IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⧖ Impact Factor 7.39 ⧖ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11759

COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR FRAUD DETECTION IN BLOCKCHAIN

Ranjitha .H[1], Prof.M.N.Chandan[2]

PG Scholar, Dept. of MCA, PES College of Engineering, Mandya, Karnataka, India[1]

Assistant Professor, Dept. of MCA, PES College of Engineering, Mandya, Karnataka, India[2]

Abstract: Long-term research has been done on the issue of spotting fake exchanges. The existence of fraudulent exchanges in the economy discourages investors from investing in bitcoin and other blockchain-based businesses. False exchanges are regularly viewed with scepticism due to the gatherings in issue or the way they are put up. To prevent them from jeopardising the trustworthiness of the neighbourhood and the blockchain network, people endeavour to identify false exchanges wherever possible. Numerous other Machine Learning approaches have been suggested to address this issue, but none of them has clearly emerged as the best one, even though some of the results show promise. This study looks at how well a few controlled AI models and a few deep learning models do at spotting bogus transactions in a blockchain network. Such a correlation exploration will assist in identifying the most efficient method given the compromise in precision and processing execution. Our goal is to pinpoint the clients and transactions that will probably resort to extortion.

A blockchain network's economics and user confidence are fundamentally damaged by fraudulent exchanges. Although it is hard to guarantee the morality of the participants or the verifiers, employing agreement methods like proof of stake or proof of work makes it possible to confirm the authenticity of an exchange. This implies that fraud in a block chain organisation is still a possible. One method to stop extortion is by using AI computations. AI facilitates both guided and independent learning. For both certified and fraudulent exchanges, we analyse managed AI solutions in this study. Additionally, we offer a full connection of several directed AI methodologies.

INTRODUCTION

Blockchain uses end-to-end encryption to produce a changeless record of exchanges, eliminating fraud and other illegal activity. Information is stored on the blockchain using a network of PCs, making it virtually impossible to hack (in contrast to ordinary PC frameworks that stores information together in servers). Additionally, blockchain can more easily address security concerns than conventional PC frameworks by encrypting information and requiring consents to limit access.

Blockchain transactions are constant because once they are recorded, they cannot be changed or reversed. Before a "block" of transactions is added to the blockchain, network clients must agree on the validity of each transaction. Fraudsters employ a variety of techniques to hide their illegal activities, including the production of fictitious records, the alteration of physical or electronic records, and the manipulation of data in an association's bookkeeping frameworks. Using a shared electronic record can help reduce extortion because it increases the openness and clarity of communications between members of a company organisation and within a production network. False trades are easier to spot since groups can track the evolution of resources and experiences.

Furthermore, to alter the transaction records on a blockchain, a single person or group would need to gain control over a larger portion of the system. In any event, not all blockchains are accessible. However, permissioned groups can be quite effective in eliminating misrepresentation because they restrict who is permitted to participate and to what extent. People from a permissioned network should be welcomed and approved before they can contribute. Access control and executive personalities are crucial in a permissioned network. Extortion is exceedingly challenging since participants in the Hyperledger Fabric blockchain execution configuration are permitted cryptographic enrollment cards that match their characters.

Objectives

The objective of the project is to create sophisticated machine learning models for analysing and predicting the data set utilising typical machine learning methods, statistics, and calculus to forecast the frequency and volume of fraudulent and legitimate transactions.

DOI: 10.17148/IJARCCE.2022.11759

LITERATURE SURVEY

Cai, Y., Zhu, D. Fraud detections for online businesses: a perspectivefrom blockchain technology.
Frauds that were both objective and subjective were discussed by Yuanfeng Cai and associates. They conclude that while blockchain is good at recognising objective fraud, it is useless at spotting subjective fraud, thus they use machine learning to remedy the weakness. [1]

Xu, J.J. Are blockchains immune to all malicious attacks
Jennifer J. Xu focused on the types of fraud that blockchain is capable of identifying as well as those that it is still vulnerable to. Thus, ideas about the considerations that a machine learning component needed to make started to develop. She makes it quite clear that while blockchain merely applies a set of predetermined rules, risks like data fraud and framework hacking continue to be attainable and difficult to stop. [2]

