Grenze International Journal of Engineering and Technology, Jan Issue

Performance Enhancement of Blockchain Systems using AI based Consensus Mechanism

Manjula K Pawar[1] and Prakashgoud Patil[2]
[1]KLE Technological University/Computer Science and Engg , Hubballi, India
Email: manjulap@kletech.ac.in
[2]KLE Technological University/MCA Department, Hubballi, India
Email: prakashpatil@kletech.ac.in

***Abstract*—Blockchain is one of the trending technologies quite popular in the 21st century. It is the brain behind the widely known and used Bitcoin networks that is implemented on a simple mechanism of a shared, immutable, and distributed network. It is ideal if the information that needs to be fetched is immediate, sharable, and completely transparent and is stored on an immutable ledger that can be accessed by only permissioned network members. It facilitates the process of recording transactions and tracking assets in a business network. A Blockchain network can do much important work like tracking orders, payments, accounts, productions, and much more. The details of the transaction can be seen from end to end, giving greater confidence. One of the properties of the Blockchain network is that it is decentralized. So, the details of the Blockchain network are substantiated by a consensus algorithm. There are some traditional consensus algorithms that are widely used, including algorithms such as PoW (Proof of Work) and PoS (Proof of Stake). They are a matter of concern due to their demerits of being computationally expensive and moving of system towards monopoly. Though there have been some optimizations on PoW and PoS, the problem still prevails. Especially the biggest challenge the Blockchain systems face is scalability. In this paper, a solution to the above problem is discussed by providing an algorithm based on Artificial Intelligence (Al) technology while reducing the demerits of the system. The main purpose Blockchain network serve is set up to generate data about the miners in order to train the machine learning model like Random Forest, KNN (K Nearest Neighbor), and Logistic Regression. The machine learning model gave outstanding results with improved efficiency and improved scalability.**

***Index Terms*— machine learning, blockchain, consensus algorithm, miners, random forest, KNN, logistic regression, scalability.**

I. INTRODUCTION

With an increasing modernization in our daily lives, society is still accepting that technology. New technologies like augmented reality and IoT have gained popularity in the past; now, the latest is Blockchain Technology[11]. Blockchain is an ingenious technology impacting different industries and was known first through the application Bitcoin. Bitcoin is nothing but a form of currency called cryptocurrency, which can be used for trading digitally. So, the success behind cryptocurrency is due to Blockchain. In a nutshell, Blockchain can be realized as a storage structure that holds records of transactions while ensuring security, transparency, and decentralization. A consensus algorithm based on such new technology could solve the traditional consensus

Grenze ID: 01.GIJET.9.1.506

algorithm's demerit, as well as maintain the efficiency of prediction of miners. It is to be evaluated whether PoAI (Proof of Artificial Intelligence) has the potential to be a good consensus algorithm or not.

Blockchain, in a nutshell, is a record of transactions in a series of blocks connected. The complete network is a series of machines, also known as nodes or miners, with no central machine present to authorize, thus, making blockchain an independent system with the features such as decentralized and democratic[1]. So, it is a Peer-to-Peer(P2P) network that provides reliability, diversity, and isolation. Since there is no central machine to coordinate the transaction, the consensus mechanism comes into play to keep the check[2]. Consensus is nothing but a process where all miners reach a common agreement on who is going to mine the block [3]. Proof of Work is a consensus algorithm where the miners have to solve a puzzle, and the winner, as a reward, gets to add transactions to the latest block of the network. The deciding role these miners play is to manage the blocks by validating them. Energy consumption is a big disadvantage here[4]. Since there are many machines across the globe, which is solving the puzzle leading to a lot of computation and energy consumption making the entire process very energy-intensive and expensive too.

The other consensus algorithm is Proof of Stake (PoS). Here, if the nodes want to participate in the mining process, then they have to put an amount as stakes. It is usually done so that only genuine miners can participate in the mining process [5]. The node having the more incredible amount of stakes has a greater probability of being getting selected. But, this process can be a major disadvantage to those miners that do not have stakes in the account [6].

The proof of Elapsed Time (PoET) algorithm works based on the lottery-based system. It is implemented on the distributed system so that the next miner of the Blockchain network can be found [7]. All miners in the distributed network need to wait for some random amount of time. This random time is called halt time. And all miners sleep during this duration. The one who gets up first after their wait time will get a chance to add block to the Blockchain, which is broadcasted to the whole network. This prevents high energy consumption. But it has a drawback: it depends on hardware for producing random time, and all nodes need to be fair. These traditional algorithms cause the scalability issues of Blockchain systems. The scalability is measured in terms of capacity, time, and throughput [8, 9, 10, 14].

