ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 13, Issue 4, 2019

COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR FRAUD DETECTION IN BLOCK CHAIN

[1]Dr.A. Avani,[2]T. Bharath Krishna

[1]Associate Professor, Department of Computer Science and Engineering

Anubose Institute of Technology, New Palvoncha-507115,

Bhadradri Kothagudem-Dist-TG

[2]Department of Computer Science and Engineering

Anubose Institute of Technology, New Palvoncha-507115,

Bhadradri Kothagudem-Dist-TG

ABSTRACT

Fraudulent transactions have a huge impact on the economy and trust of a blockchain network. Consensus algorithms like proof of work or proof of stake can verify the validity of the transaction but not the nature of the users involved in the transactions or those who verify the transactions. This makes a blockchain network still vulnerable to fraudulent activities. One of the ways to eliminate fraud is by using machine learning techniques. Machine learning can be of supervised or unsupervised nature. In this paper, we use various supervised machine learning techniques to check for fraudulent and legitimate transactions. We also provide an extensive comparative study of various supervised machine learning techniques like decision trees, Naive Bayes, logistic regression, multilayer perceptron, and so on for the above task .

I.INTRODUCTION

The problem of detecting fraudulent transactions is being studied for a long time. Fraudulent transactions are harmful to the economy and discourage people from investing in bitcoins or even trusting other blockchain-based solutions. Fraudulent transactions are usually suspicious either in terms of participants involved in the transaction or the nature of the transaction. Members of a blockchain network want to detect Fraudulent transactions as soon as possible to prevent them from harming the blockchain network's community and integrity. Many Machine Learning techniques have been proposed to deal with this problem, some results appear to be quite promising , but there is no obvious superior method. This paper compares the performance of various supervised machine learning models like SVM, Decision Tree, Naive Bayes, Logistic Regression, and few deep

ISSN 2454-9940
www.ijasem.org
Vol 13, Issue 4, 2019

learning models in detecting fraudulent transactions in a blockchain network. Such comparative study will help decide the best algorithm based on accuracy and computational speed trade-off. Our goal is to see which users and transactions have the highest probability of being involved in fraudulent transactions.

II.LITERATURE SURVEY

1.Yuanfeng Cai et al. discussed the objective and subjective frauds. They conclude that blockchain effectively detects objective fraud but not subjective fraud and thus uses Machine Learning to mitigate the weakness. Jennifer J. Xu discussed the types of fraudulent activities that blockchain can detect and the ones that blockchain is still vulnerable to. This paved a path towards ideas about what problems a Machine learning part needs to consider. She specifies that attacks like Identity theft and system hacking are still possible and challenging to detect using blockchain as it just uses some predetermined rules. Michał Ostapowicz et al. used Supervised Machine Learning methods to detect fraudulent activities. They focused on the fact that malicious actors can steal money by applying well-known malware software or fake emails. Therefore they used the capabilities of Random Forests, Support Vector Machines, and XGBoost classifiers to identify such accounts based on a dataset of more than 300 thousand accounts. Bla Podgorelec et al. devised a method using Machine Learning for the automated signing of transactions in the blockchain. Hence, it also uses a personalized identification of anomalous transactions.

Steven Farrugia et al. detected illicit accounts in the Ethereum Blockchain based on heir transaction history. They found out that ‘Time difference between first and last (Mins)’, ‘Total Ether balance’ and ‘Min value received’ are the three major contributing factors for detecting illicit accounts. Thai T. Pham et al. focused on detecting an anomaly, particularly in bitcoin transaction networks. They used k means clustering, Mahalanobis distance, and unsupervised support vector machines to detect suspicious users and trans actions. They used the dataset consisting of two graphs, one for users as nodes and another one as transactions as nodes.

Further, Patrick Monamo et al. also used unsupervised learning algorithms for detecting fraud in bitcoin networks. They specifically focused on the use of trimmed k-means for fraud-detection in a multivariate setup. Fa-Bin Shi et al. used a different method and focused on using financial index or

