iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

MACHINE LEARNING BASED FRAUD DETECTION SYSTEM FOR DIGITAL PAYMENT TRANSACTIONS

Shreyas Palsodkar, Ritika Baranwal, Achintya Warrier, Manthan Nimonkar
Department of Information Technology, D. Y. Patil College of Engineering,
Savitribai Phule Pune University, Pune, India

Mrs. Rajeshwari Thadi
Internal Guide
Department of Information Technology, D. Y. Patil College of Engineering, Pune, India

ABSTRACT
Digital payment fraud has become a critical threat to financial systems, with fraudulent transactions causing significant monetary losses globally. Traditional rule-based detection systems fail to adapt to evolving fraud strategies, as they rely on static thresholds that cannot capture complex behavioral patterns in transaction data. This paper presents a machine learning-based fraud detection system evaluated on the Credit Card Fraud Detection dataset (fraudTrain), containing 1,048,575 transactions with features including merchant category, transaction amount, cardholder demographics, and geographic information. The proposed system addresses the severe class imbalance inherent in fraud datasets through SMOTE oversampling applied exclusively on the training split, followed by feature engineering and XGBoost classification. The model achieved an accuracy of 99.48%, recall of 93.74%, F1-score of 67.58%, ROC-AUC of 0.9987, and PR-AUC of 0.9018, demonstrating strong fraud detection capability on a highly imbalanced dataset. Results confirm that combining correct oversampling methodology with gradient boosting classification provides a scalable and reliable solution for real-time financial fraud detection.

Keywords:
Fraud Detection, Machine Learning, XGBoost, SMOTE, Digital Payments, Anomaly Detection, Financial Security

INTRODUCTION

Global digital payment transaction values exceeded $20 trillion in 2025, with fraud losses projected at $56 billion annually. This scale makes automated fraud detection not just useful, but essential. With the increasing reliance on online platforms, the volume of financial transactions has grown rapidly. However, this growth has also led to a rise in fraudulent activities, posing serious risks to users and financial institutions. Fraudulent transactions can result in financial losses, compromised user data, and loss of trust in digital platforms. Traditional fraud detection systems rely on predefined rules and static thresholds, which are often insufficient to detect sophisticated and evolving fraud patterns. These systems lack adaptability and fail to identify hidden relationships within transaction data.

Unlike rule-based systems that require manual threshold updates, ML models automatically learn decision boundaries from thousands of labeled transactions — adapting to new fraud patterns without human intervention. Specifically, gradient boosting methods can capture non-linear interactions between transaction amount, time, and behavioral features that simple if-then rules completely miss. By analyzing transaction patterns, user behavior, and statistical features, machine learning models can classify transactions more effectively than traditional approaches.

However, existing approaches suffer from two persistent gaps. First, most studies either ignore class imbalance entirely or apply oversampling incorrectly — applying SMOTE before the train-test split, which causes data leakage and artificially inflates reported performance. Second, many high-accuracy models are evaluated only on accuracy, which is a misleading metric when fraud cases represent less than 0.2% of transactions. Evaluated on the Credit Card Fraud Detection dataset, the proposed system achieves a ROC-AUC of 0.9987 and recall of

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

93.74%, demonstrating that correct preprocessing combined with ensemble learning significantly outperforms conventional approaches on real-world imbalanced transaction data.

OBJECTIVES

The primary goals of this study and the resulting system are:

- Develop a machine learning pipeline that accurately detects fraudulent digital payment transactions in real time.
- Address the critical class imbalance problem inherent in financial fraud datasets using correctly applied SMOTE oversampling.
- Implement XGBoost-based classification to capture complex, non-linear patterns in transaction data that rule-based systems fail to detect.
- Demonstrate significant performance improvements over baseline classifiers including Logistic Regression, Decision Tree, and Random Forest.
- Provide a scalable and deployable fraud detection solution applicable to real-world financial institutions processing high transaction volumes.

METHODOLOGY

The proposed system aims to detect fraudulent transactions using machine learning techniques. The methodology consists of multiple stages, including data collection, preprocessing, feature engineering, model training, and evaluation.

