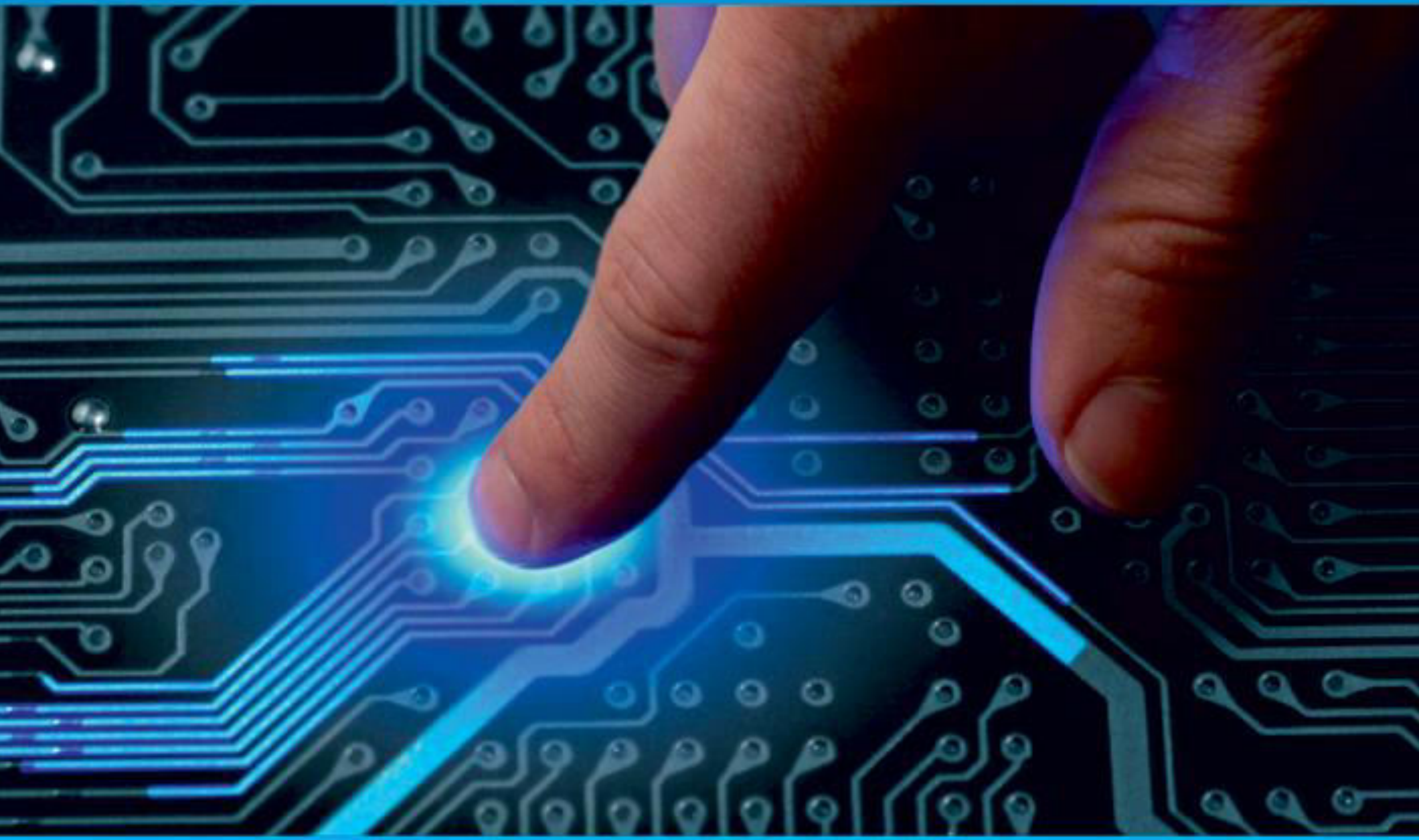
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204042 |

Efficient Fraud Detection Mechanism Based Ensemble Deep Learning Prediction Using Blockchain

