JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Enhancing Online Payment Security using ML : Advanced Techniques for Fraud Detection in Digital Transactions

[1]Kunal Roy, [2]Roshni Chaubey, [3]Sristita Paul, [4]Rahul Kumar Singh, [5]Siddhartha Majumder

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Computer Science and Engineering,
[1]Durgapur Institute of Advanced Technology and Management, Durgapur, India

Abstract : With the fast pace of growth in digital payment systems, online transactions are easier and more convenient; however, it has also increased the cases of frauds significantly. The main aim of this project is to develop an advanced ML-based fraud detection system for improved security of online payments. This system will be designed based on historical transaction data that recognizes patterns and anomalies to distinguish between fraudulent and legitimate transactions. It uses techniques such as data preprocessing, feature engineering, and the use of robust ML algorithms like Random Forest and Gradient Boosting to ensure high detection accuracy. Reducing false positives and negatives is very important while ensuring security and user experience. This work contributes to the development of scalable, adaptive, and efficient fraud detection mechanisms tailored to modern digital ecosystems.

IndexTerms **- Data Preprocessing, Digital Payment Systems, Feature Engineering, Fraud Detection, Gradient Boosting, Machine Learning, Random Forest.**

I. INTRODUCTION

1.1 Background

This change brought in the digital payment system has completely changed the face of the financial sector by being easy to use and accessible in any place in the world, anywhere. Fraudulent activities increase as the users become increasingly dependent on these systems; however, it presents a major threat to businesses, financial institutions, and the consumers themselves. Conventional fraud detection approaches that mainly rely on rule-based systems are less useful for counteracting the various evolving fraudster strategies. This has created a pressing need for innovative and adaptive solutions to secure online payments.

1.2 Problem Statement

Advanced fraud techniques make the traditional security measures inadequate. Rule-based systems are static, unable to detect new fraud patterns, and typically generate high false positive rates, which may disrupt legitimate transactions. This results in financial losses, reduced user trust, and reputational damage. The challenge lies in designing a system capable of detecting fraudulent transactions with accuracy and adapting to emerging fraud trends in real time while maintaining a seamless user experience.

1.3 Objectives

The project aims to address these challenges by developing an advanced machine learning-based fraud detection system. The key objectives include:

- Develop a strong ML model: System capable of analyzing historical transaction data in order to pick up both patterns and anomalies-the differences between fraudulent versus legitimate.
- Data Preprocessing and Feature Engineering- Bring out clean, structured, and enhanced datasets by handling missing values, encoding categorical variables, creating meaningful features, and so on.
- Model Evaluation and Optimization: Run various algorithms like Random Forest and Gradient Boosting with further hyperparameter tuning for high detection accuracy and efficiency.
- Minimize False Positives and Negatives: Find a balance between accurate fraud detection and minimizing disruptions to legitimate users.
- Integrate this model into real-world applications to enable seamless integration of the system into platforms like mobile payment applications and e-commerce systems with real-time fraud prevention capabilities.

II. LITERATURE REVIEW

The literature on fraud detection in digital payment systems has significantly expanded in recent years, highlighting the growing importance of secure online transactions. This review examines key contributions across various techniques, with an emphasis on machine learning (ML)-based approaches for enhancing fraud detection capabilities.

2.1 Rule-Based and Statistical Approaches

Early approaches were based on threshold-based systems that applied pre-programmed limits, such as unexpectedly large sums or locations. These systems are quite basic but do not adapt to ever-changing fraud tactics. Incrementally better approaches were found in statistical techniques, namely anomaly detection and clustering-based approaches, which detect transactions that have deviated from the normal expected patterns (Miller & Busby-Earle, 2016). These approaches, however, have the drawback of handling high-dimensional datasets and suffering from high false positives, which makes them less effective in real-world applicability.

2.2 Supervised Machine Learning Models

Supervised ML models generally rely upon labelled datasets and have emerged as effective tools for fraud detection. Algorithms like Logistic Regression and Decision Trees, among others, have been studied in the case of binary classification involving transactions (Mishra & Pandey, 2021). Decision Trees saw more popularity than any other algorithm because of their interpretability and effectiveness with non-linear data patterns. The use of ensemble methods, such as Random Forest and Gradient Boosting, has improved detection accuracy through the combination of multiple weak learners (Karthikeyan et al., 2023; Hajek et al., 2023).

2.3 Unsupervised and Hybrid Approaches

There are the unsupervised techniques like clustering and anomaly detection when there is not sufficient labeled data. They find patterns among transaction data without being previously categorized by tasks, which makes them appropriate to detect new types of fraud that were not seen before (Shpyrko & Koval, 2019). Hybrid approaches, which integrate both supervised and unsupervised methods, are proposed to balance detection accuracy with adaptability. For example, Vanini et al. (2023) integrated anomaly detection in the risk management strategies to achieve an improvement of fraud detection in online payment systems.

