Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Digital Transaction Fraud Detection using Machine Learning Technique

Dr. K. Nageswara Reddy[1], T. Khaleedha[2], C. Om Kiran[2], P. Lasya Devi[2], B. Manoj Kumar[2]

[1]Associate Professor, Dept. of Computer Science and Engineering, RGM College of Engineering and Technology, Nandyal (Dist), AP, India.
[2]Students, Dept. of Computer Science and Engineering, RGM College of Engineering and Technology, Nandyal (Dist), AP, India.
Email: knreddy221@gmail.com

Digital transactions play an essential role in modern economies, facilitating swift and convenient monetary exchanges. However, they are increasingly susceptible to fraudulent activities such as phishing, data breaches, and unauthorized transactions. This paper aims to minimize fraudulent practices in digital transactions using Machine Learning algorithms. The proposed system shows better accuracy compared to existing systems.
Keywords: Machine Learning, Financial Security, XGBoost Algorithm, Fraud Detection, Gradient Boosting, Digital Transactions, Transaction Fraud, Ensemble Learning, Precision and Recall, Data Pre-processing.

1. Introduction

Digital transactions have changed the face of financial operations completely, making it faster, secure, and user-friendly as a mode of payment. With these benefits come the corresponding risks in fraud cases. The growth in digital payment has correspondingly increased the rate of fraudulent activities, which include phishing, identity theft, and unauthorized transactions, among others. There is an added need to strengthen fraud detection mechanisms. Current mechanisms in fraud detection face limitations in both accuracy and scalability. To cater to these problems this paper proposes a system which uses the machine learning algorithms to achieve better precision and processing ability than the traditional algorithms of machine learning.

The features appropriate selection is an important step towards devising an efficient fraud detection model, considering the fact that online transactions have high dimensionality and contain numerous features that may not all be relevant for fraud detection.

Research on online payment fraud detection has advanced considerably in the last few years,

and several machine learning techniques are used. Fraud has been a problem since the beginning of ecommerce. However, the increase in online shopping during the COVID-19 pandemic gave the scammers a new playground. In 2020, online shopping scams accounted for 38% of all scams reported worldwide, a massive increase from 24% before the pandemic. This percentage has lowered with the easiness of the pandemic, yet security breaches carry a heavy price in dollars and cents. In 2022, the e-commerce industry lost more than $40 billion due to online payment fraud. So, the fraud detection and prevention market in e-commerce is anticipated to grow extensively with a value increase of over 100% between 2023 and 2027, eventually exceeding more than$ 100 billion [1].

Another area of interest is "idea drift," which refers to how the underlying distribution of datasets evolves over time. Just as the buying habits of customers change, so do fraudsters, updating their tactics in response. Identifying and preventing these scams requires professionals who are constantly evolving, just like the fraudsters, which may eventually lead some fraudulent tactics to become outdated. Since fraud is illegal procurement of products, there has to be some effective fraud detection systems (FDS) for the observation of transactions and monitoring any suspicious behaviour[7].

These systems review the history of a transaction and are capable of classifying it as being either a genuine or a fraudulent transaction based on the history of transaction using techniques of machine learning and data mining. The use of these approaches can on the other hand forestall the emergence of fraudulent transactions hence reducing risks associated with online payment.

2. Literature Review

The rapid growth in digital transactions has completely changed the face of finance by offering faster, more secure, and more user-friendly ways of making payments. Meanwhile, that trend brought about a substantial rise in fraud, necessitating better fraud detection techniques. Digital payments have transformed the way we direct finances. The platforms are quicker and convenient, but they introduce new risks, such as phishing, identity theft, and unauthorized transactions, representing challenges not only for the users but also for financial institutions.

Understanding how machine learning helps detect fraud involves knowing the different types of digital transaction fraud. These are credit card fraud, identity theft, phishing, and chargeback fraud. Credit card fraud occurs when someone uses stolen credit card information to make purchases without the owner's consent. Identity theft is when someone steals personal information and uses it to open fake accounts or make unauthorized transactions. Chargeback fraud occurs when a customer falsely claims that a valid transaction was unauthorized, resulting in a chargeback. Phishing occurs when fraudsters trick people into sharing sensitive information such as passwords or credit card numbers through fake emails or websites [2].