R. Sujitha, N. Yogaraja,

Assistant professor, Department of Computer Science, Annapoorna Engineering College,

Tamil Nadu, India

Department of Computer Science, Annapoorna Engineering College, Tamil Nadu, India

ABSTRACT: In conclusion, this research article proposed a deep learning-based approach for detecting phishing attacks in blockchain transaction networks. The study used long short-term memory (LSTM), bi-directional LSTM (Bi-LSTM), and convolutional neural network LSTM (CNN-LSTM) to detect phishing attacks in real-time on blockchain networks. The proposed system integrated ensemble learning techniques, such as convolutional and recurrent neural networks, with blockchain and deep learning to address low precision, latency, stability, and privacy issues. The evaluation results demonstrate the proposed approach's effectiveness in detecting phishing attacks, which can contribute to developing more secure and trustworthy blockchain-based systems. The novelty of this study lies in the proposal and evaluation of a deep learning-based approach to detect phishing attacks in blockchain transaction networks, a new and innovative application of deep learning techniques. Future research can focus on developing more effective and knowledgeable deep learning algorithms to improve the system's performance. Additionally, an improved architecture that employs deep learning principles, such as feature extraction, scaling, and classification, can be suggested in a decentralized medium to address these concerns.

KEYWORDS: blockchain; network security; phishing; attack recognition; deep learning

I. INTRODUCTION

The survey delves into machine learning's application in identifying fraud in cryptocurrency transactions, particularly within Ethereum and Bitcoin. It examines the use of machine learning, deep learning, and blockchain to counter activities like Ponzi schemes and phishing. The survey emphasizes the significance of advanced technologies, supervised and unsupervised learning, and addresses challenges such as data imbalance and privacy concerns. It also suggests future research directions, emphasizing refined feature selection, unsupervised algorithms, and tailored technologies for cryptocurrency transactions. The study [1] addresses the rising concern of fraudulent activities on the Ethereum platform. It focuses on leveraging graph neural networks to extract features from users and transactions, classifying them as fraudulent or non-fraudulent. The study highlights challenges faced by investors due to a lack of understanding of smart contracts and prevalent fraudulent activities like phishing and smart Ponzi schemes. The study also emphasizes the superiority of graph neural networks over traditional models and suggests future research directions, including combining graph neural networks with explainable artificial intelligence and exploring custom architectures for improved accuracy and computational efficiency.

The analysis [2] explores the challenge of identifying unethical and fraudulent behavior within the cryptocurrency ecosystem, particularly in the Ethereum blockchain. It delves into the analysis of fraudulent behavior using different classification techniques and machine learning algorithms, focusing on the use of k-means clustering, Support Vector Machine, and random forest classifier to construct a transaction network based on Ethereum transactional data. The study reveals that the random forest classification model achieved the best performance in identifying fraudulent behavior. The advantages of the methods used in the document include high accuracy, flexibility, and the ability to extract valuable features for analysis. The study [3] addresses the pressing need for an effective approach to identify phishing scams on the Ethereum blockchain. The authors present a three-step framework involving obtaining labeled phishing accounts and their transaction records, constructing an Ethereum transaction network, utilizing node2vec network embedding for feature extraction, and applying a one-class support vector machine (SVM) for phishing

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204042 |

classification. The study underscores the unique challenges of phishing on blockchain platforms, advocates for leveraging publicly accessible Ethereum transaction records, and emphasizes the superiority of network embedding methods in automatically extracting latent features.