Ostapowicz M., ˙ Zbikowski K. (2019) Detecting Fraudulent Accounts on Blockchain: A Supervised Approach
To detect fraud, Micha Ostapowicz et al. used supervised machine learning algorithms. The capacity of crooks to steal money using well-known software or bogus messages was their main focus. They then used the abilities of Random Forests, Support Vector Machines, and XGBoost classifiers to identify such records using a dataset of more than 300,000 records. [3]

Podgorelec, B., Turkanovi´c, M. and Karakatiˇc, S., 2020. A Machine Learning-Based Method for Automated Blockchain Transaction Signing Including Personalized Anomaly Detection.
A method for automatically certifying blockchain transactions with machine learning was developed by Blaz Podgorelec et al. As a result, it also employs a particular method to identify fraudulent transactions. [4]

Farrugia S, Ellul J, Azzopardi G. Detection of illicit accounts over the Ethereum blockchain. Expert Systems with Applications
These accounts' transaction histories allowed the author to identify phoney accounts on the Blockchain. They found that the three crucial indicators that help identify fraudulent accounts are "Time between first and last (Mins)," "Total Ether balance," and "Min value received." [5]

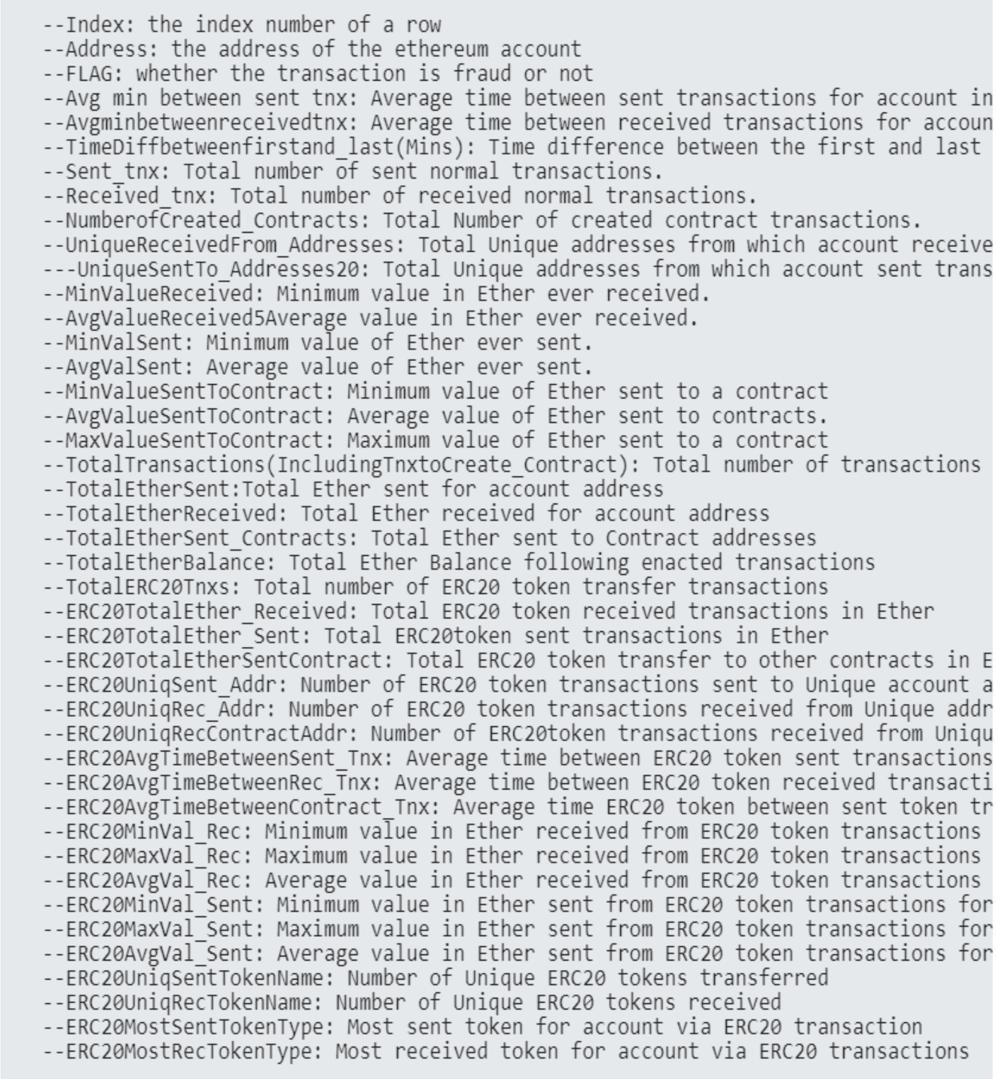
Anomaly detection in bitcoin network using unsupervised learning methods
In order to identify an anomaly in bitcoin transaction networks, Thai T. Pham et al. specifically targeted it. Kmeans clustering, Mahalanobis distance, and unsupervised support vector machines were used to identify suspect users and transactions. They made use of a set of information that was divided into two graphs, one with nodes for users and the other with nodes for transactions. [6]

