







Paper Type: Original Article

Machine Learning Model for Detecting Fraudulent Transactions on the Ethereum Blockchain

Ayman Mohamed Mostafa ^{1,2,*} , Ehab R. Mohamed ¹ , Reham Medhat ¹ , Asmaa Hanafy ¹ 

¹ Faculty of Computers and Informatics, Zagazig University, Zagazig, 44519, Egypt.

Emails: am_mostafa@zu.edu.eg; ehab.rushdy@gmail.com; rmkhalil@fci.zu.edu.eg; asmaa_289@zu.edu.eg.

² College of Computers and Information Sciences, Jouf University, Kingdom of Saudi Arabia.

Received: 04 Apr 2024

Revised: 05 Jul 2024

Accepted: 30 Jul 2024

Published: 01 Aug 2024

Abstract

Cloud-based infrastructure can offer the required processing and storage capacity to manage massive transaction data. Cloud services increase centralization by relying on a single cloud provider, which may expose risks. Even though the cloud has a strong identity and access management system to control access, security issues might still arise. We will leverage the potential of blockchain technology along with the cloud services' scalability and adaptability to tackle these issues. Although these approaches show great potential, the issue still lies in the constant evolution of fraudulent tactics within a dynamic Ethereum ecosystem. This work combines blockchain technology with machine learning algorithms to detect anomalies in Ethereum transactions. There are various scenarios in which these scams happen, including tracking actions and monitoring transaction data. It is observed that the XGBoost algorithm outperforms with an accuracy of 99.39%. Moreover, an application for cryptocurrency transactions is integrated with the fraud detection module. As a result of the experience, cryptocurrency ecosystems already have reliable fraud detection mechanisms in place. The validation metrics exhibit a similar range, indicating that the models are not over-fit. The results show that the SMOTE oversampling techniques improve the classification F1 score levels to 98.61 with an AUC of 100%. These techniques offer a 50/50 class balance for detecting Ethereum transaction fraud.

Keywords: Blockchain; Ethereum; Fraud Detection; Machine Learning; XGB Classifier.

1 | Introduction

Blockchain technology's advancement and the accompanying economic environment have led to a significant surge in cryptocurrency. The emergence of digital currencies—particularly Ethereum—has coincided with a negative climate of hype, fraud, and gambling [1]. Ethereum has experienced numerous hacking attempts as its popularity has grown. One of its primary goals is to restore people's authority and empower them to control their data and transactions. Even with this widespread use, many people are still ignorant of the dangers that come with it, such as sophisticated cybercrime that targets cryptocurrency. With Ethereum, you may transmit cryptocurrencies to anyone for a tiny fee. It uses Ether, a specialized crypto coin, to make multi-user transactions possible. Since the inception of this technology, scammers have used a variety of methods and strategies to work in tandem to undermine and obtain unauthorized access to the Ethereum architecture. Users may suffer large financial losses due to these activities, making transaction data security extremely



Corresponding Author: am_mostafa@zu.edu.eg



Licensee International Journal of Computers and Informatics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

difficult [2]. The blockchain network on which Ethereum is built is vulnerable to network-level attacks like the 51% attack.

When most of the processing capacity of the network is possessed by one person or organization. It can be exploited to alter transaction data, commit double spending, or interfere with normal network operations. These vulnerability schemes represent the main risks to these Ethereum-based transactions [3].

Detecting fraud in Ethereum transactions, particularly through smart contracts, has gained significant attention due to the rise of scams like Ponzi schemes. Numerous applications have been created globally to quantify, track, and improve Ethereum fraud detection. Our scheme required a robust model that could detect fraudulent activities on the Ethereum platform. A study of this kind utilizes machine learning models including logistic regression (LR), light gradient boost (LGBM), extreme gradient boost (XGB), k-nearest neighbors, random forest (RF), and support vector machine (SVM) [4]. A strategy for identifying a transaction as fake or authentic is presented in this paper using the XGB classifier. XGBoost, a powerful machine learning algorithm, has been employed alongside other techniques to enhance detection accuracy. This work has carefully examined and contrasted these various classifiers using a variety of performance metrics. In this work, the accuracy rate of the XGB classifier was 99.39%, whereas the accuracy rate of the LGBM model was 99.1%. The machine learning models must be trained with enough data to anticipate fraudulent activity in the transaction. The XGB algorithm can identify anomalous transactions without exposing the model to overfitting. It accurately and efficiently forecasts fraudulent transactions on the Ethereum platform with minimal memory usage. These machine-learning models are implemented with no user participation. The model opens the door for implementing reliable detection systems for fraud in cryptocurrency ecosystems by showcasing the potential of ensemble models in Ethereum fraud detection.

