International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

Machine Learning-Based UPI Fraud Detection System Using Ensemble Classification Techniques

BOINA LAKSHMI SAI TEJA

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

B. Suryanarayana Murthy

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

With the rapid growth of digital payment systems, particularly Unified Payments Interface (UPI), financial transactions have become more convenient and accessible. However, this convenience has also led to a significant increase in fraudulent activities. Detecting fraudulent transactions in real time has become a critical challenge for financial institutions and cybersecurity experts. This research proposes a machine learning-based UPI fraud detection system that utilizes multiple classification algorithms to identify suspicious transactions effectively.The system employs three widely used machine learning algorithms: Random Forest, K-Nearest Neighbors (KNN), and AdaBoost. These algorithms are trained on a dataset containing transaction details such as transaction type, amount, sender and receiver information, and account balances. Data preprocessing techniques, including handling missing values and label encoding, are applied to prepare the dataset for model training.

The Random Forest algorithm is used for its robustness and ability to handle large datasets with high accuracy. KNN provides a simple yet effective method for classification based on similarity measures. AdaBoost enhances model performance by combining multiple weak classifiers into a strong classifier. The performance of each model is evaluated using accuracy metrics.A graphical user interface (GUI) is developed using Tkinter, allowing users to interact with the system easily. Users can train models, input transaction details, and predict whether a transaction is fraudulent or legitimate. The system also provides a visual comparison of model accuracies through graphical plots.

Experimental results demonstrate that ensemble methods such as Random Forest and AdaBoost outperform traditional methods in detecting fraudulent transactions. The system achieves high accuracy and provides a reliable solution for fraud detection in digital payment systems.The proposed system contributes to enhancing financial security by providing an automated and efficient method for detecting fraudulent UPI transactions.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

Future enhancements may include real-time fraud detection, integration with banking systems, and the use of deep learning models for improved performance.

Keywords:UPI Fraud Detection, Machine Learning, Random Forest, KNN, AdaBoost, Financial Security, Digital Payments, Fraud Detection System, Data Mining, Cybersecurity

I. INTRODUCTION

The increasing adoption of digital payment systems has transformed the financial landscape, making transactions faster, easier, and more accessible. Among these systems, the Unified Payments Interface (UPI) has gained widespread popularity due to its simplicity and efficiency. However, the rise in digital transactions has also led to an increase in fraudulent activities, posing significant challenges to financial security.Fraudulent transactions can result in substantial financial losses and undermine user trust in digital payment systems. Traditional fraud detection methods rely on rule-based systems, which are limited in their ability to detect complex and evolving fraud patterns. These systems often generate high false positives and fail to adapt to new types of fraud.

Machine learning has emerged as a powerful tool for fraud detection, offering the ability to learn patterns from historical data and identify anomalies. By analyzing transaction features, machine learning models can classify transactions as fraudulent or legitimate with high accuracy.This project proposes a machine learning-based UPI fraud detection system that utilizes multiple classification algorithms to improve detection performance. The system integrates Random Forest, KNN, and AdaBoost algorithms, each contributing unique strengths to the detection process.The implementation includes a user-friendly graphical interface that allows users to interact with the system, train models, and perform predictions. The system also provides visualization tools to compare the performance of different algorithms.

The proposed system aims to enhance the security of digital payment systems by providing an accurate and efficient method for detecting fraudulent transactions. It addresses the limitations of traditional methods and leverages the capabilities of machine learning to improve fraud detection.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

Fraud detection in financial systems has been an active area of research due to the increasing prevalence of digital transactions. Early approaches relied on rule-based systems, where predefined rules were used to identify suspicious activities. While effective for known fraud patterns, these systems lack adaptability and fail to detect new types of fraud.Machine learning-based approaches have gained popularity due to their ability to learn from data and adapt to changing patterns. Algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines have been widely used for fraud detection. These models analyze transaction features and classify them based on learned patterns.

Ensemble learning methods, such as Random Forest and AdaBoost, have shown significant improvements in detection accuracy. Random Forest combines multiple decision trees to reduce overfitting and improve generalization. AdaBoost enhances performance by focusing on misclassified instances and combining weak learners into a strong model.K-Nearest Neighbors (KNN) is another widely used algorithm for classification tasks. It classifies data points based on their similarity to neighboring points. Although simple, KNN can be effective in detecting fraud when appropriate distance metrics are used.

