International Journal of Marketing and Technology
Vol. 16 Issue 6, June 2026,
ISSN: 2249-1058 Impact Factor: 8.469
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

PROACTIVE PREDICTION OF UPI TRANSACTION FAILURE USING MACHINE LEARNING ON SYNTHETIC DATA

Priyam Aggarwal[1], Dr. Ajay Kumar[2]
[1] Student, Department of Computer Science & Engineering, Chitkara University Institute of engineering & Technology, Punjab, India.
priyam2098.be22@chitkara.edu.in

[2] Associate Professor, Department of Computer Science & Engineering, Chitkara University Institute of engineering & Technology, Punjab, India.
akumar@chitkara.edu.in

Abstract:

With the rapid usage of digital payment systems in India, the Unified Payments Interface (UPI) has become one of the most widely used platforms for real time financial transactions [6] [7]. It enables seamless and instant transfer of funds between bank accounts and making it a preferred choice for millions of peoples. despite its efficiency and scalability, transaction failures continue to occur due to various reasons such as bank server overload, network instability, high transaction traffic during peak hours and delays in interbank communication.

These failures often lead to poor user experience, repeated transaction attempts and increased system load. While existing systems can detect failures after they occur, there is still a lack of proactive mechanisms to predict such failures in advance.

In this research, an attempt is made to apply machine learning techniques to analyze transaction data and predict the likelihood of UPI transaction failure before execution. A synthetic dataset is used to simulate real world transaction scenarios while maintaining privacy. Features such as transaction amount, transaction time, sender bank and receiver bank are extracted and used to train classification model including Random Forest and XGBoost [1] [2].

The results suggests that transaction contains meaningful patterns that can be used to estimate failure probability. also predictive analytics can play an important role in improving the reliability and efficiency of digital payment systems while ensuring compliance through the use of synthetic data [3].

Keywords:
UPI, Digital Payments, Machine Learning, Failure Prediction, FinTech, Synthetic Data, Predictive Analytics

1. INTRODUCTION

With the rapid expansion of digital payment systems in India, the Unified Payments Interface (UPI) has become one of the most widely used platforms for real time financial transactions. It enables seamless and instant transfer of funds between bank accounts, making it a preferred choice for millions of peoples. However, despite its efficiency and scalability, transaction failures continue to occur due to various reasons such as bank server overload, network instability, high transaction traffic during peak hours and delays in inter-bank communication.

These failures often lead to poor user experience, repeated transaction attempts and increased system load. While existing systems can detect failures after they occur, there is still a lack of proactive mechanisms to predict such failures in advance.

In this research, an attempt is made to apply machine learning techniques to analyze transaction data and predict the likelihood of UPI transaction failure before execution. A synthetic dataset is used to simulate real-world transaction scenarios while maintaining data privacy. Features such as transaction amount, transaction time, sender bank and receiver bank are extracted and used to train classification models including Random Forest and XGBoost [1] [2].

The results suggest that transaction metadata contains meaningful patterns that can be used to estimate failure probability. This study demonstrates that predictive analytics can play an important role in improving the reliability and efficiency of digital payment systems while ensuring ethical compliance through the use of synthetic data [3].

2. LITERATURE REVIEW

Machine learning has been widely applied in the financial sector for various purposes, including fraud detection, credit scoring and anomaly detection [4] [5] [9]. These applications identify patterns of transaction data and make predictions based on historical outputs and trends.

learning techniques such as Random Forest and Gradient Boosting have proven to be efficient in handling structured financial datasets [2] [10]. These models can capture complex relationships between variables and provide high accuracy on classification tasks.

In the context of digital payments, most research are focused on detecting fraud transactions or improving security mechanisms. However limited work has been done on predicting transaction failures caused by technical or infrastructural issues.

Some studies indentify that transaction outcomes are influenced by reasons such as transaction time, system load and bank specific performances [11] [12]. these learning have not been applied to build predictive models for transaction reliability.

This research bridges this gap by applying machine learning techniques to analyze transaction data and predict failure probability, contributing to a less explored area in financial space.

3. OBJECTIVES

The primary objective of this research is to see how machine learning can be used to improve the reliability of UPI transactions by predicting possible failures before they occur.

This study aims to analyze transaction data in detail to identify patterns that indicate a higher probability of failure. By examining features such as transaction amount, time of transaction and the banks involved in transaction, this research seeks to understand how these reasons influence transaction outcomes.

Another important objective is to design and implement a machine learning models like Random Forest and XGBoost to evaluate their ability to classify transactions as successful or failed [1] [2].