By way of illustration, Zoho Payments has its own models for assessing payment irregularities in order to fraud. PayPal in a similar manner employs machine learning algorithms in the assessment of transactions therefore reducing their losses connected with fraud to a very great extent. On its part, Stripe performs deep learning models in order to uncover intricate patterns of fraud throughout the course of fraud detection and guarantees safe transfer of payments on

behalf of its customers. Sophisticated machine learning technologies are incorporated into the fraud prevention solutions including IBM's Truster and FICO Falcon Platform aimed at delivering holistic fraud management. Payment providers that incorporate machine learning and other strategies guarantee that the merchants are able to protect their payment systems and customer data and have trust in digital economy [2].

Case studies from practice and from worldwide experience are informative as far as they explain how businesses and financial institutions manage to avoid and detect frauds. By analysing these examples, one may identify successful strategies and gaps in the prevention of internet scams[10].

The credit card industry's efforts to combat fraud also point to the need for continuous updating of detection techniques and the implementation of robust risk assessment frameworks. As fraudsters continue to develop new methods, advanced fraud detection systems with these capabilities are becoming essential. The growing demand for fraud detection solutions that integrate secure authentication mechanisms and can adapt to emerging threats in real-time is increasingly evident.

Digital transaction fraud detection is an area that needs constant innovation and vigilance. The previous research underlined the complexity of fraudulent activities, emphasizing the need to understand fully the strengths and limitations of current detection methods. Because fraud attempts are sophisticated and attacks can occur through various channels, it is important to identify behavioral anomalies in payment transactions. Algorithms will detect anomalies, and thus fraud can be recognized. [3]

In this paper we use Decision trees, Random Forest and XGBoost algorithms for fraud detection in digital transactions[11]. Decision Trees are successful supervised learning algorithms applied to both classification and regression tasks. The process of decision tree-based learning algorithms involves recursive partitioning of the data based on the thresholds of the features involved. This procedure builds a tree-like structure in decision-making, making them quite valuable in fraud detection, such as in Shah and Sharma's study on credit card fraud. However, it is susceptible to overfitting and has issues with noisy and imbalanced data sets, thus requires pruning among others. Nevertheless, Decision Trees typically act as a good baseline for fraud detection problems [13].

Random Forest is a technique by Biermann (2001), which, as it relates to decision trees, constructs multiple trees and aggregates their outputs to improve both accuracy and robustness. It is applied effectively in fraud detection, for example, in Wang's study on fraudulent users in peer-to-peer markets. It has good performance on high-dimensional data and minimized opportunities to overfit the models; hence, it is widely used in financial anomaly detection. However, the studies of Isangediok and Gajamannage highlight how its computational costs increase with bigger datasets. Still, Random Forest is a good candidate for fraud detection because it's reliable and a high-performance algorithm.

XGBoost is an advanced gradient boosting algorithm developed by Chen and Guestrin in 2016, which is very efficient and scalable. It grows trees sequentially by correcting errors at each iteration, and it applies regularization techniques to avoid overfitting. Research like the comparison of fraud detection methods made by Niu et al. demonstrates its ability to handle

imbalanced datasets and detect complex patterns in financial transactions. Its application in mobile payment fraud detection also shows it as versatile and strong [4].

3. Problem Statement Identification

The rapid growth of fraud committed through digital transactions is a serious concern for the consumers and businesses alike as fraudsters keep on changing their ways of exploiting weaknesses in the payment systems. The present fraud prevention systems are somewhat inadequate as they are always left behind by the ever-increasingly sophisticated attacks and as a result – there are huge financial losses and loss of customer trust. The diverse nature of fraudulent activities comprising quite a few types of frauds such as impersonation, phishing, friendly fraud, and also forced fraud, and account grabbing, etc. make detection very hard[8].

Although machine learning capability has improved tremendously such as accuracy, bottlenecks and inappropriate targeting remain and so there is still a demand for more sophisticated robust systems that require less latency while targeting and are able to reduce the false alarm rate. The objectives of this study are to investigate the creation and application of more sophisticated algorithms for fraud detection, such as, based on machine learning and behaviour analysis, with the purpose of improving the precision and efficiency of digital transaction systems over security issues.

4. Methodology

A. Dataset Information:

The dataset is selected from an online source named kaggle, which contains nearly 11 columns containing various criteria of the project. A mix of different types of values are present which gives broad view of transaction details.

The dataset is composed of several attributes that characterize the financial transactions. For instance, type indicates the classification of transaction performed and the step characteristic is in the form of time unit. The amount feature shows the total sum involved in the transaction. The account that starts the transaction is called nameOrig and the account that receives it is called nameDest. The characteristics OldbalanceOrg and NewbalanceOrig correspond to the sender's account balance before and after the transaction respectively. Equally, OldbalanceDest and NewbalanceDest correspond to the account balance of the receiver before the transaction and after the transaction respectively.