ISSN 2454-9940
www.ijasem.org
Vol 13, Issue 4, 2019

normalized logarithmic price return 539 Authorized licensed use limited to: San Francisco State Univ. Downloaded on June 18,2021 at 06:48:44 UTC from IEEE Xplore. Restrictions apply. to detect anomalies. They suggested that abnormal ask and bid prize potentially means prize manipulation or money laundering. Li Ji et al. present an exhaustive survey of data-mining techniques, including the study of deep learning techniques used for anomaly detection. They also summarized the dif ferent universal and specific detection methods. They also talk about the disadvantages and advantages of the different methods used and provide information about how this field's future may look. Bartoletti et al. also used data-mining techniques for detecting Ponzi schemes in Bitcoins. Ponzi schemes are fraudulent activities where funds from a recent investor are paid to earlier investors. They made use of features of real-life Ponzi schemes for training machine learning classifiers. Recently, Patel et al. used a sentiment analysis framework for fraud schemes detection in cryptocurrency. The decentralized framework proposed by them, KaRuNa, includes three phases of trust modeling. They employed the use of Machine Learning for measuring social trends, cryptocurrency prizes, etc. They used LSTM (Long-Short Term Memory) classifier for this purpose. Hence, many of the recent techniques used for detecting fraud are making use of machine learning for its robust nature and accuracy.

Christian Brenig et al. presents an economic analysis of money laundering using cryptocurrencies. They discuss the structure of money laundering and also propose defensive techniques. This paper further motivates our work on finding an effective way to find such fraudulent activities in cryptocur rencies and blockchain in general. Though in the presence of label scarcity, such tasks may be difficult to solve using traditional machine learning ap proaches, and thus Lorenz et al. proposed a method to detect money laundering when not enough labeled data is present. Their solution employs a real-life situation as, in many cases, labels are not present in abundance. However, their active learning solutions worked pretty well in detecting money laundering, even with very less labeled data. We studied the different types of fraudulent activities in Banking Systems, including people external to the system and employees within the Bank involved in fraudulent activities.

III.EXISTING SYSTEM

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 13, Issue 4, 2019

Many Machine Learning technique have been proposed to deal with this problem, some results appear to be quite promising, but there is no obvious superior method. We used supervised machine learning models like Decision tree, Naïve Bayes, Logistic Regression and few deep learning models in detecting fraudulent transactions in a blockchain network. Such comparative study will help decide the best algorithm based on accuracy and computational speed trade-off.

IV.PROPOSED SYSTEM :

We used supervised machine learning methods to detect fraudulent activities.They focused on the fact that malicious actors can steal money by applying well-known malware software or fake emails.Therefore, they used the capabilities of Random Forests, Support vector machines,XGBoost classifiers to identify such accounts based on a dataset of more than 300 thousand accounts.

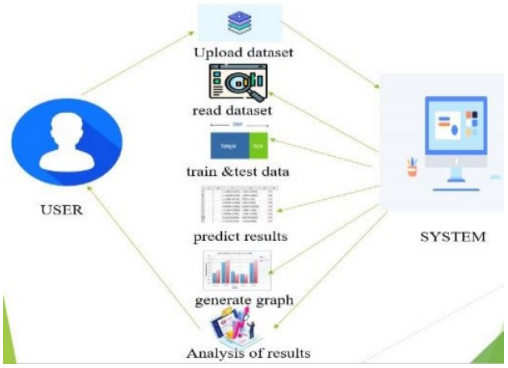
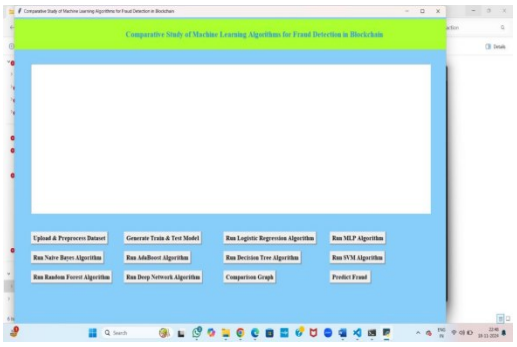
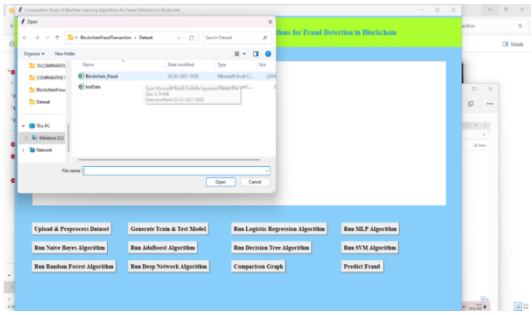


Fig.1:Architecture Diagram

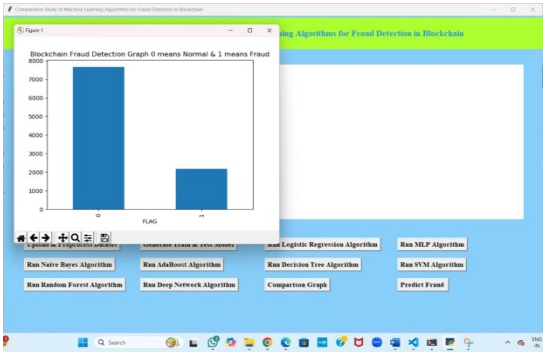
V.RESULT:



In above screen click on “upload & preprocess Dataset” button to upload and read dataset and then remove missing values.