.

In addition to this, the research also aims to demonstrate the practical use of such a predictive system. If implemented in real world applications, it could help peoples avoid failed transactions and enable banks to manage system load more effectively.

4. METHODOLOGY

This research follows a structured approach to analyze transaction data and develop a predictive model for identifying potential transaction failures. The methodology is divided into multiple steps, starting from data collection to model evaluation and prediction.

4.1. System Architecture

The proposed solution is designed as a predictive layer that can be integrated in digital payment systems. Instead of replacing existing application, the system works alongside of it by analyzing transaction details before execution and estimating the probability of failure.
The approach is simple, if a system can identify patterns from past transactions, it can use those patterns to make informed predictions about future transactions.

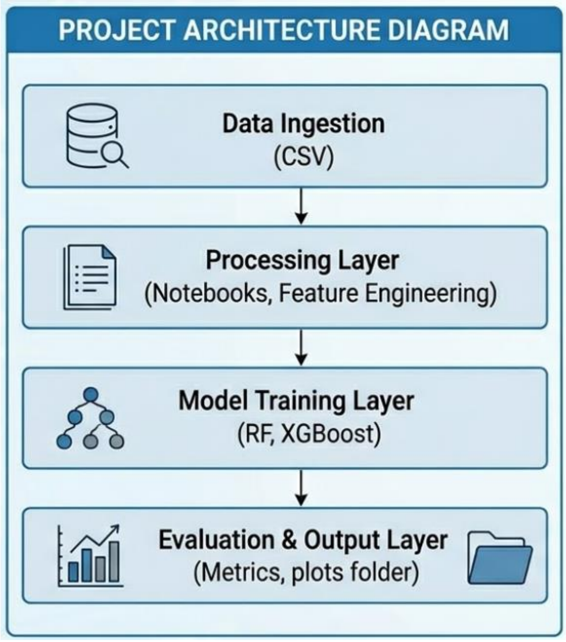


Figure 0.1: System Architecture diagram

4.2. Data Collection

The dataset used in this research is synthesized transaction dataset obtained from Kaggle. Since real banking data is highly sensitive and not publicly available, synthetic data is used to simulate realistic transaction behavior while ensuring privacy and ethical compliance.

The dataset contains the following features:

- Transaction ID
- Timestamp
- Sender UPI ID
- Receiver UPI ID
- Transaction Amount
- Transaction Status (SUCCESS / FAILED)

Although the data is synthesized, it closely resembles real world transaction patterns, making it suitable for machine learning projects. The implementation and detailed information of the model is available in an open source repository [16].

4.3. Data Pre processing

Before applying machine learning models, the dataset undergoes several pre processing steps to ensure data is of quality and consistency.

The timestamp column is converted into a proper date time format. This allows extraction of meaningful time based features such as hour and day.

missing values and inconsistencies in the dataset are handled. Since clean data is critical for accurate predictions, any incorrect or incomplete entries are either corrected or removed.

At last the dataset is structured in a format which is suitable for model training by separating input features and target labels.

4.4. Feature Engineering

Feature engineering plays a important role in improving model performance. Instead of directly using raw data, additional features are derived to capture hidden patterns.

The following features are extracted:

- Transaction Hour: Helps identify peak load periods
- Day of Week: Captures weekly transaction trends
- Sender Bank: Extracted from UPI ID
- Receiver Bank: Extracted from UPI ID

These features are important because the transaction failures are often influenced by time based load and bank specific conditions.

Categorical features like bank identifiers are encoded into numerical format so that they can be used by machine learning models.

4.5. Exploratory Data Analysis (EDA)

Exploratory Data Analysis is performed to understand the dataset available and identify patterns that are related to transaction failures.

The following analyses are carried out:

- Distribution of successful vs failed transactions
- Transaction volume across different hours
- Failure rate variation over time
- Transaction amount distribution
- Correlation between different features

This step helps in understanding which features are most important and how they influence transaction outcomes. [4] [5].

4.6. Model Selection

The machine learning models that are selected for this research:

1. Random Forest:

Random Forest is an learning algorithm [2] that combines multiple decision trees to improve

International Journal of Marketing and Technology
Vol. 16 Issue 6, June 2026,
ISSN: 2249-1058 Impact Factor: 8.469
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

prediction accuracy score. It is robust to overfitting and works well with structured datasets.

2. XGBoost:

XGBoost is a gradient boosting algorithm **[1] [10]**. known for its high performance and efficiency. It builds models sequentially and aims on minimizing prediction errors.