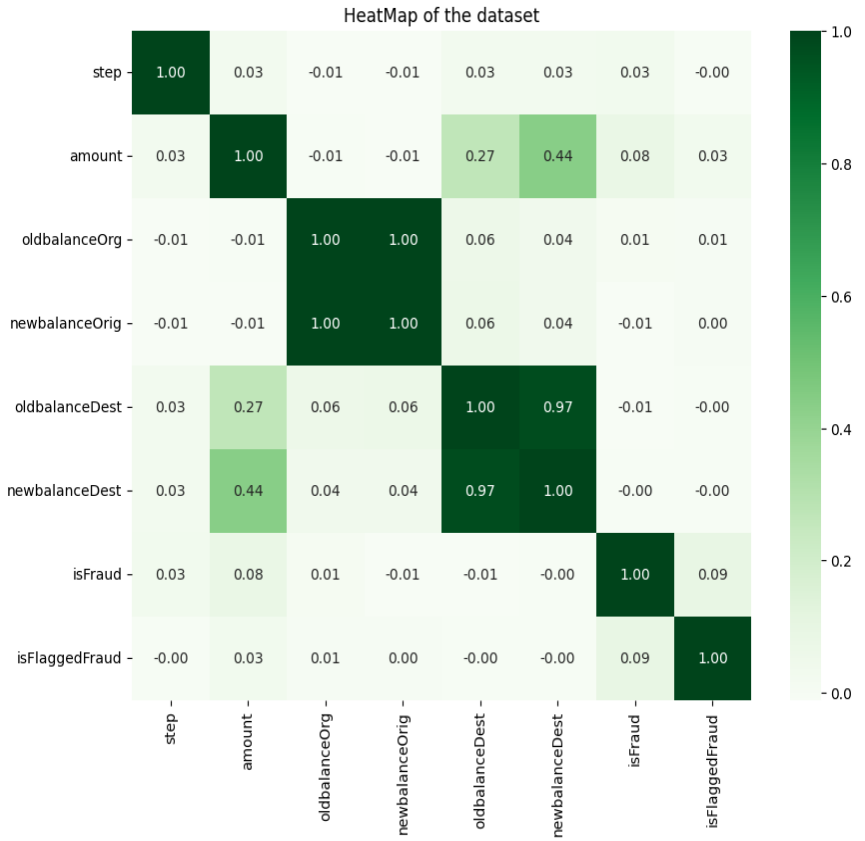

Fig. 1. Heat Map of dataset

A binary variable "isFraud" denoting whether the transaction is fraudulent or not, I've been tagged as 1 if it's fraud and 0 if not. In addition, the transactions that have been flagged as suspicious by the anti-fraud detection system are marked with the isFlaggedFraud feature (flagged 1, other 0). The overall data set provides useful information on the discovery of fraud-patterns in digital transactions, although dealing with class imbalance issues combined with various categorical and numerical features poses challenging requirements regarding pre-processing and the utilization of strong machine learning techniques in producing robust fraud detection systems.

B. Data Collection and Pre-processing:

The first step in a machine learning pipeline is data collection for model training. Data collection refers to the process of gathering information from different sources to answer relevant questions. Common issues that may arise during data collection include unreliable data, missing values, and data imbalance. These issues are addressed through data pre-processing performed on the collected data.

Most likely, inconsistencies will arise in the dataset as data is usually very partial and not systematic, thereby bringing inconsistency into the dataset and hence into the process, thus requiring pre-processing after the stage of data collection to clean and tidy it for use in building

models using machine learning.

Relevant transaction datasets, including historical and simulated fraudulent data, are collected. Pre-processing steps include handling missing values, data normalization, and feature engineering to ensure the dataset is suitable for model training. Features such as transaction amount, step, isFraud, isFlaggedFraud, and user behaviour patterns are selected for analysis.

C. Data Cleaning:

Data that has been erroneously added or misclassified can be removed either manually or automatically. Common methods for addressing missing values in machine learning systems include standard data imputation techniques such as mean, median, and standard deviation, which help balance or fill in the gaps effectively.

D. Model Selection:

Once the clusters have been created, the best model for each is selected from among three algorithms, namely Decision Trees, Random Forest, and XGBoost. GridSearch finds the optimal parameters for each of the algorithms and ensures the best performance from these models. All of these measures - Accuracy, Precision, Recall, and F1-score- are computed on each cluster as well. Then, the best model is selected based on its performance score regarding all the metrics tested, and it is used for the corresponding cluster. This way, the process of fraud detection will be customized and optimized for each specific cluster.

