

Fraud Detection in Digital Payment Systems Using Machine Learning

Gnaneswari V¹, Dr. R N Kulkarni²

¹ Department of Management Studies

² Department of Computer Science

Ballari Institute of Technology & Management, Ballari

Abstract - Digital payment platforms have become an essential part of everyday transactions, but their rapid expansion has also increased the number of fraudulent transactions, posing risks to both customers and financial institutions. This study develops a machine learning-based fraud detection system using a dataset of 7,500 digital payment transactions spanning UPI, cards, wallets, and net banking. After data preprocessing and exploratory analysis, the SMOTE technique is applied to correct class imbalance in the training data. Four algorithms - Logistic Regression, Decision Tree, Random Forest, and XGBoost - are trained and compared using Accuracy, Precision, Recall, F1-score, and ROC-AUC. While XGBoost achieved the highest accuracy (82.5%), Logistic Regression achieved the highest recall (30.6%), making it more effective at catching fraud despite its lower overall accuracy, and it was therefore selected as the preferred model. SHAP (SHapley Additive exPlanations) is used to interpret its predictions, identifying payment mode, account age, and prior failed attempts as the leading fraud indicators. The fraud probability is further converted into a 0-100 risk score with Low/Medium/High risk categories and a corresponding early warning alert, giving the system a practical, actionable output. The study shows that recall, rather than accuracy alone, should guide model selection in imbalanced fraud detection tasks, and it contributes a working risk-scoring framework for real-time digital payment monitoring.

Key words: *Digital Payments, Fraud Detection, Machine Learning, Risk Scoring, SHAP, SMOTE*

1. INTRODUCTION

Digital payment systems have transformed how individuals and businesses in India transact, driven largely by the Unified Payments Interface (UPI), launched by the National Payments Corporation of India in 2016, alongside wallets, cards, and net banking. Rising smartphone penetration, affordable internet access, and government initiatives such as Digital India have accelerated this shift toward cashless payments.

This growth has, however, been accompanied by a rise in fraudulent transactions carried out through phishing, fake payment links, SIM-swap fraud, and unauthorized use of card details. Traditional rule-based detection systems, which flag transactions using fixed thresholds, are increasingly ineffective because fraudsters can readily find ways around static rules. This motivates a machine learning-based approach that can learn patterns from historical transaction data and adapt to evolving fraud behaviour.

This study develops and compares four machine learning models - Logistic Regression, Decision Tree, Random Forest, and XGBoost - to classify digital payment transactions as genuine or fraudulent. Because fraud cases are a small minority of all transactions, the Synthetic Minority Oversampling Technique (SMOTE) is applied to balance the training data. SHAP (SHapley Additive exPlanations) is then used to interpret the selected model's predictions. In line with the Reserve Bank

of India's 2025 guidelines [11] on authentication for digital payments, which call for additional security checks only on transactions that appear risky, the system also assigns each transaction a Low, Medium, or High risk level with a corresponding recommended action - approval, OTP verification, or blocking for review.

Objectives: (1) collect transaction data covering UPI, Wallet, Card, and Net Banking payments; (2) preprocess, encode, and balance the data; (3) develop and compare machine learning models for fraud detection; (4) evaluate models using Accuracy, Precision, Recall, F1-score, and ROC-AUC to select the best-performing model; and (5) design a risk-scoring and alert mechanism based on the selected model's predictions.

2. LITERATURE REVIEW

Early work on payment fraud relied on descriptive and rule-based analysis. Mahdi et al. [1] examined online banking fraud trends in the UK and found phishing and card-not-present fraud to be major threats, but did not apply machine learning. Al-Khatib [2] reviewed classical techniques - neural networks, decision trees, and regression - concluding that no single method was sufficient on its own and that combinations performed better, while flagging data imbalance as a persistent challenge.