A Survey on Blockchain Anomaly Detection UsingData Mining Techniques.
Deep learning methods for data mining anomaly detection are fully analysed by Li Ji et al. They also provided concise summaries of the main general and targeted detection approaches. Discussed along with predicted advancements in this field are the advantages and disadvantages of the various techniques currently in use. [7]

Economic analysis of cryptocurrency backed money laundering
By Christian Brenig et al., a money laundering scheme based on cryptocurrencies is examined economically. They cover the fundamentals of money laundering and provide ways to avoid it. This research provides us with even more inspiration to find out about such fraudulent activities involving cryptocurrency and blockchain technology in general. [8]

DATA SET

This collection contains rows of Ethereum transactions, both honest and fraudulent. Consider the unevenness of the dataset during modelling. Ethereum is a software platform that distributes copies of smart contracts to thousands of users worldwide in order to decentralise data. With Ethereum serving as a means of exchange, value may be transferred anywhere in the world without interference from outside parties. The main threat to trade security, nevertheless, comes from a variety of fraudulent practises, such as bribery, money laundering, and phishing. This work proposes the Light Gradient Boosting Machine technique for efficiently detecting fraudulent transactions. In order to categorise the Ethereum fraud detection dataset with constrained features, it also looks at the metrics of several models, including Random Forest, Multi-Layer Perceptron, and others, based on machine learning and soft computing techniques.



```
--Index: the index number of a row
--Address: the address of the ethereum account
--FLAG: whether the transaction is fraud or not
--Avg min between sent tnx: Average time between sent transactions for account in
--Avgminbetweenreceivedtnx: Average time between received transactions for accoun
--TimeDiffbetweenfirstand_last(Mins): Time difference between the first and last
--Sent_tnx: Total number of sent normal transactions.
--Received_tnx: Total number of received normal transactions.
--NumberofCreated_Contracts: Total Number of created contract transactions.
--UniqueReceivedFrom_Addresses: Total Unique addresses from which account receive
---UniqueSentTo_Addresses20: Total Unique addresses from which account sent trans
--MinValueReceived: Minimum value in Ether ever received.
--AvgValueReceived5Average value in Ether ever received.
--MinValSent: Minimum value of Ether ever sent.
--AvgValSent: Average value of Ether ever sent.
--MinValueSentToContract: Minimum value of Ether sent to a contract
--AvgValueSentToContract: Average value of Ether sent to contracts.
--MaxValueSentToContract: Maximum value of Ether sent to a contract
--TotalTransactions(IncludingTnxtoCreate_Contract): Total number of transactions
--TotalEtherSent:Total Ether sent for account address
--TotalEtherReceived: Total Ether received for account address
--TotalEtherSent_Contracts: Total Ether sent to Contract addresses
--TotalEtherBalance: Total Ether Balance following enacted transactions
--TotalERC20Tnxs: Total number of ERC20 token transfer transactions
--ERC20TotalEther_Received: Total ERC20 token received transactions in Ether
--ERC20TotalEther_Sent: Total ERC20token sent transactions in Ether
--ERC20TotalEtherSentContract: Total ERC20 token transfer to other contracts in E
--ERC20UniqSent_Addr: Number of ERC20 token transactions sent to Unique account a
--ERC20UniqRec_Addr: Number of ERC20 token transactions received from Unique addr
--ERC20UniqRecContractAddr: Number of ERC20token transactions received from Uniqu
--ERC20AvgTimeBetweenSent_Tnx: Average time between ERC20 token sent transactions
--ERC20AvgTimeBetweenRec_Tnx: Average time between ERC20 token received transacti
--ERC20AvgTimeBetweenContract_Tnx: Average time ERC20 token between sent token tr
--ERC20MinVal_Rec: Minimum value in Ether received from ERC20 token transactions
--ERC20MaxVal_Rec: Maximum value in Ether received from ERC20 token transactions
--ERC20AvgVal_Rec: Average value in Ether received from ERC20 token transactions
--ERC20MinVal_Sent: Minimum value in Ether sent from ERC20 token transactions for
--ERC20MaxVal_Sent: Maximum value in Ether sent from ERC20 token transactions for
--ERC20AvgVal_Sent: Average value in Ether sent from ERC20 token transactions for
--ERC20UniqSentTokenName: Number of Unique ERC20 tokens transferred
--ERC20UniqRecTokenName: Number of Unique ERC20 tokens received
--ERC20MostSentTokenType: Most sent token for account via ERC20 transaction
--ERC20MostRecTokenType: Most received token for account via ERC20 transactions
```