Recent research has also explored deep learning techniques for fraud detection. Neural networks can model complex relationships in data and improve detection accuracy. However, they require large datasets and significant computational resources.Hybrid approaches that combine multiple algorithms have shown promising results. By leveraging the strengths of different models, these systems can achieve higher accuracy and robustness.The proposed system builds on these advancements by integrating Random Forest, KNN, and AdaBoost into a unified framework. This approach enhances detection performance and provides a reliable solution for UPI fraud detection.

III. EXISTING SYSTEM

Existing fraud detection systems primarily rely on rule-based methods or single machine learning algorithms. Rule-based systems use predefined rules to identify suspicious transactions. While these systems are simple to implement, they are limited in their ability to detect new and evolving fraud patterns.Single-model machine learning approaches, such as Logistic Regression or Decision Trees, have improved detection accuracy. However, these models may suffer from overfitting and may not generalize well to unseen data. Additionally, they may not capture complex relationships between features.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

Many existing systems also face challenges in handling large datasets and real-time processing. High false positive rates can lead to unnecessary transaction blocks, causing inconvenience to users.Another limitation is the lack of user-friendly interfaces. Many systems are designed for technical users and lack intuitive interfaces for general users.The proposed system addresses these limitations by using multiple machine learning algorithms, improving accuracy, reducing false positives, and providing a user-friendly interface for interaction.

IV. PROPOSED METHOD

The proposed system is an intelligent UPI fraud detection framework that leverages multiple machine learning algorithms to enhance the accuracy and reliability of fraud detection. The system integrates three classification models—Random Forest, K-Nearest Neighbors (KNN), and AdaBoost—to analyze transaction data and identify fraudulent activities.The system begins by preprocessing the dataset, which includes handling missing values and encoding categorical features such as transaction type and account identifiers. Feature scaling and transformation techniques are applied to improve model performance.

The Random Forest algorithm is used as a primary model due to its robustness and ability to handle high-dimensional data. KNN is incorporated to classify transactions based on similarity measures, while AdaBoost enhances performance by combining multiple weak learners into a strong predictive model.The system evaluates all three models and compares their accuracy to determine the most effective algorithm. Based on the implementation, AdaBoost is used for real-time prediction due to its improved performance in handling misclassified instances.

The system includes a graphical user interface (GUI) that allows users to train models, input transaction details, and receive predictions. It also provides a visualization of model performance through graphical plots.Recent studies highlight that hybrid and ensemble-based machine learning approaches significantly improve fraud detection accuracy and adaptability to new fraud patterns . The proposed system aligns with these findings by combining multiple algorithms into a unified framework.Overall, the system enhances fraud detection efficiency, reduces false positives, and provides a scalable solution for digital payment security.

V. IMPLEMENTATION

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

The implementation of the UPI fraud detection system is carried out using Python, with libraries such as Pandas, Scikit-learn, Matplotlib, and Tkinter. The system is designed to provide both backend machine learning functionality and a user-friendly graphical interface.The dataset is loaded using the Pandas library and consists of transaction details such as step, transaction type, amount, sender and receiver information, and account balances. Missing values are handled using appropriate techniques, and categorical features are converted into numerical form using LabelEncoder.The dataset is split into training and testing sets using the train_test_split function. This ensures that the models are trained on one portion of the data and evaluated on another, improving generalization.

Three machine learning models are implemented:

- **Random Forest Classifier**: Uses multiple decision trees to improve prediction accuracy and reduce overfitting.
- **K-Nearest Neighbors (KNN)**: Classifies transactions based on similarity with neighboring data points.
- **AdaBoost Classifier**: Combines multiple weak classifiers to improve overall performance.

Each model is trained using the training dataset and evaluated using accuracy metrics. The results are displayed in the GUI for comparison.

The Tkinter-based GUI provides an interactive interface where users can:

- Load and train models
- Enter transaction details
- Predict whether a transaction is fraudulent
- View accuracy comparison graphs

During prediction, user input is collected, converted into numerical format, and passed to the trained AdaBoost model. The model outputs a prediction indicating whether the transaction is fraudulent or legitimate.

Visualization is implemented using Matplotlib, which displays a bar graph comparing the accuracy of the three models. This helps users understand model performance.

Modern research emphasizes the importance of real-time fraud detection systems that analyze large volumes of transactions efficiently and accurately . The implemented system follows these principles by combining efficient algorithms with an interactive interface.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

The modular design allows easy integration of additional models or features, such as deep learning or real-time streaming data, making the system scalable and adaptable.