Later studies moved toward supervised learning and resampling. Lima and Pereira [3] combined feature selection with undersampling to improve detection accuracy, and Roy et al. [4] applied deep learning architectures (ANN, RNN, LSTM, GRU) to credit card fraud, finding GRU most accurate but computationally expensive. Diadiushkin et al. [5] and Victoria Priscilla and Padma Prabha [6] both used Random Forest and XGBoost with resampling (SMOTE/ADASYN) for instant-payment and credit-card fraud, reporting accuracy gains but limited discussion of real-time deployment.

More recent work has emphasized ensembles and explainability. Muhal et al. [7] combined Random Forest, AdaBoost, and LSTM with SMOTE to exceed 99% accuracy, while Jain et al. [8] found gradient boosting and ANN to be the strongest performers among several classical methods. Lande and Sharma [9] optimized Logistic Regression specifically for online payment fraud and achieved very high classification accuracy, underscoring that simpler, interpretable models remain competitive. Ebenezer et al. [10] proposed a Bi-LSTM with attention for UPI and wallet fraud, achieving high accuracy but noting persistent challenges around data imbalance and explainability.

Research gap: Most existing studies optimize primarily for predictive accuracy on benchmark or secondary datasets, with comparatively little attention paid to explaining model decisions or converting predictions into actionable, risk-tiered outputs. This study addresses that gap by combining SMOTE-based balancing, a four-model comparison evaluated on recall as well as accuracy, SHAP-based explainability, and a risk-scoring and early-warning framework suited to real-world monitoring.

3. PROPOSED METHODOLOGY

The proposed pipeline consists of six stages: data collection, preprocessing (missing-value handling, encoding, SMOTE, feature scaling), exploratory data analysis, machine learning model training, model explainability using SHAP, and a fraud risk-scoring and early-warning alert framework.

Dataset: The study uses the Digital Payment Fraud Detection Dataset of India (Kaggle), comprising 7,500 labelled transactions across UPI, Wallet, Card, and NetBanking modes, with no missing values or duplicate rows. Fraudulent transactions make up only 6.5% of the dataset (391 of 5,609+391 in the training split), confirming a strong class imbalance that SMOTE is used to correct in the training set only, after the train-test split, to avoid data leakage.

Tools: The system was implemented in Python 3 on Google Colab, using pandas and NumPy for data handling; Matplotlib and Seaborn for visualization; scikit-learn for Logistic Regression, Decision Tree, Random Forest, train-test splitting, encoding, scaling, and evaluation metrics; imbalanced-learn for SMOTE; XGBoost for gradient boosting; and SHAP for model explainability.

Models and evaluation: Logistic Regression, Decision Tree, Random Forest, and XGBoost were trained on the SMOTE-balanced, scaled training data and evaluated on a common held-out test set of 1,500 transactions (1,402 genuine, 98 fraud) using Accuracy, Precision, Recall, F1-score, and ROC-AUC. The best model was then explained using SHAP feature-importance and summary plots, and its fraud probabilities were converted into a 0-100 risk score mapped to Low (<35), Medium (35-59), and High (≥ 60) risk categories, each linked to a recommended action - approve, request OTP verification, or block for investigation.

Risk Scoring Framework: The selected model's fraud probability is converted into a 0-100 risk score. Scores below 35 are classified Low Risk (Approve, no alert); 35-59 as Medium Risk (Approve with OTP / additional verification); and 60 and above as High Risk (Block for investigation, critical alert), aligning the system's output with the Reserve Bank of India's risk-proportionate authentication guidance (Table 1).

Score	Risk Level	Action
< 35	Low Risk	Approve (No Alert)
35 - 59	Medium Risk	OTP Verification
≥ 60	High Risk	Block for Investigation

Table -1: Fraud risk scoring framework

4. RESULTS AND DISCUSSION

Exploratory analysis confirmed the class imbalance: genuine transactions accounted for 93.5% of the dataset against 6.5% fraud (Fig. 1). Fraud was distributed fairly evenly across transaction type, payment mode, device type, city, IP risk score, transaction amount, hour, and account age, with no single categorical feature dominating - reinforcing the need for a multivariate model rather than simple threshold rules.