The dataset used in this study is the Credit Card Fraud Detection dataset (fraudTrain), containing 1,048,575 transaction records. The dataset includes 22 features: trans_date_trans_time, cc_num, merchant, category, amt, first, last, gender, street, city, state, zip, lat, long, city_pop, job, dob, trans_num, unix_time, merch_lat, merch_long, and the binary target variable is_fraud. Fraudulent transactions represent less than 1% of the total records, making this a severely imbalanced classification problem. The dataset was chosen for its realistic transaction attributes including cardholder demographics, geographic coordinates, and merchant categories, which provide rich signal for fraud pattern learning.

Data preprocessing was performed in four steps. First, the dataset was examined for missing values and duplicate transaction records, which were removed to ensure data integrity. Second, categorical variables including merchant name, transaction category, and cardholder gender were encoded using label encoding and one-hot encoding as appropriate. Third, numerical features such as transaction amount amt and geographic coordinates were normalized using Min-Max scaling to bring all features to a comparable range. Finally, the trans_date_trans_time column was parsed to extract time-based features used in subsequent feature engineering.

The fraudTrain dataset exhibits severe class imbalance, with fraudulent transactions comprising less than 1% of all records. Applying oversampling before splitting would allow synthetic samples derived from test data to appear in training, artificially inflating performance metrics through data leakage. To prevent this, the dataset was first divided into an 80% training set and 20% test set using stratified sampling to preserve the original fraud ratio across both splits. SMOTE was then applied exclusively to the training set, generating synthetic minority class samples by interpolating between existing fraud instances in feature space. The test set was kept in its original imbalanced state to reflect real-world evaluation conditions accurately.

Feature engineering was performed to extract fraud-relevant signals beyond the raw transaction attributes. Four categories of derived features were created. First, time-based features were extracted from trans_date_trans_time, including hour of day, day of week, and a binary flag indicating whether the transaction occurred outside standard business hours (9AM–6PM), as fraudulent activity is disproportionately observed at unusual hours. Second, geographic distance features were computed using the Haversine formula applied to cardholder coordinates (lat, long) and merchant coordinates (merch_lat, merch_long), since an unusually large distance between cardholder location and merchant location is a strong fraud indicator. Third, transaction frequency features were calculated per cardholder over rolling windows of 1 hour and 24 hours to detect sudden spikes in activity. Fourth, a relative amount feature was derived by comparing each transaction amount against the cardholder's historical average, flagging transactions that deviate significantly from established spending behavior.

RESULTS AND DISCUSSION

The proposed fraud detection system was evaluated on a held-out test set of 209,715 transactions drawn from the fraudTrain dataset. Since fraudulent transactions account for less than 1% of all records, accuracy alone is

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

not a meaningful measure of performance — a model that blindly classifies every transaction as legitimate would still exceed 99% accuracy. For this reason, PR-AUC, Recall, and F1-Score are used as the primary indicators throughout this evaluation.

Table I summarises the performance metrics obtained by the XGBoost classifier after training on SMOTE-balanced data.

TABLE I. Performance Evaluation Metrics

Metric	Value	Metric	Value
Accuracy	99.48%	Recall	93.74%
Precision	52.84%	F1-Score	67.58%
ROC-AUC	0.9987	PR-AUC	0.9018

The ROC-AUC of 0.9987 indicates that the model separates fraudulent from legitimate transactions with near-perfect consistency across all decision thresholds. The PR-AUC of 0.9018 is particularly significant — this metric is sensitive to minority class performance, meaning the classifier genuinely learned fraud-specific patterns rather than defaulting to majority class prediction. A recall of 93.74% confirms that the system successfully caught the vast majority of actual fraud cases, which is the most operationally critical outcome in financial security contexts.

The precision of 52.84% reflects a moderate false positive rate, where some legitimate transactions were flagged as suspicious. In practice, this trade-off is acceptable — the financial and reputational cost of missing a genuine fraud transaction far outweighs the cost of reviewing a false alert. Institutions can further adjust the classification threshold depending on their tolerance for false positives versus missed detections.

Table II compares the proposed XGBoost model against three standard baseline classifiers trained under identical preprocessing conditions on the same dataset.