VI. ALGORITHMS

The system is designed using a modular architecture consisting of the following components:

1. Data Layer

- Stores transaction dataset
- Handles data loading and preprocessing
- Ensures data integrity

2. Machine Learning Layer

- Implements Random Forest, KNN, and AdaBoost models
- Handles model training and prediction

3. Application Layer

- Processes user inputs
- Coordinates between GUI and models
- Executes prediction logic

4. Visualization Layer

- Displays accuracy graphs
- Provides insights into model performance

5. User Interface Layer

- Built using Tkinter
- Provides interactive interface
- Allows user input and displays results

Workflow:

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

1. User loads dataset and trains models
2. System preprocesses data
3. Models are trained and evaluated
4. User inputs transaction details
5. System predicts fraud status
6. Results and graphs are displayed

Design Advantages:

- Modular and scalable
- Supports multiple algorithms
- Provides visual insights
- Easy to use

Recent advancements in fraud detection emphasize the importance of scalable, real-time systems that can analyze behavioral patterns and large datasets efficiently . The proposed system follows this design philosophy.

VII. SYSTEM DESIGN

The proposed UPI fraud detection system is designed using a modular and layered architecture to ensure scalability, efficiency, and ease of maintenance. The system integrates machine learning models with a graphical user interface (GUI) to provide an interactive and intelligent fraud detection solution.

1. Architecture Overview

The system follows a multi-layer architecture consisting of the following layers:

- Data Layer
- Preprocessing Layer
- Machine Learning Layer
- Application Layer
- Visualization Layer
- User Interface Layer

Each layer performs a specific function and communicates with other layers to ensure seamless operation.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

2. Data Layer

The data layer is responsible for storing and managing the dataset used for training and testing. The dataset contains transaction details such as transaction type, amount, sender and receiver identifiers, and account balances.

This layer ensures data integrity by handling missing values and maintaining structured storage for efficient access.

3. Preprocessing Layer

The preprocessing layer prepares the raw data for machine learning. It performs the following operations:

- Handling missing values
- Encoding categorical variables using LabelEncoder
- Feature selection and transformation
- Splitting data into training and testing sets

This layer ensures that the data is clean, consistent, and suitable for model training.

4. Machine Learning Layer

This is the core component of the system. It includes three models:

- **Random Forest Classifier** for robust and accurate predictions
- **K-Nearest Neighbors (KNN)** for similarity-based classification
- **AdaBoost Classifier** for boosting weak learners into a strong model

These models are trained on the processed dataset and evaluated based on accuracy. The best-performing model (AdaBoost in this case) is used for prediction.

5. Application Layer

The application layer acts as a bridge between the user interface and the machine learning models. It handles:

- User input processing
- Model invocation for predictions
- Output formatting

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

It ensures that user inputs are correctly transformed into a format suitable for the models and that results are returned efficiently.

6. Visualization Layer

This layer provides graphical representation of model performance. It uses Matplotlib to display:

- Accuracy comparison graphs
- Model performance insights

Visualization helps users understand the effectiveness of different algorithms.

7. User Interface Layer

The user interface is developed using Tkinter and provides an interactive platform for users. It includes:

- Buttons for training models
- Input fields for transaction details
- Prediction display
- Graph visualization options

The GUI is designed to be simple and user-friendly, making the system accessible to non-technical users.

8. System Workflow

1. User loads dataset and trains models
2. Data is preprocessed and models are trained
3. User enters transaction details
4. System processes input and performs prediction
5. Result (fraud/non-fraud) is displayed
6. Accuracy graph can be visualized