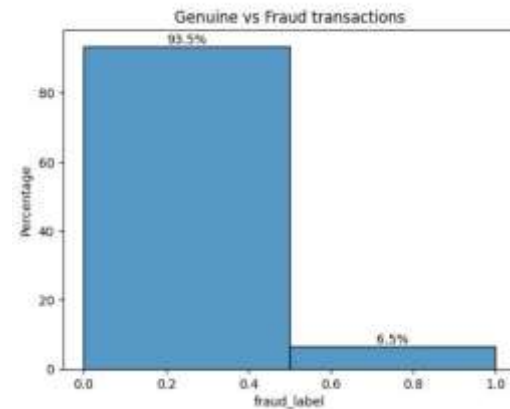


Fig -1: Distribution of genuine vs. fraud transactions

Table 2 summarizes the performance of the four models on the held-out test set.

Table -2: Model performance comparison

Model	Acc.	Prec.	Rec.	F1	AUC
Logistic Reg.	0.639	0.060	0.306	0.100	0.494
Decision Tree	0.771	0.058	0.163	0.085	0.488
Random Forest	0.791	0.068	0.173	0.098	0.512
XGBoost	0.825	0.063	0.122	0.084	0.509

XGBoost achieved the highest overall accuracy (82.5%), but its confusion matrix showed it correctly flagged only 12 of 98 fraud cases in the test set - the fewest of all four models - because it conservatively favours the majority (genuine) class. Logistic Regression, by contrast, recorded the highest recall (30.6%), correctly identifying 30 of 98 fraud cases, at the cost of more false alarms on genuine transactions (474 misclassified as fraud out of 1,402). Since missing a fraudulent transaction is typically more costly than an extra verification step, Logistic Regression was selected as the preferred model despite its lower accuracy. Random Forest offered the most balanced trade-off among the remaining models, catching more fraud than Decision Tree while keeping false alarms lower than Logistic Regression.



Fig -2: Confusion matrix - Logistic Regression

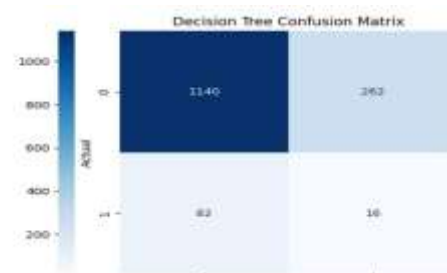


Fig -3: Confusion matrix - Decision Tree

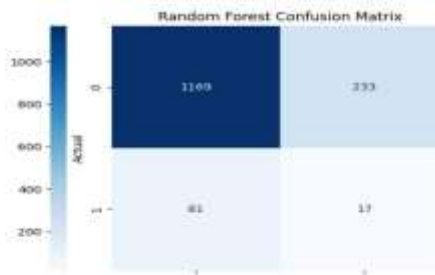


Fig -4: Confusion matrix - Random Forest

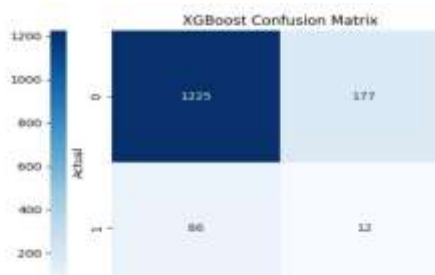


Fig -5: Confusion matrix - XGBoost

SHAP analysis of the Logistic Regression model (Fig. 6) identified payment mode, account age, and prior failed login/payment attempts as the three strongest predictors of fraud, followed by device type and transaction type. IP risk score and the international-transaction flag contributed comparatively little on their own, consistent with the exploratory finding that no single feature separates fraud from genuine activity - supporting the case for a multivariate learned model over static rules.