Similar-scope researchers required a model to characterize the factors that usually affect fraudulent transactions. Another study uses XGB to classify fraud with a 96.82% accuracy rate [5]. With their explainable AI (XAI) description of the machine learning classification, these studies serve as inspiration for the proposed study. Another study created a deep learning model that combined a genetic algorithm with the cuckoo search [6]. The results of this study improve the XGB classifier's performance by 99.7% [7]. For Ethereum fraud detection, a pre-trained binary encoding and recording approach model is also utilized [8].

1.1 | Research Gap

Current Ethereum fraud detection methods don't solve the following issues. Consequently, the following serve as the driving forces behind the suggested work:

- The sole goal of the current research is to enhance the model's ability to identify fraudulent transactions.
- The oversampling methods offer no opportunity to rectify the class imbalance.
- There is an inability to describe how attributes affect the model's performance, and none of these models are transparent.
- The anomalous accounts that give rise to these illicit activities are not being identified.
- In the dataset used for detecting Ethereum fraud, handling class imbalance is not demonstrated.

1.2 | Contributions to the proposed work

Comparing the proposed work to the prior implementations, there are numerous improvements, such as the following:

- The utilization of XAI models and oversampling strategies in the suggested work enhances the dependability and credibility of the fraud detection procedure.

- Identifying the addresses of the accounts involved in the transaction that was flagged.
- It is possible to identify and reduce the features that often affect fraud detection.
- Several oversampling strategies address the class imbalance and improve the model's performance.

The remainder of the paper is coordinated as follows. Section 2 contains the literature summary for this study. The categorization models and fundamental ideas of the suggested approach are discussed in Section 3. Experimental setup, efficiency measures, and data processing for multiple approaches are covered in Section 4. The suggested study's conclusion and a plethora of suggestions for future further research are presented in Section 5.

2 | Literature Review

To provide frameworks for detecting fraud in Ethereum transactions, recent research has looked into several neural networks and artificial intelligence models [9-13] highlighting the possible interpretation of Ethereum fraud as any illicit technique or activity that results in measurable losses or profits. The Ethereum scam is found using a detection system based on data mining. This study illustrates researchers that use the XGBoost classifier in fraud detection and demonstrate other comparative techniques.

2.1 | XGB classifier in Fraud Detection

Kumar et al. [5] identify suspicious accounts on the Ethereum blockchain by analyzing both EOA and smart contract accounts. The model identified malignant nodes using the XGB classifier with an accuracy of 96.82% based on fraud detection in the account's transaction data. The suggested method has produced excellent results by examining both accounts, with a false positive rate of only 3%. A convolutional neural network and an XGBoost classifier are combined by Dutta et al. [7] to provide a highly successful method for detecting fraudulent accounts on the Ethereum blockchain. With an average AUC of 0.9988 and an astounding accuracy of 99.7%, the model demonstrated its potent ability to discriminate between legitimate and fraudulent accounts. The model not only improves accuracy but also enhances scalability and generalizability.

Rathore et al. [16] employ a specific XGBoost model to locate the bogus addresses in the Ethereum cryptocurrency. Vast trials have proven that the model is highly effective in operation in a real environment, as evidenced by its accuracy of over 96% on test data. Using XGB classifiers, Walavalkar et al. [17] suggest a token-based method to identify fraud in Ethereum transactions that integrates the ERC20 standard. After a thorough assessment, the XGB classifier produced the best results, detecting fraud with an accuracy of 95%. Crisostomo et al. [18] tackle the problem of detecting malicious accounts for the Ethereum blockchain by combining the techniques of an auto-encoder and an XGBoost. With an accuracy rate of 98.86%, the proposed model outperforms the model tested by XGBoost without using an auto-encoder. The experiment's positive outcomes showed a 12.07% improvement in precision-recall AUC measurements. Sallam et al. [19] propose a solution that employs various ML techniques to detect and identify patterns of fraudulent accounts with the Ethereum network. Using the XGB model, this investigation achieved an accuracy of 96.8% with an average AUC of 0.99%. The authors admit that the size and extent of the dataset employed were constrained, making it insufficient to increase the reliability of the findings.

2.2 | Complementary Techniques

Pahuja and Kamal [1] introduce a classification model based on the CRISP-DM framework to identify fraudulent transactions on the Ethereum blockchain. Notably, the LGBM classifier achieved an impressive accuracy of 99.2%. The authors also claim that their proposed method surpasses existing state-of-the-art techniques when applied to similar datasets. With the help of the Euclidean distance methodology, Aziz et al. [4] established a method for detecting Ethereum fraud transactions with limited characteristics, LGBM. With a slightly higher performance of 99.17%, the modified LGBM technique outperforms the other recommended comparable models.