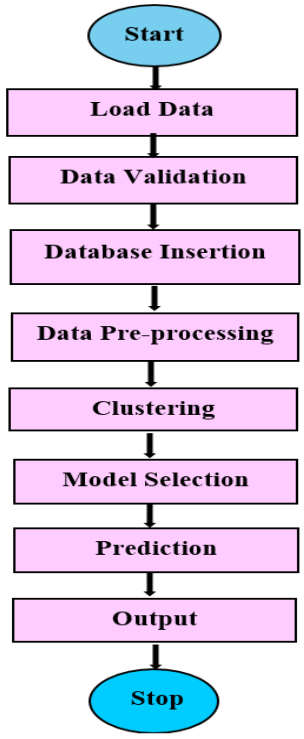

Fig. 2. Flow chart of proposed system.

E. Training and Validation:

The dataset is split into training and testing sets in an 80:20 ratio. The training process involves hyper parameter tuning to achieve optimal learning rates, tree depths, and minimum child weights. Cross-validation techniques are employed to validate model performance.

F. Evaluation Metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FP}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$ERR = \frac{FP + FN}{P + N}$$

5. Results and Discussion

TABLE I: Comparison of Models with Accuracy, F1 Score, Precision, and Recall

S.NO	Model	Accuracy	F1 Score	Precision	Recall
1	Decision Tree	96.88750	45.161	77.78	31.8181
2	Random Forest	97.9187	68.292	73.684	63.6364
3	XGBoost	99.93750	75.000	83.33333	68.1818

To evaluate the performance of the machine learning algorithms (Decision Tree, Random Forest, XGBoost), following metrics were used: Accuracy, Precision, Recall, F1-score. For each cluster XG Boost has given better results when compared with Decision Tree and Random Forest algorithms.