BACKGROUND STUDY

Cross-line exchanges, which may result in double taxation in both the exchanged and receiving countries, receive the majority of their attention. The various expense forms of extortion are examined as well. This brought up some important points for an AI component to think about. Although blockchain just maintains a predetermined set of rules, she claims that assaults like data fraud and system hacking are still feasible and difficult to detect when using blockchain.

METHODOLOGY

When the Blockchain network accepts a trade after passing all first checks, our suggested arrangement basically jumps straight into it and performs additional checks to see whether the exchange may be fake. As a result of the Blockchain organization's ability to instantly discredit exchanges without putting in further effort, this technique eliminates the requirement to evaluate exchanges.

Our suggested arrangement enters the image and makes additional tests to determine if the exchange might be false after the Block Chain network has approved a transaction after each significant check. This approach ensures that there isn't even a small additional cost for trade confirmation because of how straightforward nullification is within the blockchain organisation.

DOI: 10.17148/IJARCCE.2022.11759

Three phases can largely be used to categorise the work:

1. Pre-handling stage

We preprocess using network node embedding and the node2vec method. The combined ratings dataset is then read to produce a data frame.

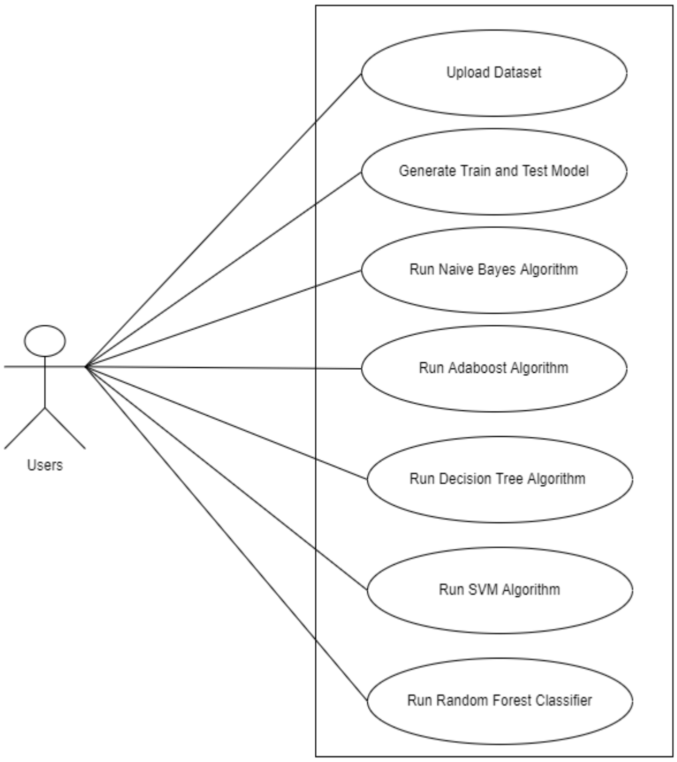
The node2vec method's outputs are then normalised, and the normalised values are then saved in a file. When a transaction is discovered to be fraudulent, we assign it a score of 1, and when it is not, we assign it a score of 0. Then the mean and SD of the node features are calculated, and the outcomes are saved to a CSV file. Next, train and test sets of the gathered data were created.