Aziz et al. [6] introduce a novel approach that combines deep learning with a unique optimization strategy known as GA-CS. This hybrid method aims to improve the detection of fraudulent transactions by addressing the limitations of existing techniques. The model demonstrates a high accuracy rate of 99.71%, outperforming other traditional ML techniques.

Fraud may be efficiently identified and mitigated in Ethereum transactions using an ensemble model called CAT Boost, as demonstrated by Ravindranath et al. [14]. After being oversampled, the model achieves remarkable accuracy, ranging from 97% to 98.42%. The suggested work uses Ethereum data to create policies for data consistency tests. Zhang et al. [15] developed an enhanced version of the Light GBM algorithm to identify Ethereum Ponzi schemes using characteristics from opcodes, bytcodes, and user data. The method allows for faster processing, making it more efficient for real-world applications where timely detection is critical. Investors who want to quickly evaluate smart contracts and reduce the dangers connected with Ponzi schemes will find this model review crucial.

3 | The Proposed Methodology

Several actions are taken in this architecture, as shown in Figure 1, to find the fraudulent account on the Ethereum network.

3.1 | Data Acquisition

The dataset can be obtained using external APIs by scraping logs from various web servers or census databases. However, the Ethereum fraud detection dataset collected from the Etherscan and Kaggle platforms [20] consists of 9841 instances with 2179 fraudulent accounts (22.1%). Additionally, 7662 legal accounts (77.9%) are included, and 51 features overall per instance were employed in this work. Figure 2 depicts that the suggested dataset is imbalanced and needs efficient oversampling techniques for handling the issue. The extracted features consist of the total number of sent and received regular transactions, the time difference between the first and last transactions, and the average time in minutes between sent and received transactions, etc.

3.2 | Data Preprocessing

The act of merging, arranging, and structuring data is known as data cleaning. A variety of irregularities are often found in data, such as empty columns, missing values, and non-uniform data formatting. For this purpose, data must be analyzed, investigated, and prepared before creating the necessary model. Finding the pertinent data that must be incorporated into analytics to guarantee information delivery is another aspect of data cleaning. Some of the causes of data unreliability include missing values, duplicate examples, incorrect labels, and inaccurate feature values. For example, the model manages NAN values and deletes features like the ERC20 token, index, and address, which contain the least amount of information. We plan to preprocess the data by developing a transformer class that facilitates the Scikit-learn API. It creates new features, codes categorical features, fills in gaps, and handles outliers while processing numerical features. Therefore, utilizing the traditional scaler method, the proposed strategy scaled the feature for machine learning algorithms.

3.3 | Data Visualization

Because the mind is so transparent, visualization is essential to the computation of linguistic theory. The data has 42 dimensions and can be shown in a variety of ways. Using techniques for dimension reduction is a strong solution. This work summarizes the key features of datasets using box plots and histograms. This could highlight oddities, correlations, and tendencies that aren't always immediately apparent. Dimensionality reduction techniques are utilized to decrease the dimensions of a high-dimensional dataset to two or three dimensions.