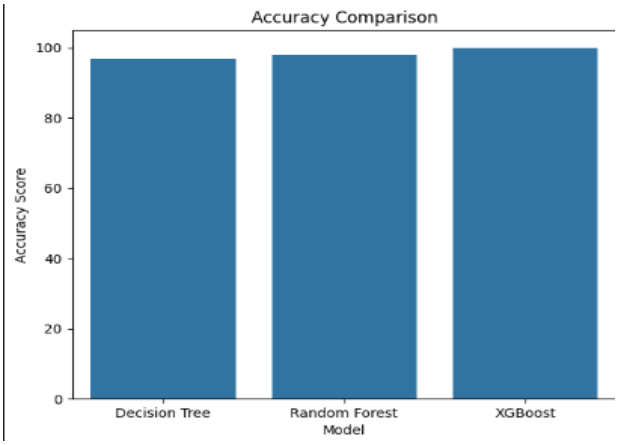

Fig. 3. Accuracy Graph

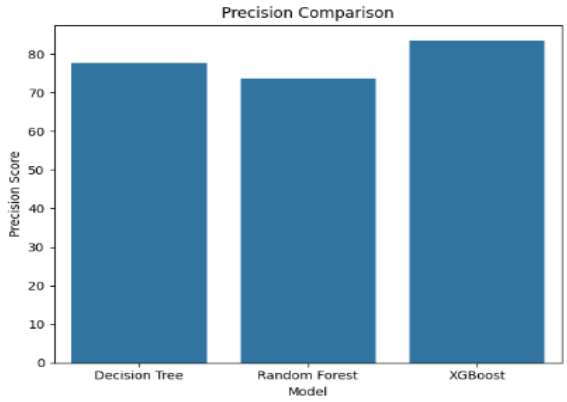

Fig. 4. Precision Graph

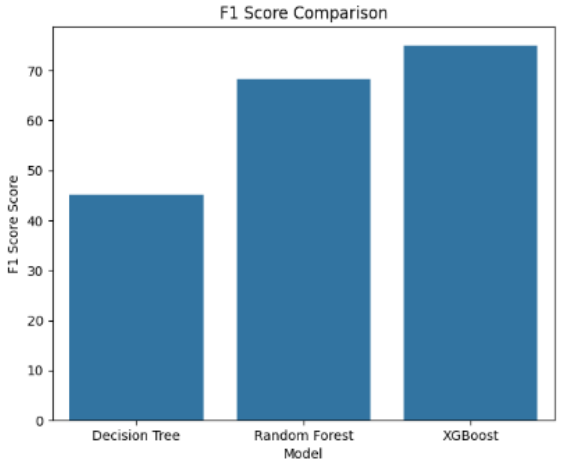

Fig. 5. F1 Score Graph

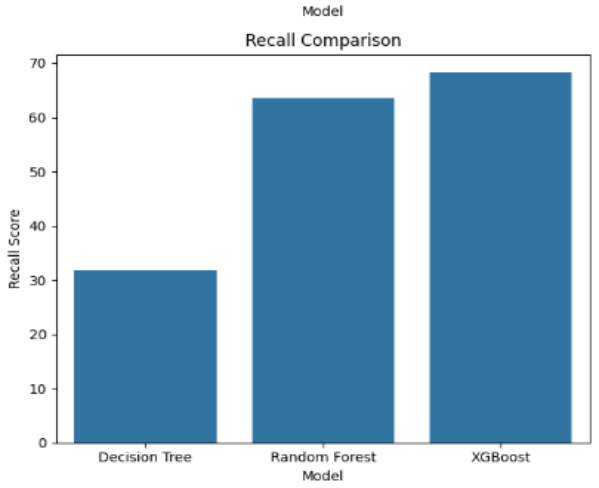

Fig. 6. Recall Graph

6. Conclusion

In this project, Decision Trees, Random Forest, and XGBoost were used to detect fraud using metrics such as Accuracy, Precision, Recall, and F1-score. XGBoost outperformed the rest with the highest accuracy at (99.94%).In summary, XGBoost is the most reliable model to detect fraudulent transactions.

References

[1] S. R. Department. "E-commerce fraud overview." (2024), [Online]. Available: https://www.statista.com/topics/9240/e-commerce-fraud/#topicOverview

[2] Z. Payments. "Machine learning fraud detection." (), [Online]. Available: https://www.zoho.com/in/payments/academy/fraud-and-risk-management/machine-learning-fraud-detection.html

[3] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," IEEE Access, vol. 10, 2022. doi: 10.1109/ ACCESS.2022.3166891.

[4] S. LEI, K. XU, Y. HUANG, and X. SHA, "An xgboost based system for financial fraud detection," E3S Web of Conferences, vol. 214, p. 02 042, Dec. 2020. doi: 10 . 1051/e3sconf/202021402042.

[5] M. Farouk, N. Ragab, D.-D. Salama, et al., "Fraud$detectionml$: Machine learning based on online payment fraud detection," Journal of Computing and Communication, vol. 3, pp. 116–131, Feb. 2024. doi: 10.21608/jocc.2024.339929.

[6] E.-A. Minastireanu and G. Mesnita, "Light gbm machine learning algorithm to online click fraud detection," J. Inform. Assur. Cybersecur, vol. 2019, p. 263 928, 2019. doi: 10.5171/2019.263928.

[7] I. Vejalla, S. P. Battula, K. Kalluri, and H. K. Kalluri, "Credit card fraud detection using machine learning techniques," in 2nd International Conference on Paradigm Shifts in Communications, Embedded Systems, Machine Learning and Signal Processing (PCEMS), 2023, isbn: 979-8-3503-1071-9.

[8] K. D. Kadam, M. R. Omanna, S. S. Neje, and S. S. Nandai, "Online transactions fraud detection using machine learning," International Journal of Advances in Engineering and Management (IJAEM), vol. 5, no. 6, pp. 545–548, Jun. 2023. [Online]. Available: https://www.ijaem.net.

[9] A. N. Ahmed and R. Saini, "A survey on detection of fraudulent credit card transactions using machine learning algorithms," in 2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT), Manipal University, Jaipur, 2023.

[10] A. Ali, S. A. Razak, S. H. Othman, et al., "Financial fraud detection based on machine learning: A systematic literature review," Applied Sciences, vol. 12, no. 19, p. 9637, 2023. doi: 10 . 3390 / app12199637. [Online]. Available: https://doi.org/10.3390/app12199637.

[11] V. Chang, L. M. T. Doan, A. Di Stefano, Z. Sun, and G. Fortino, "Digital payment fraud detection methods in digital ages and industry 4.0," Computers and Electrical Engineering, vol. 100, p. 107 734, 2022, issn: 0045-7906.doi:https://doi.org/10.1016/j.compeleceng.2022.107734.[Online].Available:https://www.sciencedirect.com/science/article/abs/pii/S0045790622000465

[12] A. A. Almazroi and N. Ayub, "Online payment fraud detection model using machine learning techniques, IEEE Access, vol. 11, pp. 137 188–137 203, 2023. doi : 10.1109/ACCESS.2023.3339226 .

[13] R. O. Shah, "Online payment fraud detection using machine learning techniques," IEEE Access, vol. 8, pp. 150 999–151 010, 2020.