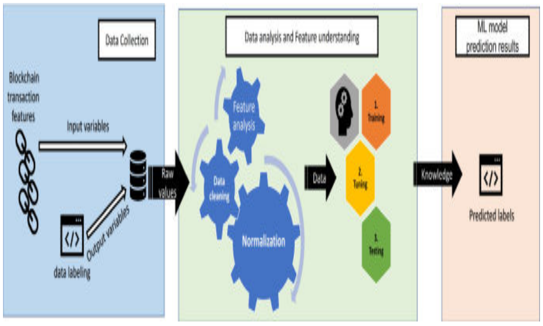


Fig 1: Block Diagram

The analysis [4] extensively explores the application of Light Gradient Boosting Machine (LGBM) for the detection of fraudulent activities within the Ethereum network. The study addresses potential threats such as Ponzi schemes, money laundering, and phishing, and proposes LGBM as an effective solution, showcasing its superior accuracy in comparison to other models like Random Forest and Multi-Layer Perceptron (MLP). The analysis also sheds light on the challenges of Ethereum fraud detection, suggesting future research directions such as refining feature selection methods and enhancing model accuracy and scalability. The study [5] delves into the application of unsupervised learning algorithms for anomaly detection in Bitcoin transactions. It evaluates various unsupervised learning algorithms, including Multivariate Gaussian distribution, One-Class SVM, Two-phase Clustering, and Isolation Forest, in the context of detecting anomalous behavior within cryptocurrency transactions. The study compares the performance of these algorithms through graphical representations and accuracy evaluations, ultimately highlighting the superiority of the Multivariate Gaussian algorithm. The analysis also includes a literature survey on related research works, showcasing diverse approaches that leverage unsupervised learning models for Bitcoin fraud detection and anomaly detection in blockchain electronic transactions. The advantages of unsupervised learning algorithms, such as their ability to handle large and complex datasets without labeled data, are outlined, along with considerations for their potential drawbacks, such as challenges in imbalanced datasets and interpretability issues. The study [6] offers a comprehensive analysis of cryptocurrencies within the context of the digital economy. It aims to assess the potential consequences of the further proliferation of cryptocurrencies in the global financial system and to identify strategies to address them. The authors delve into the historical evolution of monetary systems, from the gold standard to modern conditions, and examine the emergence and development of cryptocurrencies. They also discuss the potential use of cryptocurrencies in fraudulent schemes and evaluate the possibilities and challenges associated with the regulation of cryptographic money circulation. The document emphasizes the need to develop a legal and economic framework to regulate cryptocurrencies in the digital economy's development and highlights the opportunities and threats posed by virtual money from both economic and financial perspectives. The study also offers a critical examination of the regulatory landscape for cryptocurrencies, drawing attention to the varying approaches adopted by different countries, such as Japan, Germany, and the United States, in recognizing and regulating virtual currencies.

The analysis [7] addresses the pressing issue of identifying Ponzi schemes on the Ethereum blockchain to curb fraudulent activities causing substantial losses to investors. The study emphasizes the urgency in strengthening regulatory measures and monitoring within the blockchain market. The proposed CTRF model introduces a Code and Transaction Random Forest, leveraging features from smart contract code and transaction data to enhance the recall value for Ponzi contract identification. The document underscores the advantages of the model, including improved recall, effective data preprocessing techniques, and insightful feature analysis. However, it acknowledges challenges such as an imbalanced dataset and potential overfitting. Future research directions are suggested, including regulatory enhancements, exploration of deep learning algorithms, and addressing dataset limitations for improved generalization. The analysis [8] introduces a novel framework for detecting fraud in Bitcoin transactions, focusing on the efficiency of anomaly detection. The framework employs a stacking model with machine learning classifiers, combining Decision Tree, Naive Bayes, K-Nearest Neighbors, and Random Forest. Through ensemble learning and hyperparameter tuning using random search, the proposed model achieves impressive performance metrics, including a 97% accuracy, 98%

recall, and 97% F1-score, outperforming individual classifiers. The study systematically evaluates dataset aspects, balancing techniques, hyperparameter tuning, and model construction, emphasizing the effectiveness of the ensemble Bitcoin detector (EBD) model. The analysis [9] delves into the application of machine learning algorithms, specifically the Light Gradient Boosting Machine (LGBM), for the detection of fraudulent activities within the Ethereum network. Emphasizing the surging demand for cryptocurrencies like Ethereum, the study addresses potential threats such as Ponzi schemes, money laundering, and phishing.