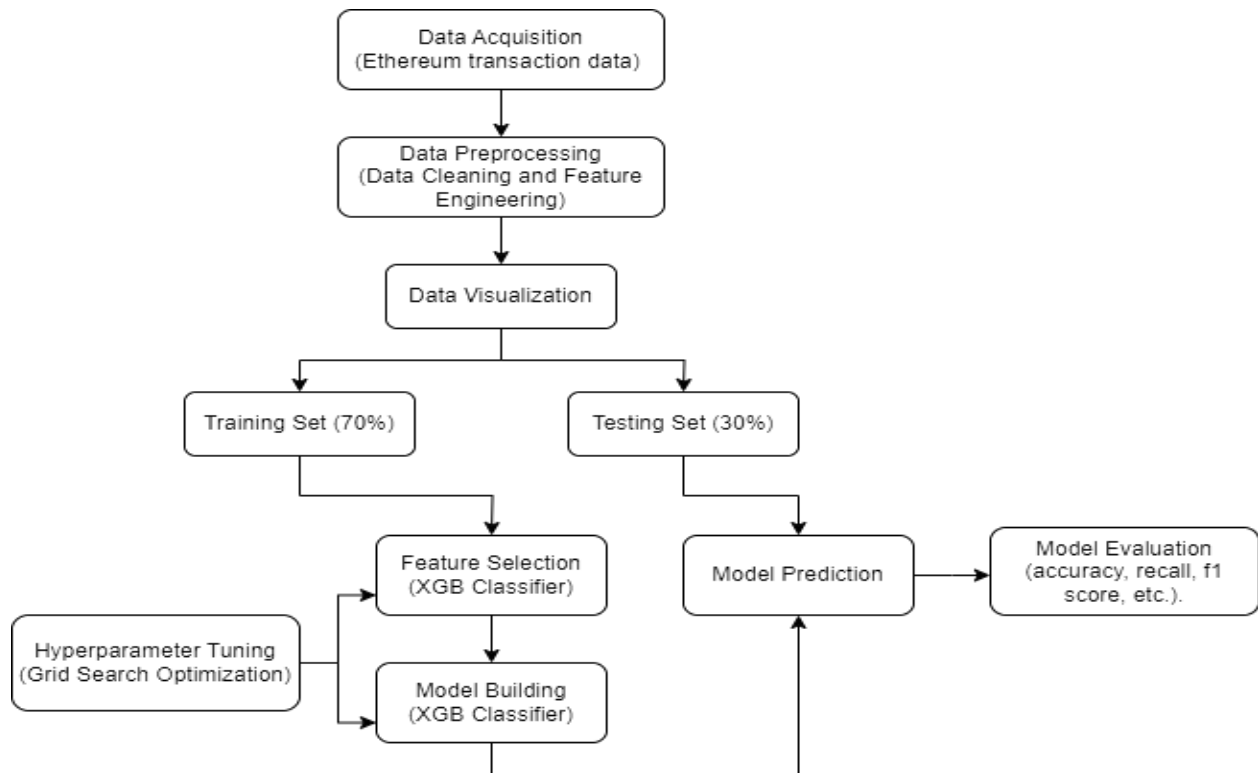


Figure 1. The flowchart of the proposed methodology.

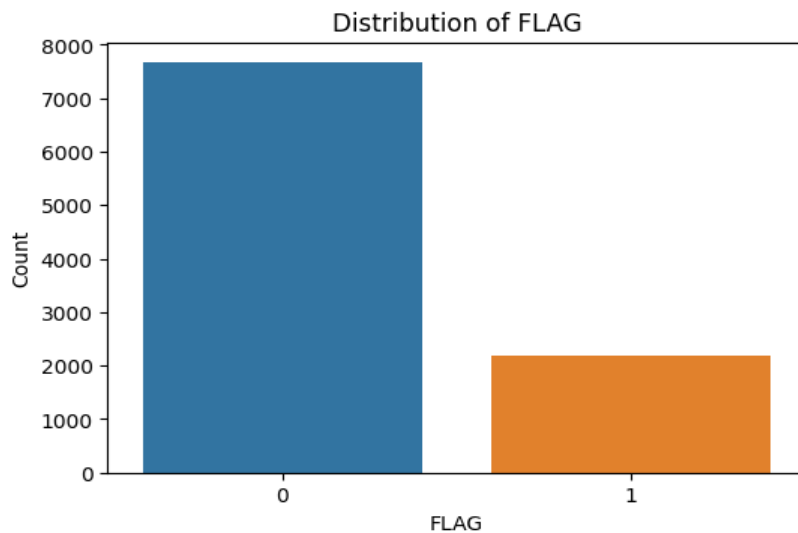


Figure 2. The distribution of the target variable.

3.4 | Model Building

When training and testing models for machine learning, this stage is crucial to achieving precise and excellent findings. Out of the 51 features, 17 are superfluous and would be removed from the data because they have no bearing on performance. After eliminating the null and duplicate values, 31 attributes are kept for additional examination. In this case, we suggested applying an oversampling method based on SMOTE algorithms to correct the imbalance in the data [21]. Following SMOTE analysis, it was discovered that each class had the same number of samples with the same ratio of 1:1.

The aforementioned phases are essential for constructing the suggested model in the subsequent phases:

- 30% of the dataset is used for testing, while the remaining 70% is for training.

- A classification model (XGB classifier) is trained on the training dataset. The study adjusted eleven hyperparameters to assess the XGB model's performance like a base estimator. The model results were achieved after numerous experiments along with tuning some hyperparameters which are:

Maximum Tree Depth=3 and Number of Trees=200.

Train the model and assess its performance to forecast the outcome for every combination of hyperparameters.

The following machine learning algorithms are recommended in this study:

- In particular, the XGB classifier—a scalable machine-learning method for tree boosting—was suggested by this study [22]. XGB is a member of the GB algorithm-based ensemble learning class. XGB focuses on minimizing the computing complexity of the best split, which in decision trees is a laborious procedure.
- Three different individual classifiers—LR, KNN, and SVM—were employed to categorize the data in the experiment's initial step. Accurate classification has also been achieved by using two different bagging and boosting ensemble classifiers: LGBM, and RF.