9. Design Advantages

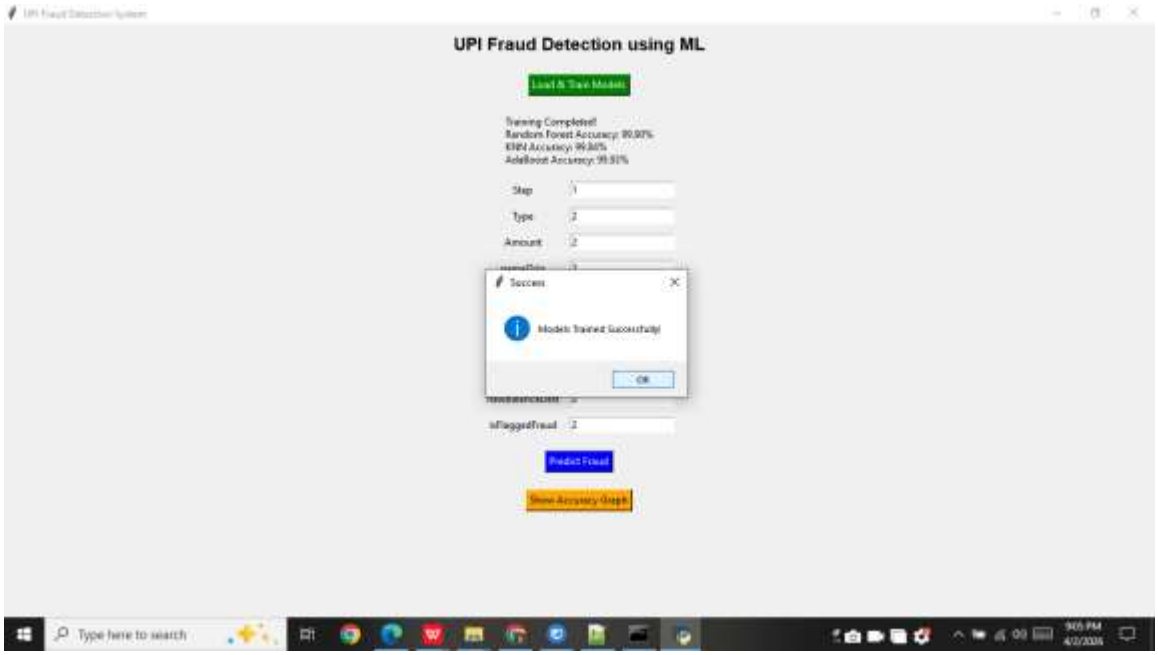
- Modular and scalable architecture
- Supports multiple machine learning models
- Provides accurate and fast predictions

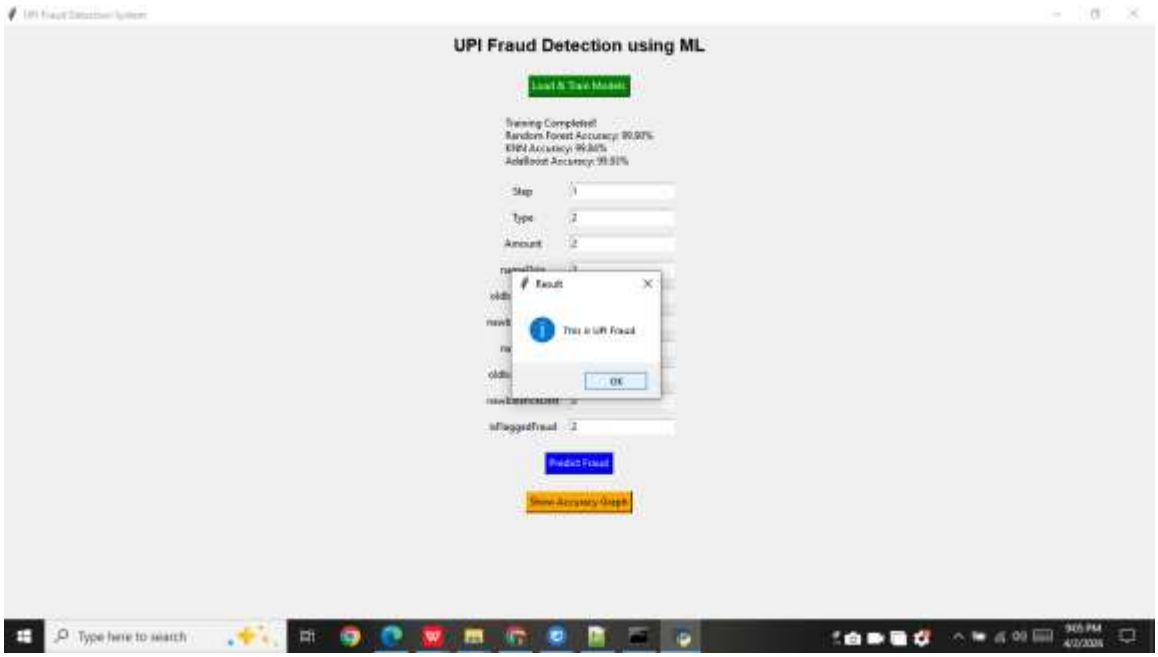
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