In above screen selecting and uploading dataset and then click on ‘Open’ button to load dataset and get below output.



In above screen dataset loaded and dataset contains some non-numeric data and ML algorithms will not take such data so we need to remove and graph x-axis contains type of transaction and y-axis contains number of records and now

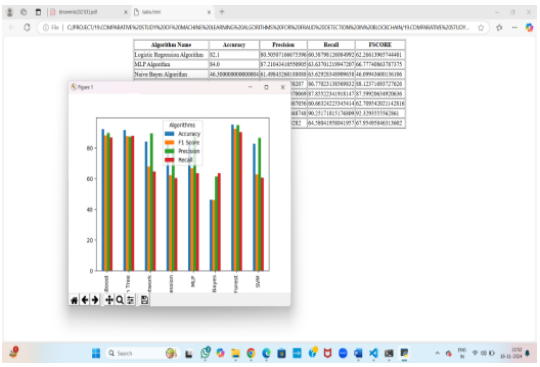
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

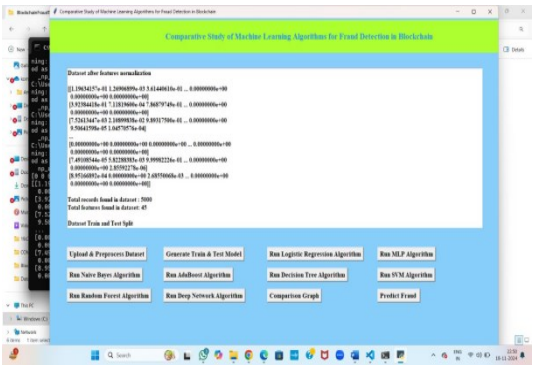
www.ijasem.org

Vol 13, Issue 4, 2019

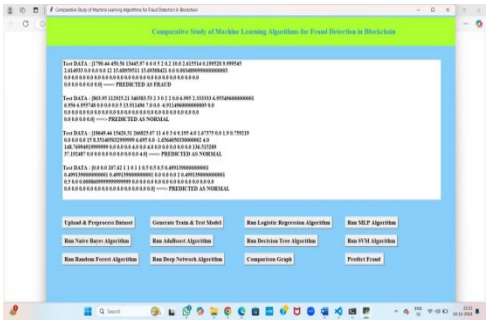
close above graph and then click on 'Generate Train & Test Model' button to get below output.



In above screen we can see the accuracy, precision, recall and FSCORE of each algorithm in graph and tabular format and in all algorithms Random Forest giving better result .



In above screen we can see all data converted to numeric format and we can see total records found in dataset with total columns and then split dataset into train and test and now train and test data is ready and now click on each button to run all algorithms and get below output.



After uploading dataset it reads the dataset and predicts data as fraud or normal.

VI.CONCLUSION

A method has been proposed for the detection of fraudulent transaction in a block chain network using machine learning.In this method,various supervised learning approaches like support vector machines,decision trees,logistic regression, and dense neural networks were analyzed.A thorough comparative analysis of the approaches is performed through accuracy.This work can be extended for the comparative study of unsupervised algorithms like clustering.In future,we also plan to do an exhaustive study on fraudulent activities in a private bloackchain. In this comparative study, we explored the efficacy of various machine learning (ML) algorithms in detecting fraudulent activities within blockchain networks. By evaluating multiple algorithms on a comprehensive dataset, we aimed to identify the most suitable approach for

enhancing the security and integrity of blockchain systems. Our findings indicate that certain ML algorithms consistently outperformed others in terms of accuracy, precision, recall, and F1-score. Notably, Random Forest and XGBoost emerged as strong contenders, demonstrating exceptional performance in classifying fraudulent transactions with high accuracy. These algorithms' ability to handle complex patterns and feature interactions proved advantageous in identifying subtle anomalies that may be indicative of fraudulent behavior. In conclusion, this study highlights the potential of ML algorithms in safeguarding blockchain networks from fraud. By carefully selecting and fine-tuning appropriate algorithms, we can significantly improve the security and reliability of these emerging technologies. Future research should focus on exploring advanced ML techniques, such as deep learning and ensemble methods, to further enhance fraud detection capabilities and address evolving threats.

VII.FUTURE ENHANCEMENTS

Advanced ML Techniques:

1.Recurrent Neural Networks (RNNs): To capture temporal dependencies in transaction sequences.