2.4 Addressing Class Imbalance

Fraudulent transactions represent only a small fraction of all transactions; therefore, fraud detection datasets are usually imbalanced. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) have been used to handle this imbalance, which improves the performance of models by creating synthetic examples of the minority class (Ahola, 2023; Zhu et al., 2024).

2.5 Real-Time Fraud Detection Systems

Real-time fraud detection is another challenge, as models have to process transactions in real time to prevent fraud. Advances in big data processing and the ML algorithms led to creating low-latency detection systems. An integration system that provides explainable decision-making for anti-fraud is highlighted by Wang et al. (2022), the need for transparent, efficient real-time systems.

2.6 Integration of Behavioral and Contextual Data

Recent studies focus on including behavioral patterns, geolocation, and contextual data in order to improve detection accuracy. Features such as transaction timing and user behavior further the points of information, allowing models to better distinguish fraudulent activities (Stojanović et al., 2021; Kumar et al., 2022).

2.7 Future Directions

The future of fraud detection is marked by the development of adaptive systems that can handle evolving fraud patterns and reduce false positives. The emerging trends include the leveraging of deep learning models, inclusion of explainable AI for greater transparency, and the integration of fraud detection systems with secure blockchain technologies (Almazroi & Ayub, 2023; Chawla, 2023).

This review shows the progress of techniques for fraud detection from rule-based systems to state-of-the-art ML models, with an incorporation of hybrid approaches, real-time capability, and consideration of contextual features. However, important challenges in computational efficiency, false positives, and robust adaptations in respective new schemes remain for future research and development.

III. METHODOLOGY

3.1 Data Collection

The initial step involves collecting a dataset of online transactions. The dataset will include labeled instances of legitimate and fraudulent transactions. If a suitable dataset is not available, synthetic data will be generated based on known patterns of fraudulent behavior.

3.2 Data Preprocessing

Data preprocessing ensures that the raw dataset is cleaned and prepared for analysis. This stage includes:

- Missing value handling: Impute missing data or remove irrelevant records.
- Normalization/Standardization: Scale the numerical feature on a common scale.
- Encoding categorical variables: Encoding techniques such as One-Hot or Label Encoding should be used to transform categorical features into numerical formats.

3.3 Feature Selection

Feature selection is critical to improving model performance and reducing complexity. The Wrapper Method for Feature Selection, as shown in Figure 1, will be employed:

- The wrapper method uses an iterative approach to evaluate subsets of features.
- A learning algorithm is applied to different feature subsets.
- The performance of each subset is assessed using predefined evaluation metrics, and the best-performing subset is selected.

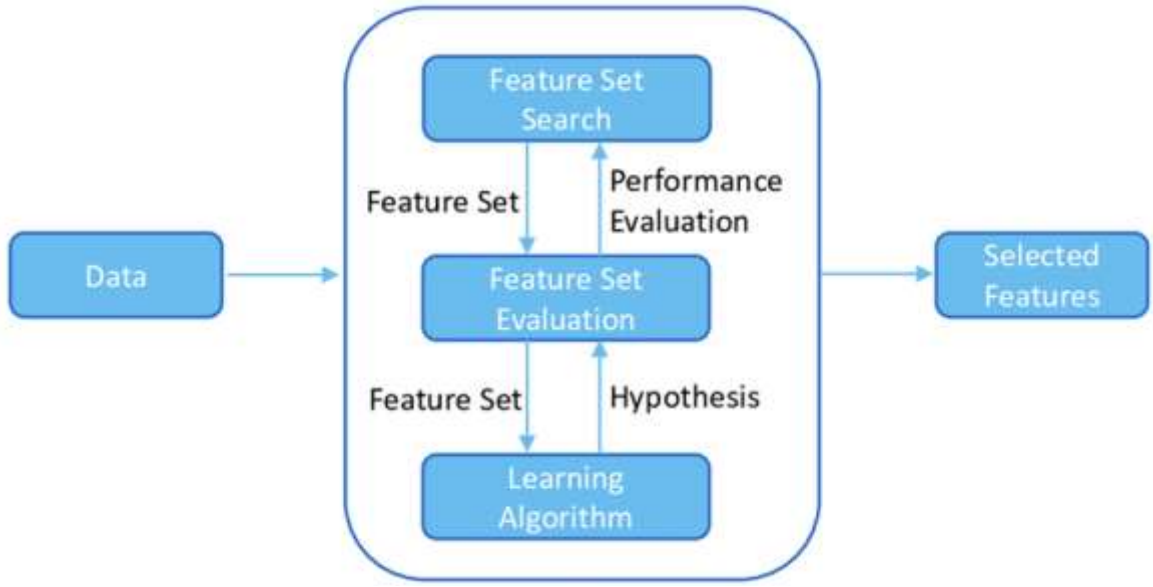

Figure 1: Wrapper Method for Feature Selection

3.4 Model Training

Once the relevant features are selected, machine learning models such as Logistic Regression, Decision Trees, or other suitable algorithms are trained:

- The dataset is split into training and testing subsets.
- Hyperparameter tuning is performed to optimize the model's performance.