3.5 | Model Evaluation Metric

The assessment metrics used in this study are the confusion matrix, F1 score, accuracy, recall, specificity, precision, and log loss. The optimal classifier used in this study is determined by comparing these specific criteria.

- Confusion Matrix: To demarcate the relationship between the model's expected and actual values, this matrix is employed.

		Predicted Label	
		Negative or 0 or genuine	Positive or 1 or fake
Actual Label	Negative or 0 or genuine	TN (The scenario when the classifier correctly anticipated that the transaction was authentic, and it is.)	FP (The instance wherein the classifier identified the transaction as fraudulent while it is not.)
	Positive or 1 or fake	FN (The instance where the classifier assumed the transaction to be real when, in fact, it wasn't.)	TP (The case where the transaction is genuinely fraudulent, as the classifier had anticipated.)

- Accuracy: It is described as the appearance of accurate predictions for every sample that is being tested that is accessible. Additionally, it computes the proportion of cases that are effectively classified.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \tag{1}$$

- Sensitivity (Recall): the proportion of erroneous transactions to total actual positives or real positives.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{2}$$

- Precision: In addition to proving the definitive propriety, it ascertains affirmative class measures (i.e., unlawful transactions) that are being genuinely acknowledged by a certain classifier.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

- F1 Score: Based on recall and precision results, the F_β coefficient aids in evaluating the model's predictive efficacy.

$$F_{\beta} = \frac{(1+\beta^2)xy}{(\beta^2x)+y}, \text{ where } x \text{ is precision and } y \text{ is recall}$$

$$F1 = \frac{2xy}{x+y} \quad (4)$$

- Log Loss: is a "soft" accuracy metric that takes probabilistic certainty into account. Thus, the following is the binary classifier's log-loss:

$$\text{Log Loss} = -\frac{1}{n} \sum_1^n y_i \log b_i + (1 - E_i) \log(1 - b_i) \quad (5)$$

where b_i Shows how likely it is that the i th instance is a member of class 1. $E_i \in \{0,1\}$ which depicts the true label. In other words, it represents the cross-entropy between the predictions and the genuine label distribution. Consequently, a classifier's accuracy increases with decreasing log loss or cross-entropy values.

4 | Results and Discussion

Figure 3 shows the confusion matrix of the XGB classifier after model evaluation on the testing dataset.

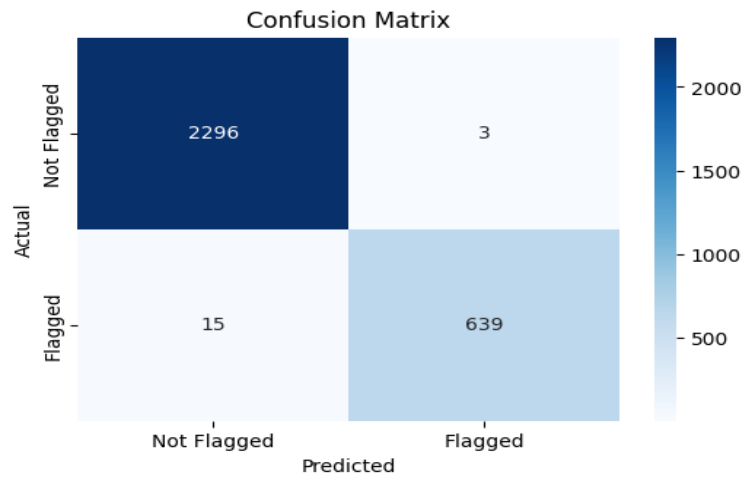


Figure 1. The confusion matrix of the XGB algorithm.

Six different classifiers i.e., KNN, LR, XGB, LGBM, RF, and SVM have been used in this work to identify such features to detect and classify the fraud account and to obtain more accurate results of detection due to their efficiencies and auto-learning abilities. With each classifier, extensive trials were conducted using the fraud detection dataset [20], which has over forty characteristics.

Each classifier's performance is assessed using the following metrics: F1 score, recall, accuracy, and precision, as shown in Table 1. As per Table 1's evaluation metrics, the XGB classifier has achieved 99.39% accuracy, making it the most effective detector for fraudulent accounts. Respectively, the LGBM also achieved a high result with 98.12% accuracy, while the RF obtained 97.51% accuracy. In conclusion, the KNN yielded a credible result with 96% accuracy, the LR had 88.52%, and the SVM scored 88%.

Table 1. Score comparison between the proposed XGB classifier and other commonly used models.