TABLE II. Comparative Model Performance

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	97.8%	31.2%	61.4%	41.4%	0.951
Decision Tree	98.9%	39.7%	74.3%	51.8%	0.871
Random Forest	99.1%	44.6%	79.2%	57.1%	0.981
XGBoost (Proposed)	99.48%	52.84%	93.74%	67.58%	0.9987

The results confirm a clear advantage for the proposed approach. XGBoost outperformed all three baselines across every metric, with the most notable improvements in Recall (+14.54% over Random Forest) and ROC-AUC (+0.0177 over Random Forest). These gains reflect the ability of gradient boosting to model non-linear interactions between transaction features — patterns that simpler classifiers consistently miss.

A key contribution of this work is the correct placement of SMOTE in the pipeline. By applying oversampling exclusively to the training set after the stratified split, the model learned fraud patterns from genuinely balanced data without any synthetic test samples influencing the evaluation. This methodological discipline is what separates the reported results from the inflated metrics commonly seen in prior literature. Combined with the engineered features capturing transaction timing, geographic distance, spending frequency, and relative amount deviations, the pipeline delivers a robust and practically deployable solution for real-time financial fraud detection.

CONCLUSION

This paper presented a machine learning-based fraud detection pipeline built on the fraudTrain dataset, which comprises over one million credit card transactions. The work makes a deliberate methodological contribution: SMOTE oversampling is applied strictly after the train-test split, closing a data leakage gap that has quietly inflated performance figures across a significant portion of existing fraud detection literature.

The XGBoost classifier, trained on SMOTE-balanced data and evaluated on an untouched test set, achieved a ROC-AUC of 0.9987, PR-AUC of 0.9018, recall of 93.74%, and F1-score of 67.58%. Comparison against Logistic Regression, Decision Tree, and Random Forest baselines confirmed that the proposed system outperforms conventional approaches across all primary metrics. The engineered features — time-of-day

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

signals, Haversine-based geographic distance, cardholder spending frequency, and relative transaction amount — proved effective in surfacing behavioral anomalies that raw transaction attributes alone cannot capture.
The main limitation of the current system is its precision of 52.84%, which produces a non-trivial false positive rate. Threshold calibration and cost-sensitive learning are natural next steps to bring this into production-ready range. Evaluation was also conducted on a single dataset, so cross-institutional generalisability remains to be established.

REFERENCES

1) Dhanush Sai, C. C., Amaresh, C. N., and Jancy, S. (2025). Online Payment Fraud Detection Using Exploratory Data Analysis. *Proc. International Conference on Machine Learning and Autonomous Systems (ICMLAS)*.
2) Sharma, A., Gupta, R., and Verma, S. (2024). Aquila Optimization Algorithm with Random Forest for Real-Time Fraud Detection in Financial Transactions. *Proc. ICDSCNC*.
3) Atia, H. A., Aboul-Ela, M., Reyad, C. A., and Awad, N. A. (2024). Online Payments Fraud Detection Using Machine Learning Techniques. *2024 Intelligent Methods, Systems, and Applications (IMSA)*, IEEE, pp. 402–409.
4) Johnson, M. and Wang, L. (2024). Machine Learning Techniques for Financial Fraud Detection: A Comprehensive Survey. *IEEE Access*.
5) S. N., Patil, P., Vidyashree, S. A. M., and A. P. (2025). Enhancing Banking Fraud Prevention Using ML Technologies. *2025 3rd IEEE International Conference on Knowledge Engineering and Communication Systems (ICKECS)*.
6) Kirar, J. S., Kumar, D., Chatterjee, D., Patel, P. S., and Yadav, S. N. (2021). Exploratory Data Analysis for Credit Card Fraud Detection. *2021 International Conference on Computational Performance Evaluation (ComPE)*, pp. 157–161.
7) Gee, S. (2014). *Fraud and Fraud Detection: A Data Analytics Approach*. John Wiley & Sons.
8) Moreira, M. A. L., et al. (2022). Exploratory Analysis and Implementation of Machine Learning Techniques for Predictive Assessment of Fraud in Banking Systems. *Procedia Computer Science*, vol. 214, pp. 117–124.