II. LITERATURE SURVEY

Fraudulent Behaviour Identification in Ethereum Blockchain The research paper explores the challenge of identifying unethical and fraudulent behaviour within the cryptocurrency ecosystem, particularly in the Ethereum blockchain. The study aims to address the absence of regulation and transparency in transactions, which may lead to an increased number of fraudulent cases. The research delves into the analysis of fraudulent behaviour using different classification techniques and machine learning algorithms. It focuses on the use of k-means clustering, Support Vector Machine, and random forest classifier to construct a transaction network based on Ethereum transactional data. The results revealed that the random forest classification model achieved the best performance in identifying fraudulent behavior. The advantages of the methods used in the document include high accuracy, flexibility, and the ability to extract valuable features for analysis. However, the disadvantages include limitations in clustering algorithms, potential issues with false positives, and challenges related to data imbalance and network heterogeneity. The paper emphasizes the importance of data pre-processing and feature extraction in training and comparing the proposed models. The paper also outlines future plans to improve model reliability, such as increasing the number of fraudulent and non-fraudulent wallets for analysis and exploring the use of XG-Boost method. Additionally, the study intends to conduct a statistical significance test to ascertain differences between results and further enhance the proposed model's accuracy. Overall, the study provides a comprehensive overview of the research conducted to identify fraudulent behavior in the Ethereum blockchain, highlighting the significance of machine learning techniques and outlining future research directions.

Cryptocurrencies in the Global Financial System: Problems and Ways to Overcome them (2020) The research paper presents a comprehensive analysis of cryptocurrencies within the context of the digital economy. The study aims to assess the potential consequences of the further proliferation of cryptocurrencies in the global financial system and to identify strategies to address them. The authors delve into the historical evolution of monetary systems, from the gold standard to modern conditions, and examine the emergence and development of cryptocurrencies. They also discuss the potential use of cryptocurrencies in fraudulent schemes and evaluate the possibilities and challenges associated with the regulation of cryptographic money circulation. The document emphasizes the need to develop a legal and economic framework to regulate cryptocurrencies in the digital economy's development and highlights the opportunities and threats posed by virtual money from both economic and financial perspectives. Furthermore, the authors provide an insightful overview of the various types of fraud associated with cryptocurrencies, such as fake wallets, investment schemes, and phishing, underscoring the risks and challenges in the cryptocurrency market. The study also offers a critical examination of the regulatory landscape for cryptocurrencies, drawing attention to the varying approaches adopted by different countries, such as Japan, Germany, and the United States, in recognizing and regulating virtual currencies. Overall, the study provides a comprehensive analysis of the complexities surrounding cryptocurrencies, shedding light on their potential impact on the global financial system and the challenges posed by their unregulated circulation.

Detecting Phishing Scams on Ethereum Based on Transaction Records [3] The research paper addresses the pressing need for an effective approach to identify phishing scams on the Ethereum blockchain. The authors present a three-step framework involving obtaining labeled phishing accounts and their transaction records, constructing an Ethereum transaction network, utilizing node2vec network embedding for feature extraction, and applying a one-class support vector machine (SVM) for phishing classification. Experimental results reveal the model's effectiveness with an F-score of 0.846. The paper underscores the unique challenges of phishing on blockchain platforms, advocates for leveraging publicly accessible Ethereum transaction records, and emphasizes the superiority of network embedding methods in automatically extracting latent features. It discusses issues like data imbalance and network heterogeneity, introducing the one-class SVM as a solution. Furthermore, the paper delves into the significance of time and amount features, explores varying embedding dimensions, and suggests future research directions. In summary, the paper contributes a comprehensive framework for detecting Ethereum phishing scams, showcasing promising results and paving the way for advancements in blockchain security and fraud detection research.