Model Name	Accuracy	F1 Score	Precision	Recall	FNR	FPR
SVM	88%	92%	97%	88%	0.9028	0.1164
LR	88.52%	76.70%	67.88%	88.15%	0.1185	0.1138
KNN	96%	97%	97%	96%	0.9502	0.0414
RF	97.51%	94.28%	92.87%	95.73%	0.0427	0.0200
LGBM	98.12%	95.64%	94.86%	96.44%	0.0356	0.0142
XGB	99.39%	98.61%	99.53%	97.71%	0.0229	0.0013

Table 2 analyzes the evaluation metrics of FP, TP, FN, and TN for the six suggested machine-learning models. Out of 654 flagged accounts, the XGB classifier has successfully identified 639 of them as fake, as seen in Figure 4. The suggested model discovered the following addresses as samples of fraudulent accounts:

0xe79392c79832287f9a07d0af9fa87fd150014e18

0x82603fec3241b319ac1a0470f2c08bd90461a355

0x1d64ea27764164debb4e891eb04d524f42904b08

Table 2. The contrasted analysis of the evaluation metrics.

Model Name	FN	FP	TP	TN
SVM	41	180	381	1367
LR	50	176	372	1371
KNN	21	64	401	1483
RF	18	31	404	1516
LGBM	15	22	407	1525
XGB	15	3	639	2296

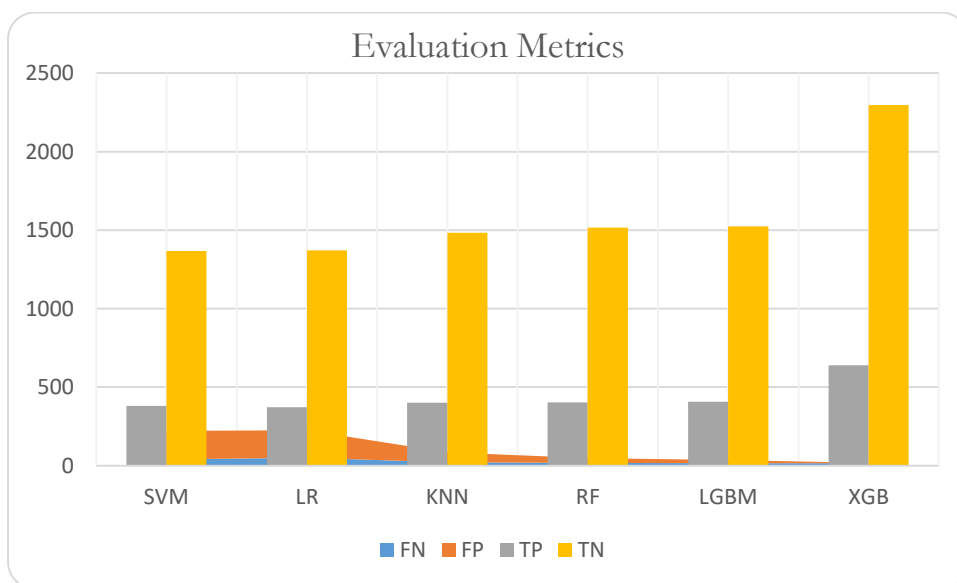


Figure 2. The comparative analysis of evaluation metrics.

The AUC results displayed in Figure 5 further suggest that the XGB classifier is capable of identifying fraudulent accounts on the Ethereum network. As illustrated in Figure 6, the proposed methodology that utilizes the XGB model demonstrates higher accuracy compared to the other studies [5, 7, 16-19]. This study indicates the importance of tailored feature selection and rigorous hyperparameter tuning in leveraging the XGB classifier effectively. While the proposed model's accuracy was 99.39%, the innovative model [7] attained a high accuracy of 99.7%; nonetheless, it may not perform well with unseen data. Therefore, it achieves the highest accuracy while training on data. The suggested model can verify and test new data to ensure model reliability in practical applications.

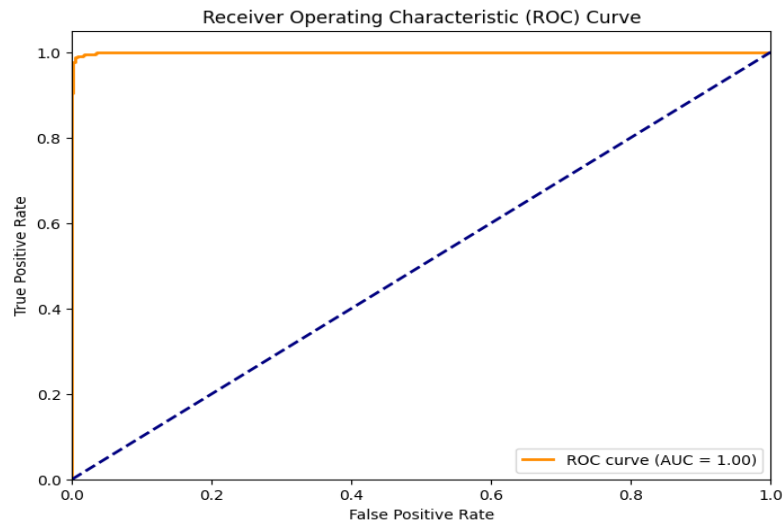


Figure 3. AUC performance of the XGB model.