2. Building and preparing different models

Test (0.2) and train (0.8) data were included in our analysis. Then, in our train and test sets, we evaluate the ratio of honest to dishonest transactions. We use machine learning and deep learning techniques to predict the likelihood that a transaction will be successful.

3. Performance assessment of the relative multitude of models

We evaluate each of our classification models. In order to estimate a parameter, sampling in machine learning involves selecting a sample of data from the dataset with replacement. So, we start by choosing the bootstrap sample size. The model's efficacy is next evaluated using the mean of all accuracy values obtained in this way, after which the sample size is determined


Upload Dataset
Generate Train and Test Model
Run Naive Bayes Algorithm
Run Adaboost Algorithm
Run Decision Tree Algorithm
Run SVM Algorithm
Run Random Forest Classifier
Users

CONCLUSION

For a block chain organisation, it has been suggested using AI to identify false exchanges. The administered learning methods that were used in this cycle were support vector machines, choice trees, and strategic relapse. A comprehensive assessment of all methodologies is taken into account for accuracy. To discriminate between unaided methods like grouping, this exploration might be used. We also intend to lead an in-depth investigation of criminal activity on a private block chain in the future.

A detailed focus on phoney exercises in a private blockchain is planned for later study. There will also be a thorough focus on phoney activities in a private blockchain. The examination involved is too extensive for us to complete on our own, despite the fact that it is something we are very motivated by. As a result, to assist us, we have gotten in touch with a group of experts that are experts in this industry.

DOI: 10.17148/IJARCCE.2022.11759

BIBLIOGRAPHY

1. Cai, Y., Zhu, D. Fraud detections for online businesses: a perspectivefrom blockchain technology.
2. Xu, J.J. Are blockchains immune to all malicious attacks
3. Ostapowicz M., ˙ Zbikowski K. (2019) Detecting Fraudulent Accounts on Blockchain: A Supervised Approach
4. Podgorelec, B., Turkanovi´c, M. and Karakatiˇc, S., 2020. A Machine Learning-Based Method for Automated Blockchain Transaction Signing Including Personalized Anomaly Detection.
5. Farrugia S, Ellul J, Azzopardi G. Detection of illicit accounts over the Ethereum blockchain. Expert Systems with Applications
6. Pham, Thai, and Steven Lee. ”Anomaly detection in bitcoin network using unsupervised learning methods
7. Li, Ji, et al. ”A Survey on Blockchain Anomaly Detection UsingData Mining Techniques.
8. Brenig, Christian, and G¨unter M¨uller. ”Economic analysis of cryptocurrency backed money laundering.” (2015)
9. Monamo, Patrick, Vukosi Marivate, and Bheki Twala. ”Unsupervised learning for robust Bitcoin fraud detection.” 2016 Information Security for South Africa (ISSA). IEEE, 2016.
10. Anomaly Detection in Bitcoin Network Using Unsupervised Learning Methods
11. Data mining for credit card fraud: A comparative study S. Bhattacharyya, Sanjeev Jha, K. Tharakunnel, J. Westland
12. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature E. Ngai, Yong Hu, Y. Wong, Yijun Chen, Xin Sun
13. Graph based anomaly detection and description: a survey L. Akoglu, Hanghang Tong, Danai Koutra