III. METHODS

Phishing attacks are one such security concern observed in blockchain transactions. Phishing is a malicious technique attackers use to gain sensitive information, such as usernames, passwords, and private keys, from unsuspecting victims. Phishing attacks can be carried out via email, messaging apps, or social media platforms, leading to financial loss, identity theft, and other serious consequences. For example, fraudsters are taking advantage of flaws in major search engines' ad presentations to drive consumers to phishing sites. In contrast to a common strategy of sending emails to everyone, links to phishing sites in search engine advertising may be more convincing, especially if the domain they are targeting is listed as the destination. To guard against phishing, users' credentials are actively protected, which involves storing password hashes rather than passwords. Even though many firms spend large sums of money researching phishing and developing unique tools and algorithms for its detection and blocking, no products provide complete protection from such assaults. Because phishing attacks are typically carried out with the victim's involvement, anti-phishing defense uses technical and social engineering tools. Phishing detection has been intensively researched in recent decades, with several approaches presented. However, despite the properties of blockchain, there has been little study on phishing fraud detection. Andryukhin classifies the significant types and schemes of phishing attacks on the blockchain project and proposes ways of phishing attack defense from the blockchain project's perspective. Unlike them, we are focusing overall blockchain ecosystem and alerting users to phishing schemes as soon as they appear. AI and deep learning methods can be applied to detect network intrusions and recognize malware and botnet attacks. To improve recognition accuracy, ensemble learning architectures have been successfully adopted.

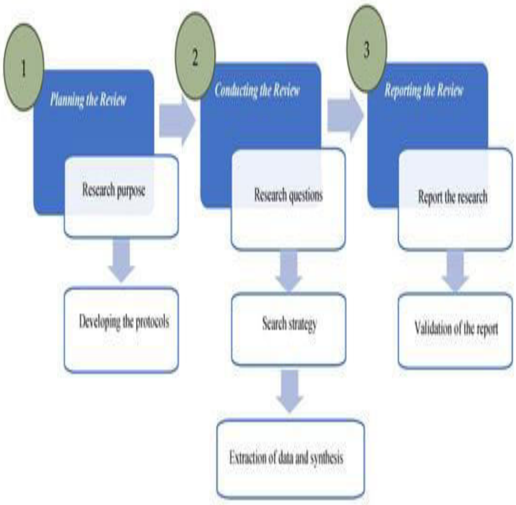


Fig 2: Work Flow

The motivation for this paper is to propose a deep-learning-based approach to detect phishing attacks in blockchain transaction networks. The use of deep learning methods, such as long short-term memory (LSTM), bi-directional LSTM (Bi-LSTM), and convolutional neural network LSTM (CNN-LSTM), can help to identify and prevent phishing attacks in real time. Deep learning is increasingly recognized as a powerful tool for advancing big data technologies, including those related to the internet of things (IoT). However, developing effective deep learning models for IoT applications requires addressing several critical issues, such as single points of vulnerability, privacy leakage, lack of usable knowledge, and data poisoning attacks. In addition, phishing scams are a severe threat to the financial security of users in blockchain ecosystems. To address this issue, this paper proposes a systematic approach to detecting phishing scams in the Ethereum ecosystem using an improved ensemble-based LSTM architecture combining deep learning and blockchain technology. The proposed system aims to provide possible solutions for mitigating phishing scams in IoT networks by integrating ensemble learning techniques, such as convolutional and recurrent neural networks, with blockchain and deep learning.

As input, we use Blockchain transaction addresses which are preprocessed and transformed into the numerical domain using word embedding. The classification uses three types of neural network models (LSTM, BI-LSTM, and CNN-LSTM). We have selected Bi-LSTM over other alternatives such as GRU (Gated Recurrent Unit) because Bi-LSTM is

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204042 |

generally considered better than GRU in tasks where modeling long-term dependencies is important. This is because Bi-LSTM has a more complex architecture that allows it to remember information from earlier in the sequence for longer periods. Additionally, the bidirectional aspect of Bi-LSTM allows it to capture dependencies in both directions, which can be especially useful in tasks where the ordering of the data is important. Finally, the results are aggregated using ensemble voting to produce the final prediction (malicious or benign transaction)

IV. RESULT ANALYSIS

Word embedding refers to representing words in a high-dimensional vector space, where each word is assigned a unique vector that captures its semantic and syntactic properties. The main idea behind word embedding is that similar words should have identical vector representations and that the distance between two vectors can be used to measure the similarity between their corresponding words. Word embedding typically starts with a neural network used to learn the vector representations of each word in the training corpus. The neural network is trained to predict the context in which each word appears based on its neighboring words. The resulting vector representations capture the meaning of each word based on the context in which it is used and can be used as input to many natural languages processing tasks, such as sentiment analysis, machine translation, and information retrieval.