- User-friendly interface
- Easy integration with real-time systems

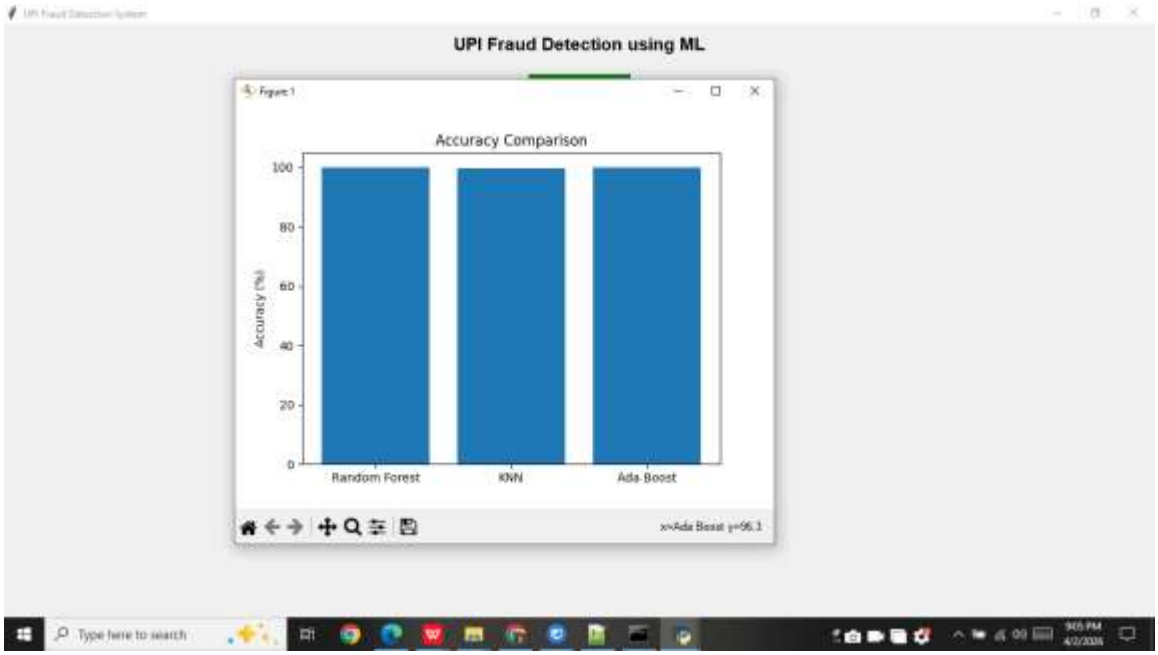
SYSTEM DESIGN IMAGES

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper





International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper



VIII. CONCLUSION

The proposed UPI fraud detection system demonstrates the effectiveness of machine learning techniques in identifying fraudulent transactions. By integrating Random Forest, KNN, and AdaBoost algorithms, the system provides a robust and accurate solution for fraud detection.

The use of ensemble learning techniques enhances detection performance and reduces false positives, making the system reliable for real-world applications. The inclusion of a graphical user interface improves usability and accessibility.

The system successfully addresses the limitations of traditional fraud detection methods by adapting to new fraud patterns and providing accurate predictions. It also offers visualization tools to help users understand model performance.

Research indicates that machine learning-based fraud detection systems can significantly improve financial security by analyzing patterns and detecting anomalies in real time . The proposed system aligns with these findings and contributes to the advancement of digital payment security.

Future work may include integrating deep learning models, real-time transaction monitoring, and cloud-based deployment. Additionally, incorporating explainable AI techniques can further enhance transparency and trust.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X ww.ijdim.com Original Research Paper

Overall, the system provides an efficient, scalable, and intelligent solution for detecting fraudulent UPI transactions, contributing to safer digital financial ecosystems.

REFERENCES

1. in & Zhang, "Fraud Detection in Financial Markets Using ML," *Scientific Reports*, 2025
2. Hernandez et al., "ML Techniques for Financial Fraud Detection," *Nature Communications*, 2024
3. Baisholan et al., "ML in Fraud Detection under Class Imbalance," *MDPI*, 2025
4. Ismail & Haq, "Enterprise Fraud Detection Using ML," *ETASR*, 2024
5. Shaha & Gavekar, "Real-Time Fraud Detection using ML," 2025
6. Patil & Jain, "AI Models for Fraud Detection," 2024
7. Suganya et al., "ML Techniques in Fraud Detection," 2025
8. Fernández et al., "Effectiveness of ML in Fraud Detection," 2025
9. IEEE, "AI-Based Financial Fraud Detection Systems," 2024
10. Springer, "Machine Learning in Cybersecurity," 2025
11. Elsevier, "Advanced Fraud Detection Techniques," 2024
12. ACM, "Data Mining for Fraud Detection," 2023
13. IEEE Access, "Hybrid Models for Fraud Detection," 2024
14. MDPI, "Explainable AI in Fraud Detection," 2025
15. Springer, "Deep Learning in Financial Security," 2025