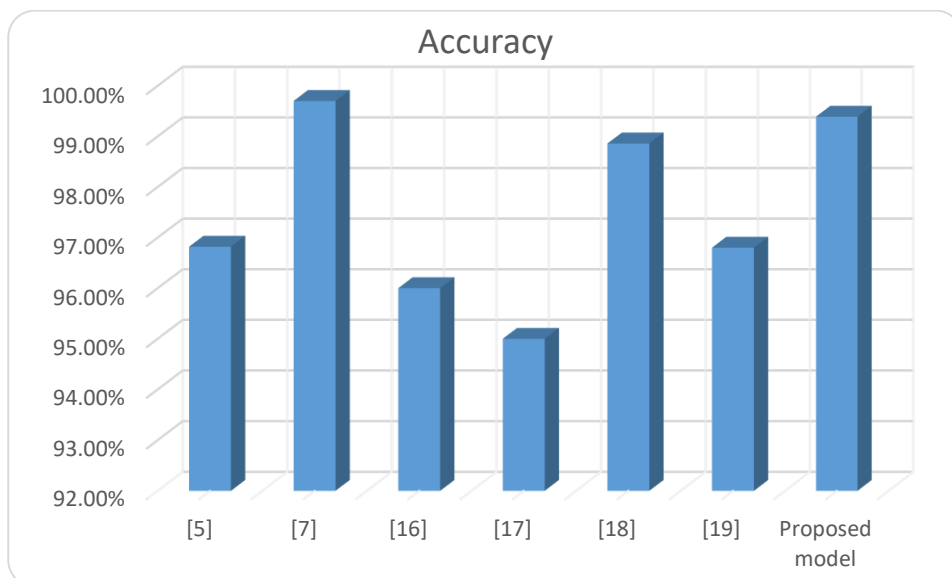


Figure 4. The contrastive analysis of the proposed XGB classifier accuracy with other recommended studies.

5 | Conclusions and Future work

This article employed a comprehensive technique together with a carefully selected dataset to focus on Ethereum fraud detection. The process involves data cleaning, exploratory analysis of data, and self-optimized machine-learning models. Concerning precisely spotting illicit activities and their accounts, XGB and LGBM in particular have shown to be incredibly efficient. Using oversampling strategies to address class imbalance enhances the effectiveness of the model without causing overfitting. Metrics that are consistent throughout the test and validation datasets verify this. The results of this investigation offer significant perspectives on the efficiency of ensemble models in identifying fraudulent activity in the Bitcoin space. Features significance analysis, assessment metrics, and model flexibility enable a thorough understanding of the Ethereum fraud detection process. The effectiveness of this study's execution will have a big impact on how secure and reliable the Ethereum ecosystem is.

Cohort analysis can be used in future research to examine the causes and consequences of fraudulent transactions in greater detail. Including counterfactual justifications can improve the comprehension of the model. This will render fraud detection systems more reliable and transparent.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Author Contributions