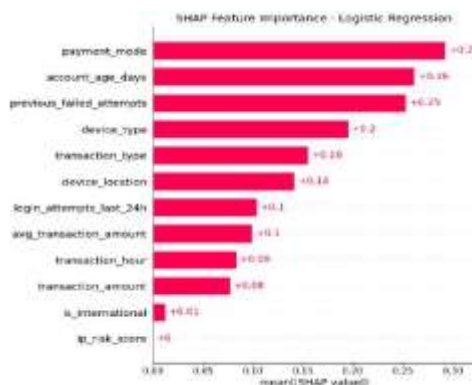


Fig -6: SHAP feature importance - Logistic Regression

The selected model's fraud probabilities were converted into a 0-100 risk score: scores below 35 are Low Risk (no alert), 35-59 are Medium Risk (early warning, additional verification such as OTP), and 60 and above are High Risk (critical alert, immediate investigation). In a sample of ten scored transactions, one was Low Risk, seven were Medium Risk, and two were High Risk, illustrating how the framework converts a raw probability into a directly actionable recommendation for a payment system.

5. CONCLUSIONS

This study developed and compared four machine learning models for fraud detection in digital payment transactions. XGBoost achieved the highest accuracy, but Logistic Regression achieved the highest recall and was selected as the preferred model, showing that accuracy alone is not an adequate criterion

for imbalanced fraud detection tasks - recall and the precision-recall balance matter more when the cost of missing fraud is high. SHAP analysis improved the interpretability of the selected model, and the resulting risk-scoring framework and early-warning alert system make the output directly usable in a real payment monitoring workflow.

The low precision and ROC-AUC values obtained across all models indicate that the models are not yet ready for production use; this is attributable to the modest dataset size, its class imbalance, and the absence of richer behavioural or sequential features. Future work should explore larger and more diverse transaction datasets, behavioural and temporal features, and more advanced algorithms to build a reliable, real-time fraud detection system for digital payments.

ACKNOWLEDGEMENT

The authors thank the Department of Management Studies, Ballari Institute of Technology & Management, for its support and guidance during this study.

REFERENCES

1. Mahdi, M. D. H., Rezaul, K. M., Rahman, M. A.: Credit Fraud Detection in the Banking Sector in UK: A Focus on E-Business. Int. Conf. on Innovative Computing Technology (2010)
2. Al-Khatib, A. M.: Payment Fraud Detection Techniques. International Journal of Computer Applications (2012)
3. Franca Lima, R., Machado Pereira, A. C.: Feature Selection Approaches to Fraud Detection in E-Payment Systems. E-Commerce and Web Technologies (2015)
4. Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., Beling, P.: Deep Learning Detecting Fraud in Credit Card Transactions. Systems and Information Engineering Design Symposium (2018)
5. Diadiushkin, A., Sandkuhl, K., Maiatin, A.: Fraud Detection in Instant Payments as Contribution to Digitalization in Banks. Int. Conf. on Business Informatics Research (2019)
6. Victoria Priscilla, C., Padma Prabha, D.: Influence of Optimizing XGBoost to Handle Class Imbalance in Credit Card Fraud Detection. Int. Conf. on Smart Electronics and Communication (2020)
7. Muhal, H., et al.: An Ensemble Approach for Credit Card Fraud Detection Using Champion-Challenger Analysis. Int. Conf. on Machine Learning and Data Engineering (2022)
8. Jain, N., Chaudhary, A., Kumar, A.: Credit Card Fraud Detection Using Machine Learning Techniques. Int. Conf. on Computing, Communication, and Intelligent Systems (2022)
9. Lande, J., Sharma, J.: Optimizing Logistic Regression for Fraud Detection in Digital Payments. International Journal of Computer Science and Engineering (2025)
10. Ebenezer, Mohan, S., Subalya, S., Jansi Rani, T.: AI-Powered Fraud Detection Framework for UPI and Digital Wallet Transactions Using Bi-LSTM with Attention. Int. Conf. on Artificial Intelligence and Smart Systems (2025)
11. Reserve Bank of India: Authentication Mechanisms for Digital Payment Transactions Directions, 2025. RBI (2025)