Here, we used word embedding with an external neural network to train to connect words with their context, yielding a set of numeric vectors with the necessary dimensions.

CNN may utilize max-pooling layers. Max pooling performs nonlinear down-sampling that minimizes the longitudinal size of the conventional layer's output using the production subdivision in rectangular boxes, which is the best value for each filter. By bringing down the quantity of connection and computational cost, max-pooling lessens overfitting by making features more spatially accessible. Max pooling layers can be numbered after every convolutional layer or after some layers.

Using dropout, during the training process, randomly selected neurons are ignored. The operation means that we "drop out" some neurons randomly. In easy words, the contribution of neurons is temporarily removed on the forward pass during the activation of downstream neurons, and no weight updates are done to neurons on the backward pass.

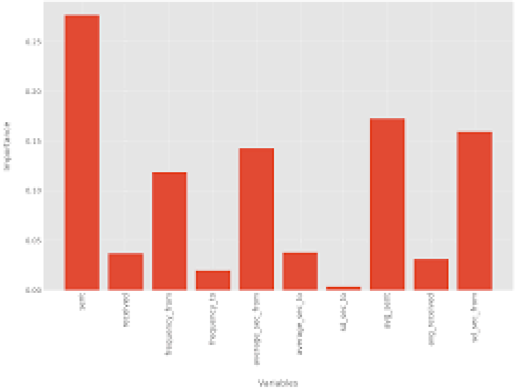


Fig 3: Fraudulent Behaviour Identification in Ethereum Blockchain Result Analysis

Ensemble learning is a technique used in machine learning that involves training multiple imperfect models, also known as "weak learners," to perform the same task. Then, their results are combined to achieve a better outcome. These weak learners are then used to build more complex models by combining multiple them. Typically, these base models do not perform independently due to either bias or variability, making them less robust. Ensemble techniques address this issue by merging multiple models to create a strong learner with improved performance. More accurate and reliable models can be generated by adequately combining weak models. A stacking ensemble model is designed using base models and a meta-learner, the final stage classifier that utilizes the predictions made by the base models (see Algorithm 1). The base models are trained on the training data and used to generate predictions, which are then used as input for the meta-learner. The meta-learner is trained on the base model predictions using new, unseen data, to aggregate the base model predictions and make the correct output prediction. This is achieved by feeding the meta-learner with input and output pairs of data from the base learners. Ensemble classification models are potent machine learning methods that have the potential to perform exceptionally well and generalize well to novel, new datasets. An ensemble classifier's benefit is that combining the predictions of several classifiers may correct for mistakes produced by any of them, improving total accuracy.

V. CONCLUSION

The papers provide a comprehensive and in-depth analysis of fraudulent activities within the cryptocurrency ecosystem, with a particular focus on the Ethereum blockchain. It encompasses various research papers that explore the identification, analysis, and impact of Ponzi schemes, phishing scams, and other fraudulent behaviors. The studies employ a range of methodologies, including machine learning algorithms, graph neural networks, and ensemble learning, to detect and classify fraudulent transactions and smart contracts. Additionally, the papers address the challenges and opportunities associated with blockchain technology, the potential consequences of the proliferation of cryptocurrencies in the global financial system, and the need for regulatory measures to combat fraudulent activities. Overall, the papers offer valuable insights into the complexities of cryptocurrency scams, the potential impact on the financial system, and the ongoing efforts to enhance security and fraud detection in blockchain technology. The document serves as a valuable resource for researchers, practitioners, and policymakers in the evolving field of cryptocurrency fraud mitigation.

REFERENCES

[1]. G. Luchkin, O. L. Lukasheva, N. E. Novikova, V. A. Melnikov,,A. V. Zyatkova, and E. V. Yarotskaya, ‘‘Cryptocurrencies in the global financial system: Problems and ways to overcome them,’’ in Proc. Russian Conf. Digit. Economy Knowl. Manag. (RuDEcK), 2020.