2.Convolutional Neural Networks (CNNs): To extract features from complex transaction graphs.

3.Graph Neural Networks (GNNs): To model the intricate relationships between entities within the blockchain network.

Boosting: To combine multiple weak learners into a strong classifier.

Bagging: To reduce variance and improve generalization.

Stacking: To create a hierarchical model that leverages the strengths of different algorithms.

Enhanced Feature Engineering:

1.Community Detection: To identify groups of nodes with similar behavior.

2.Link Prediction: To anticipate future relationships between entities.

3.User Profiling: To characterize normal and anomalous user behavior.

4.Anomaly Detection: To identify deviations from established patterns.

5.Smart Contract Analysis: To scrutinize the logic and execution of smart contracts.

6.Token Flow Analysis: To track the movement of digital assets across the network.

Hybrid Approaches:

1.Blockchain-Powered ML Models: To ensure transparency, immutability, and security of the learning process. ML-

2.Enhanced Blockchain Systems: To improve the efficiency and security

of blockchain operations.

Real-Time Fraud Detection:

1.Streaming Algorithms: To process large volumes of data in real time.

2.Online Learning: To adapt to evolving fraud patterns.

3.Distributed Systems: To scale fraud detection systems across multiple nodes.

Collaboration and Data Sharing:

1.Public-Private Partnerships: To foster collaboration between industry, academia, and government agencies.

2.Data Sharing Initiatives: To facilitate the development of robust ML models

VII.REFERENCES

[1] Cai, Y., Zhu, D. Fraud detections for online businesses: a per spective from blockchain technology. Financ Innov 2, 20(2016).https://doi.org/10.1186/s40854-016-0039-4

[2] Hyv¨arinen, H., Risius, M. & Friis, G. A Blockchain-Based Approach Towards Overcoming Financial Fraud in Public Sector Services. Bus Inf Syst Eng 59, 441–456 (2017). https://doi.org/10.1007/s12599-017-0502 4

[3] Xu, J.J. Are blockchains immune to all malicious attacks?. Finance Innov 2, 25 (2016). https://doi.org/10.1186/s40854-016-0046-5

[4] Ostapowicz M., ˙ Zbikowski K. (2019) Detecting Fraudulent Accounts on Blockchain: A Supervised Approach. In: Cheng R., Mamoulis N., Sun Y., Huang X. (eds) Web Information Systems Engineering– WISE 2019. WISE 2020. Lecture Notes in Computer Science, vol 11881. Springer, Cham. https://doi.org/10.1007/978-3-030-34223-4 2

[5] Podgorelec, B., Turkanovi´c, M. and Karakati˘c, S., 2020. A Machine Learning-Based Method for Automated Blockchain Transaction Signing Including Personalized Anomaly Detection. Sensors, 20(1), p.147.

[6] Farrugia S, Ellul J, Azzopardi G. Detection of illicit accounts over the Ethereum blockchain. Expert Systems with Applications. 2020 Jul 15;150:113318.

[7] Pham, Thai, and Steven Lee. ”Anomaly detection in bitcoin network using unsupervised learning methods.” arXiv preprint arXiv:1611.03941 (2016).

[8] Monamo, Patrick, Vukosi Marivate, and Bheki Twala. ”Unsupervised learning for robust Bitcoin fraud detection.” 2016 Information Security for South Africa (ISSA). IEEE, 2016.

[9] Shi, Fa-Bin, et al. ”Anomaly detection in Bitcoin market via price

ISSN 2454-9940

www.ijasem.org

Vol 13, Issue 4, 2019

return analysis.”PloS one 14.6 (2019): e0218341.

[10] Li, Ji, et al. ”A Survey on Blockchain Anomaly Detection Using Data Mining Techniques.” International Conference on Blockchain and Trustworthy Systems.Springer, Singapore, 2019.

[11] P. N. Sureshbhai, P. Bhattacharya and S. Tanwar, ”KaRuNa: A Blockchain-Based Sentiment Analysis Framework for Fraud Cryptocur rency Schemes,” 2020 IEEE International Conference on Communica tions Workshops (ICC Workshops), Dublin, Ireland, 2020, pp. 1-6, doi:10.1109/ICCWorkshops49005.2020.9145151.

[12] Brenig, Christian, and G¨unter M¨uller. ”Economic analysis of cryptocur rency backed money laundering.” (2015).

[13] Lorenz, Joana, et al. ”Machine learning methods to detect money laundering in the Bitcoin blockchain in the presence of label scarcity.” arXiv preprint arXiv:2005.14635 (2020).