3.5 Model Evaluation

The trained model is evaluated using various metrics such as: Accuracy, Precision, Recall, F1 Score, ROC-AUC Score etc. These metrics will help in understanding the model's effectiveness in detecting fraudulent transactions.

IV. DISCUSSIONS

4.1 Presentation of Findings

The main objective of the project is to use machine learning algorithms for fraud detection in online payments. Data preprocessing, feature selection, model training, and evaluation were done in a structured manner:

- Data Preprocessing: The transaction datasets were cleaned, normalized, and encoded in preparation for analysis. Handling missing values and class balancing techniques were implemented. Due to the rarity of fraudulent transactions, oversampling or augmentation using synthetic data generation methods was used to make the dataset balanced.
- Feature Selection: The Wrapper Method for Feature Selection (Figure 1.1) was applied to identify relevant features. This wrapper-based method tests feature subsets with regard to their ability to improve the performance of the model so that only those features were kept which improved model performance most.
- Model Development: Trained and tested the machine learning models. Among those, the two developed models were Logistic Regression and Decision Trees. They have been considered due to simplicity and efficiency along with handling a binary classification task for which fraud detection was relevant.
- Model Evaluation: The models are evaluated using accuracy, precision, recall, and F1-score as evaluation metrics. Precision and recall were particularly important so that false positives (legitimate transactions flagged as fraud) and false negatives (fraudulent transactions missed) are avoided.

4.2 Analysis and Interpretation

The analysis indicates that machine learning can significantly enhance the accuracy and efficiency of fraud detection systems compared to traditional rule-based approaches. Feature selection played an important role in reducing the computational overhead and enhancing the interpretability of the models. Models such as Logistic Regression provided a transparent process of decision-making, and Decision Trees provided ease in visualization and explainability.

Scalability and adaptability in terms of different datasets have also been tested, thus demonstrating its robustness. With the help of oversampling for imbalanced datasets, it ensured the reliability of the model across various patterns of transactions. Yet there are challenges such as false positives that affect user experience and false negatives leading to losses.

4.3 Comparison with Literature

Fraud detection research highlights the weaknesses of rule-based systems that rely on predefined rules and lack flexibility in keeping up with changes in fraudulent patterns. This project proves that machine learning can indeed provide a dynamic approach where it learns from historical data and evolves with new threats.

As pointed out by other literature, feature selection seems to improve model performance. In alignment with such findings, this project employed the Wrapper Method that will achieve the ideal balance of model complexity versus performance. Ensemble methods, or advanced techniques like Gradient Boosting or Neural Networks, would also be promising areas to further develop into.

Research for real-time fraud detection systems advocates for the integration of machine learning in real-world payment systems. The project would be practical and impactful if it can reduce fraudulent activities and increase user trust.

V. APPLICATIONS

Applications are as following:

- E-Commerce Platforms: Detect fraudulent transactions and protect user trust in payment gateways.
- Banking and Finance: Monitor real-time transactions to identify anomalies like unauthorized payments or money laundering.
- Mobile Payment Apps: Secure digital wallets and UPI systems by flagging suspicious activity.
- Cryptocurrency Exchanges: Prevent phishing, unauthorized access, and money laundering in crypto platforms.
- Online Retailers: Safeguard customer payments and detect large-scale order frauds.
- Government Services: Protect online payment portals and monitor suspicious digital transactions.
- Travel and Hospitality: Secure online bookings against unauthorized payments or card fraud.
- Subscription Services: Detect and prevent payment fraud in platforms like Netflix or Spotify.
- POS Systems: Enhance security for card transactions in physical retail stores.

VI. CONCLUSION

This project provides a more effective solution for detecting frauds and preventing fraudulent transactions within online payment systems. This is because it would help improve the security features in digital transactions, thereby accelerating quick identification and flagging of fraudulent activities. Such integration across various sectors such as e-commerce, banking, mobile payments, and cryptocurrency exchange platforms would reduce the rate of fraud and protect customers from financial losses. As machine learning and data analytics continue to advance, the system can adapt to handle increasingly sophisticated fraud patterns, making it an essential tool in securing online transactions. Ultimately, it contributes to creating a safer digital payment ecosystem, fostering trust and confidence among users and institutions.