[2]. K. Lašas, G. Kasputyté, R. Užupyté, and T. Krilavičius, ‘‘Fraudulent behaviour identification in Ethereum blockchain,’’ in Proc. CEUR Workshop, Inf. Soc. Univ. Stud., Kaunas, Lithuania, 23, Apr. 2020.

[3]. Q. Yuan, B. Huang, J. Zhang, J. Wu, H. Zhang, and X. Zhang, ‘‘Detecting phishing scams on Ethereum based on transaction records,’’ in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), Oct. 2020.

[4]. R. M. Aziz, M. F. Baluch, S. Patel, and A. H. Ganie, ‘‘LGBM: A machine learning approach for Ethereum fraud detection,’’ Int. J. Inf. Technol.,vol. 14, no. 7, pp. 3321–3331, Dec. 2022, doi: 10.1007/s41870-022- 00864-6

[5]. G. D. Arya, K. V. S. Harika, D. V. Rahul, S. Narasimhan, and A. Ashok, ‘‘Analysis of unsupervised learning algorithms for anomaly mining with Bitcoin,’’ in Machine Intelligence and Smart Systems. Berlin, Germany: Springer, 2021.

[6]. Mwanza, Charity, "Graph neural networks for ethereum fraud detection" (2023). Theses. 449, https://louis.uah.edu/uah-theses/449

[7]. Xuezhi He , Tan Yang , and Liping Chen "CTRF: Ethereum-Based Ponzi Contract Identification", Hindawi,Security and Communication Networks,Volume 2022, Article ID 1554752, https://doi.org/10.1155/2022/1554752

[8]. M. Bhowmik, T. S. S. Chandana, and B. Rudra, ‘‘Comparative study of machine learning algorithms for fraud detection in blockchain,’’ in Proc. 5th Int. Conf. Comput. Methodologies Commun. (ICCMC), Apr. 2021.

[9]. W. Chen, Z. Zheng, E. C.-H. Ngai, P. Zheng, and Y. Zhou, ‘‘Exploiting blockchain data to detect smart Ponzi schemes on Ethereum,’’ IEEE Access, vol. 7, pp. 37575–37586, 2019.

[10]. Ross Phillips and Heidi Wilder, "Tracing Cryptocurrency Scams: Clustering Replicated Advance-Fee and Phishing Websites",2020,IEEE

[11]. Ogundokun, R.O., Arowolo, M.O., Damaševičius, R. and Misra, S., 2023, May. Phishing Detection in Blockchain Transaction Networks Using Ensemble Learning. In Telecom. Ashfaq, T, Khalid, R.Yahaya, A.S.; Aslam, S.; Azar, A.T.;

[12]. Alsafari, S, Hameed, I.A. "A Machine Learning and Blockchain Based Efficient Fraud Detection Mechanism.", Sensors 2022, 22, 7162. https://doi.org/10.3390/s22197162

[13]. Bartoletti, Massimo & Carta, Salvatore &Cimoli, Tiziana & Saia, Roberto. (2017). Dissecting Ponzi schemes on Ethereum: identification, analysis, and impact.

[14]. Bartoletti, Massimo & Lande, Stefano &Loddo, Andrea &Pompianu, Livio &Serusi, Sergio. (2021). Cryptocurrency Scams: Analysis and Perspectives. IEEE Access. 9. 1-1. 10.1109/ACCESS.2021.3123894.

[15]. Nayyer, Noor & Javaid, Nadeem & Akbar, Mariam &Aldegheishem, Abdulaziz &Alrajeh, Nabil & Jamil, Mohsin. (2023). A New Framework for Fraud Detection in Bitcoin Transactions Through Ensemble Stacking Model in Smart Cities. 10.1109/ACCESS.2023.3308298.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

☎ 9940 572 462 6381 907 438 ✉ ijircce@gmail.com

Scan to save the contact details

www.ijircce.com