All authors contributed equally to this work.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] Pahuja, L. and Kamal, A. (2023) 'enlefd-dm: Ensemble learning based Ethereum Fraud Detection using crisp-dm framework', *Expert Systems*, 40(9). doi:10.1111/exsy.13379.
- [2] Jimmy, F. (2024) 'Enhancing data security in financial institutions with Blockchain technology', *Journal of Artificial Intelligence General science (JAIGS) ISSN:3006-4023*, 5(1), pp. 424–437. doi:10.60087/jaigs.v5i1.217.
- [3] Aziz, R.M. et al. (2022) 'LGBM: A machine learning approach for ethereum fraud detection', *International Journal of Information Technology*, 14(7), pp. 3321–3331. doi:10.1007/s41870-022-00864-6.
- [4] Aziz, R.M., Baluch, M.F., Patel, S. and Kumar, P. (2022) 'A machine learning based approach to detect the Ethereum fraud transactions with limited attributes', *Karbala International Journal of Modern Science*, 8(2), pp. 139–151. doi:10.33640/2405-609x.3229.
- [5] Kumar, N. et al. (2020) 'Detecting malicious accounts on the Ethereum blockchain with supervised learning', *Lecture Notes in Computer Science*, pp. 94–109. doi:10.1007/978-3-030-49785-9_7.
- [6] Aziz, R.M. et al. (2023) 'Modified genetic algorithm with deep learning for fraud transactions of Ethereum Smart Contract', *Applied Sciences*, 13(2), p. 697. doi:10.3390/app13020697.
- [7] Dutta, S., Sharma, A. and Rajgor, J. (2024a) 'Ethereum Fraud Prevention: A supervised learning approach for fraudulent account recognition', 2024 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST), abs/1603.02754, pp. 1–8. doi:10.1109/ictest60614.2024.10576142.
- [8] Hu, S. et al. (2023) 'Bert4eth: A pre-trained transformer for Ethereum Fraud Detection', *Proceedings of the ACM Web Conference 2023 [Preprint]*. doi:10.1145/3543507.3583345.
- [9] Wen, T. et al. (2023) 'A novel hybrid feature fusion model for detecting phishing scam on Ethereum using Deep Neural Network', *Expert Systems with Applications*, 211, p. 118463. doi:10.1016/j.eswa.2022.118463.
- [10] Chen, L. et al. (2020) 'Phishing scams detection in Ethereum Transaction Network', *ACM Transactions on Internet Technology*, 21(1), pp. 1–16. doi:10.1145/3398071.
- [11] Tan, R. et al. (2023) 'Ethereum fraud behavior detection based on Graph Neural Networks', *Computing*, 105(10), pp. 2143–2170. doi:10.1007/s00607-023-01177-7.
- [12] Sabry, F. et al. (2020) 'Cryptocurrencies and artificial intelligence: Challenges and opportunities', *IEEE Access*, 8, pp. 175840–175858. doi:10.1109/access.2020.3025211.
- [13] Motie, S. and Raahemi, B. (2024) 'Financial fraud detection using graph neural networks: A systematic review', *Expert Systems with Applications*, 240, p. 122156. doi:10.1016/j.eswa.2023.122156.

-
- [14] Ravindranath, V. et al. (2024) 'Evaluation of Performance Enhancement in Ethereum fraud detection using oversampling techniques', *Applied Soft Computing*, 161, p. 111698. doi:10.1016/j.asoc.2024.111698.
- [15] Zhang, Y. et al. (2021) 'Detecting Ethereum Ponzi Schemes Based on Improved LightGBM Algorithm,' *IEEE Transactions on Computational Social Systems*, 9(2), pp. 624–637. <https://doi.org/10.1109/tcss.2021.3088145>.
- [16] Rathore, M.M. et al. (2023) 'Detection of Fraudulent Entities in Ethereum Cryptocurrency: A Boosting-based Machine Learning Approach,' *GLOBECOM 2023 - 2023 IEEE Global Communications Conference* [Preprint]. <https://doi.org/10.1109/globecom54140.2023.10437184>.
- [17] Walavalkar, P. et al. (2024) 'A Token-based Approach to Detect Fraud in Ethereum Transactions,' *International Journal for Research in Applied Science and Engineering Technology*, 12(4), pp. 34–42. <https://doi.org/10.22214/ijraset.2024.59690>.
- [18] Crisostomo, J., Lobo, V. and Bacao, F. (2023) 'Detecting Fraudulent Wallets in Ethereum Blockchain Combining Supervised and Unsupervised Techniques - Using Autoencoders and XGBoost,' in *Lecture notes in networks and systems*, pp. 224–233. https://doi.org/10.1007/978-3-031-45155-3_23.
- [19] Sallam, A. et al. (2022) 'Fraudulent Account Detection in the Ethereum's Network Using Various Machine Learning Techniques,' *International Journal of Computer Systems & Software Engineering*, 8(2), pp. 43–50. <https://doi.org/10.15282/ijsecs.8.2.2022.5.0102>.
- [20] Aliyev, V. "Ethereum Fraud Detection Dataset," Kaggle, <https://www.kaggle.com/datasets/vagifa/ethereum-frauddetection-dataset> (accessed Jan. 26, 2024).
- [21] Li, J. et al. (2021) 'A novel oversampling technique for class-imbalanced learning based on SMOTE and natural neighbors,' *Information Sciences*, 565, pp. 438–455. <https://doi.org/10.1016/j.ins.2021.03.041>.
- [22] Demir, S. and Sahin, E.K. (2022) 'An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost,' *Neural Computing and Applications*, 35(4), pp. 3173–3190. <https://doi.org/10.1007/s00521-022-07856-4>.