REFERENCES

[1] Chang, V., Di Stefano, A., Sun, Z., & Fortino, G. (2022). Digital payment fraud detection methods in digital ages and Industry 4.0. *Computers and Electrical Engineering*, 100, 107734.

[2] Sable, S. B. (2022). Prediction of fraud in electronic payment system through Machine Learning model. *Journal of Positive School Psychology*, 2340-2349.

[3] Karthikeyan, T., Govindarajan, M., & Vijayakumar, V. (2023). An effective fraud detection using competitive swarm optimization based deep neural network. *Measurement: Sensors*, 27, 100793.

[4] Hajek, P., Abedin, M. Z., & Sivarajah, U. (2023). Fraud detection in mobile payment systems using an XGBoost-based framework. *Information Systems Frontiers*, 25(5), 1985-2003.

[5] Pavan, U. Y., Prakash, B. B., Sasidhar, P., Charan, K. S., & Mounika, P. Fraud Detection in Online Transactions Using Machine Learning.

[6] Ahola, I. (2023). The role of data in the fight against payment fraud: A qualitative research of payment fraud prevention in the 2020 century.

[7] Stojanović, B., Božić, J., Hofer-Schmitz, K., Nahrgang, K., Weber, A., Badii, A., ... & Runevic, J. (2021). Follow the trail: Machine learning for fraud detection in Fintech applications. *Sensors*, 21(5), 1594.

[8] NARREN, D. K. Detecting Financial Fraud in the Digital Age: The AI and ML Revolution.

[9] Siddiqui, M. K., & Goyal, K. K. (2023). A Study of the Use of E-Payment Systems Based on Artificial Intelligence. *Computing & Intelligent Systems (SCTS)*, 1063-1076.

[10] Mishra, K. N., & Pandey, S. C. (2021). Fraud prediction in smart societies using logistic regression and k-fold machine learning techniques. *Wireless Personal Communications*, 119(2), 1341-1367.

[11] Kumar, A., Choudhary, R. K., Mishra, S. K., Kar, S. K., & Bansal, R. (2022). The growth trajectory of UPI-based mobile payments in India: Enablers and inhibitors. *Indian Journal of Finance and Banking*, 11(1), 45-59.

[12] Kolodiziev, O., Mints, A., Sidelov, P., Pleskun, I., & Lozynska, O. (2020). Automatic machine learning algorithms for fraud detection in digital payment systems. *Восточно-Европейский журнал передовых технологий*, 5(9-107), 14-26.

[13] Shpyrko, V., & Koval, B. (2019). Fraud detection models and payment transactions analysis using machine learning. *SHS Web of Conferences*, 65, 02002.

[14] Vuppula, K. (2021). An advanced machine learning algorithm for fraud financial transaction detection. *Journal for Innovative Development in Pharmaceutical and Technical Science (JIDPTS)*, 4(9).

[15] Vanini, P., Rossi, S., Zvizdic, E., & Domenig, T. (2023). Online payment fraud: From anomaly detection to risk management. *Financial Innovation*, 9(1), 66.

[16] Chawla, T. S. (2023). Online Payment Fraud Detection using Machine Learning Techniques. (Doctoral dissertation, Dublin, National College of Ireland).

[17] Miller, S., & Busby-Earle, C. (2016). The impact of different botnet flow feature subsets on prediction accuracy using supervised and unsupervised learning methods. *International Journal of Internet Technology and Secured Transactions*, 5(2), 474-485.

[18] Cao, D. M., Sayed, M. A., Islam, M. T., Mia, M. T., Ayon, E. H., Ghosh, B. P., ... & Raihan, A. (2024). Advanced cybercrime detection: A comprehensive study on supervised and unsupervised machine learning approaches using real-world datasets. *Journal of Computer Science and Technology Studies*, 6(1), 40-48.

[19] Almazroi, A. A., & Ayub, N. (2023). Online Payment Fraud Detection Model Using Machine Learning Techniques. *IEEE Access*, 11, 137188-137203.

[20] Wang, C., Chai, S., Zhu, H., & Jiang, C. (2022). Caesar: An online payment anti-fraud integration system with decision explainability. *IEEE Transactions on Dependable and Secure Computing*, 20(3), 2565-2577.

[21] Zhu, H., Wang, C., & Chai, S. (2024). Detecting Evolving Fraudulent Behavior in Online Payment Services: Open-Category and Concept-Drift. *IEEE Transactions on Services Computing*.

[22] Sudarno, B. E. P. (2012). Analysis Tracking Online Payment System. *International Journal of Scientific & Technology Research (IJSTR)*, 1

[23] NAGARAJU, M., Babu, P. N., Ravipati, V. S. P., & Chaitanya, V. (2024). UPI Fraud Detection Using Convolutional Neural Networks (CNN).