These models are chose because they are widely used in financial data analysis and provide reliable results.

4.7. Model Training

The dataset is divided into two parts:

Training Set (80%) → used to train the model

Testing Set (20%) → used to evaluate performance

During training, the model learns patterns from data by analyzing relationships between features and from transaction outcomes.

Hyper parameters of the models can be tuned to improve accuracy and performance.

4.8. Model Evaluation

To evaluate a model performance, the following metrics are used:

Accuracy: It Measures overall correctness

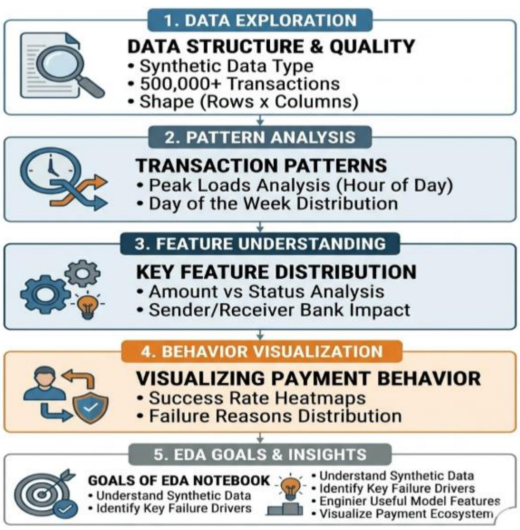
Precision: It measures correctness of positive predictions

Recall: It measures ability to detect failures

F1 Score: It balances between precision and recall

ROC-AUC Score: It measures classification performance

These metrics provide a detailed understanding of how well the model performs. **[4] [5]**.



4.9. Prediction Mechanism

Once the model is trained, it can be used to predict probability for new transactions.

The prediction process works like this:

1. User enters transaction details
2. Feature are extracted from input
3. Model processes the input
4. Failure Probability is calculated
5. System outputs prediction (SUCCESS / FAILURE likelihood)

This allows system to act as a **decision support** rather than replacing the actual transaction system.

5. RESULTS

The experiment results shows that machine learning models can effectively identify patterns of transaction failures. **[4] [5] [11]**.

Key findings include:

Higher transaction amounts may increase the probability of failure. Peak transaction hours have higher failure rates. Certain bank combinations affect transaction outcomes. Time based patterns play a important role **[12]**. The models achieved satisfactory performance, indicating that transaction metadata has meaningful predictive information.

6. CONCLUSION

This research demonstrates that machine learning can be used to predict UPI transaction failures using transaction metadata based on historical transactions.

This suggests that predicting models can be used to find patterns that influence transaction outcomes and provide early learning. This approach can be used to improve user experience and system efficiency.

Although the dataset used is synthetic, the results highlight the potential implementation of predictive systems in digital payments [11] [13].

7. FUTURE SCOPE

This research shows that machine learning can be used to predict UPI transaction failures, but there are several more ways to improve it further. In future the model can be enhanced by integrating **real time data** such as bank server status and network conditions that can improve prediction accuracy. The system can also be

deployed in actual **UPI applications or banking platforms** to provide real time feedback to peoples before they complete a transaction.
Further improvements can be made by using **advanced machine learning models** with deep learning techniques to capture more complex patterns in transaction data. Adding more features such as user behavior, device type and location can make the predictions more reliable.
Another use is **explainable AI**, which can help in understanding why a transaction is predicted to fail. This can improve trust and usability of the system.
Finally, the system can be developed into a **smart payment assistant** that not only predicts failures but also suggests actions like by retrying later or using a different payment method or different bank.

8. REFERENCES

[1] T. Chen and C. Guestrin, "XGBoost," 2016.
[2] L. Breiman, "Random Forests," 2001.
[3] Kaggle Dataset.
[4] IEEE ML in Finance, 2019.
[5] Springer FinTech Models, 2020.
[6] NPCI UPI Report, 2023.
[7] RBI Digital Payments Report, 2022.
[8] IEEE Payment Failures Study, 2021.
[9] Statistical Learning Journal, 1997.
[10] Gradient Boosting, 2001.
[11] McKinsey AI Banking, 2021.
[12] Deloitte AI Finance, 2020.
[13] World Economic Forum, 2022.
[14] IBM AI Banking, 2021.
[15] IEEE Digital Payments, 2020.
[16] GitHub Repository, 2026. Available: "https://github.com/priyamaggarwal18/transaction-failure-research"