classifying UPI transactions as fraudulent or valid. The results are shown in the bar chart above, which compares the performance of each model.

**Table 2: ML Classifiers**

****

To evaluate the effectiveness of various machine learning algorithms in detecting fraudulent UPI transactions, six classifiers were trained and tested on the dataset: Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbors (KNN) and Random Forest (RF). Among these, the Random Forest model outperformed all others, achieving the highest accuracy of 99.99%, along with a precision of 0.996, a recall of 0.945, and an F1-score of 0.970. The decision tree model was followed closely, showing strong performance with an accuracy of 99.98% and an F1-score of 0.873. In contrast, while Logistic Regression and SVM demonstrated high accuracy scores (approximately 99.96%), their recall values were relatively lower (0.691 and 0.680, respectively), indicating limited effectiveness in capturing fraudulent instances. The KNN and Naive Bayes models, although achieving respectable accuracy, exhibited significantly lower recall and F1-scores, making them less reliable for fraud detection in this context. The Random Forest classifier demonstrated the best balance across all performance metrics, making it the most suitable choice for accurate and robust UPI fraud detection.



***Figure 3: Model Evaluation Metrics***

The Random Forest (RF) model achieved the highest accuracy of 99.99%, making it the best-performing algorithm for this UPI fraud detection task. Random Forest combines multiple decision trees, making it robust against overfitting and very effective in handling complex fraud patterns. Decision Tree (DT) also performed very well with 99.98% accuracy, slightly below Random Forest. It is simple and interpretable but slightly less stable than RF. Logistic Regression (LR) and Support Vector Machine (SVM) both showed excellent accuracy at 99.96%, indicating that even linear models can effectively capture patterns in the transaction data. K-Nearest Neighbors (KNN) and Naive Bayes (NB) had slightly lower accuracies at 99.92% but still performed well, showing the dataset's overall separability between fraudulent and valid transactions.

1. **CONCLUSION**

The experimental results demonstrate that machine learning models can effectively detect fraudulent UPI transactions with high accuracy. Among the models evaluated, the Random Forest classifier consistently outperformed others, achieving the highest values across all key performance metrics, including accuracy, precision, recall, and F1-score. This indicates its superior ability to not only correctly classify transactions but also minimize false negatives—an essential factor in fraud detection tasks. While models such as Logistic Regression and Support Vector Machine delivered competitive accuracy, their relatively lower recall values suggest a reduced capability in identifying fraudulent instances. Conversely, algorithms like K-Nearest Neighbors and Naive Bayes, despite yielding high accuracy, suffered from poor recall and F1-scores, making them less practical for real-world fraud detection scenarios where missing fraudulent cases can be costly.

In conclusion, the Random Forest algorithm emerges as the most reliable and effective model for UPI fraud detection, offering a strong balance between precision and recall. These findings support its suitability for deployment in real-time financial fraud prevention systems.

**ACKNOWLEDGEMENTS**

I would like to acknowledge the contributors of the Fraud Detection Dataset available on Kaggle, which served as the primary data source for training and evaluating the machine learning models in this project. The dataset, containing both valid and fraudulent transaction records. I would like to express my sincere gratitude to Dr. Tejashree Moharekar, my project guide, for her invaluable guidance, continuous support, and encouragement throughout the completion of this project titled "UPI Fraud Detection System". Her expert advice and constructive feedback helped me at every stage of the work and contributed significantly to the successful development of this project. I also extend my heartfelt thanks to the Yashwantrao Chavan School of Rural Development (YCSRD), Shivaji University, Kolhapur, for providing the necessary resources and an excellent academic environment that enabled me to carry out this project effectively.

1. **REFERENCES**
2. V. S. S. J. Sindhu, "Implementation of ML algorithms for detecting irregular UPI transactions," International Journal of Engineering Research and Science & Technology, p. 57–67, 2024.
3. G. I. A. K. A. S. a. D. B. J. Kavitha, "Detection of fraudulent transactions in UPI systems using supervised learning methods," EPRA International Journal of Research and Development, pp. 45-52, 2024.
4. A. S. S. S. R. K. M. Nazmoddin, "Developing a hybrid machine learning framework to detect UPI frauds," Journal of Computational Analysis and Applications, p. 1192–1200, 2024.
5. P. P. S. Bodade, "An overview of machine learning strategies for identifying UPI payment fraud," International Journal for Research in Applied Science and Engineering Technology, p. 30–35, 2023.
6. S. K. R. P. O. Awosika, "Federated learning and interpretability in fraud detection for UPI payments," Transactions on Secure Financial Computing, p. 88–97, 2023.
7. N. M. A. Gupta, "Applying ML techniques to flag UPI-based transactional fraud," International Journal of Innovative Research in Computer and Communication Engineering, p. 55–60, 2024.
8. A. V. D. G. M. D. a. V. S. K. S. S. e. a. S. Jagadeesan, "Design and evaluation of a UPI fraud detection system using machine learning," in Challenges in Information, Communication and Computing Technology, p. 755–759, 2024.
9. V. R. A. Sharma, "Convolutional neural networks for secure and intelligent UPI transaction monitoring," ResearchGate, 2024.
10. K. D. A. R. N. Sridevi, "Comparative study on ML and DL techniques for fraud detection in UPI systems," International Journal for Research in Applied Science and Engineering Technology, pp. 18-22, 2024.
11. D. A. K. R. D. S. A. M. V. Vijaykumar, "Machine learning-enabled fraud detection framework for secure digital transactions in UPI and card-based systems," International Journal of Advanced Research in Engineering and Science Management, pp. 112-118, 2024.