To solve the traditional consensus algorithm's demerits, this paper uses a novel approach for the selection of miners by using machine learning. This new method, called Proof of AI, gives results like PoW and PoS while overcoming the weakness [12].

II. Related Work

In paper[13], the author has proposed the AI-based consensus algorithm for improving the efficiency of Blockchain systems. Here, the author has identified that most of the consensus algorithm like POW(Proof of Work) and POS (Proof of stakes) wastes a lot of resources and are monopolies, respectively. And if all nodes participate in mining, there is a waste of resources from the remaining nodes since only one node is the winner. To handle this issue AI based consensus algorithm using the Deep Neural Network(DNN) model is used, which is trained and modelled to predict the miners. This also has some drawbacks, such as to train the model takes a lot of time.

In paper[14] author provides various solutions for improving the scalability of Blockchain systems. The author proposed different techniques like off-chain, on-chain, side-chain, and inter-chain. There are various off-chain techniques like lightning network and plasma, where transaction processing is done outside the main chain, and results are stored in the main chain after processing. On-chain techniques where transactions are processed and results are stored in the main chain. In the side-chain, assets and functions from different Blockchains are exchanged. In inter-chain, communication between Blockchain is established.

In paper[15], the author proposes a Deep Reinforcement Learning(DRL) based algorithm integrated with permissioned Blockchain for vehicular edge content cashing. The proposed method improves security and privacy. This technique has been adapted to a real dataset of Uber networks, giving better results.

In the paper[16] author proposes an optimization framework for Blockchain-enabled Industry Internet of things(IIoT).The author improves the throughput of the system with novel deep reinforcement learning(DRL) based performance algorithm. This scalability is enhanced without affecting the basic characteristics like decentralization, latency, and security.

III. PROPOSED METHODOLOGY

A. Proof of AI

A machine learning model is made which helps in choosing the winner miner based on the following features[13]:

- No. of Blocks Mined -The number of blocks successfully mined by a miner.
- Faulty Transaction - It stores the information about the number of attempts taken by a miner in order to mine a particular block or modify any data or record in the transaction pool.
- Time to Mine - The time taken by a miner to successfully mine a block. It is the time between the miner selection phase and the phase in which the block is mined.
- Age - Age of a miner shows the time spent by a miner when it is not involved in the block mining.

The above parameters can play an important role in diminishing the disadvantages of PoS and PoW like centralized authority, energy consumption, and time consumption [13,17,18]. The general topology showed in figure1, this architecture represents four steps such as initially peer to peer network is established, the dataset generated for miner selection, generated dataset is used for training the machine learning algorithm.
Lastly miner is selected based on AI algorithm.

B. Implementation of Blockchain Network

In architecture that follows peer to peer mechanism, each node participates equally. It is unlike the client-server model, where there is no main head authority in a peer-to-peer architecture. Peers here are both senders and receivers. So, it can be said that no single machine is there to track the state of the whole network or system. The main objective of the proposed approach is to design a machine learning algorithm to select the miner considering fair parameters. According to the proposed method, every node gets an equal opportunity to mine the block so that the fairness of the system can be maintained. The data that is generated
from this network mainly consists of 5 features. These features are blocks, faulty transactions , time to mine, age, and output. Following are their respective meanings [13].

- **Number of Block mined**: Measures Goodness of miners.
- **Faulty Transactions**: Tells about how malicious a miner is.
- **Time to mine:** The time miner takes to mine the block.
- **Age:** It is time a miner spent without mining the blocks.
- **Output** : Whether the miner has mined a block or not with 0 being not mined and 1 being mined 15 successfully.

The implementation works on api interface only hence postman is required to execute POST and GET request. Here, port is depending on the selected port to run the server by default is set to 5000. Following are the get request ports:

http://127.0.0.1:port/validate
For each port chain, nodes of staker,stakers details can be displayed.
http://127.0.0.1:port/transactions/new JSON-data-input : sender, recipient, amount

All servers that are running in each terminal must be of different ports. Each port also automatically acts as a Staker. Following is the step-by-step process.

1. **Starting with one Server** : Switch on the Server on the Terminal, open a terminal, and run the below command to begin python blockchain.py. By default, it will start a localhost server at port 5000.https://127.0.0.1:5000.The mentioned URL is where the server is running. The above URL has no output response when opened [Yet to Build].
2. **Running with servers/ stakers:** Execute each of these in different Terminals with following command python blockchain.py –port {portnumber},

Now run multiple servers as to have more stakers :https://127.0.0.1:{portnumber}\newline

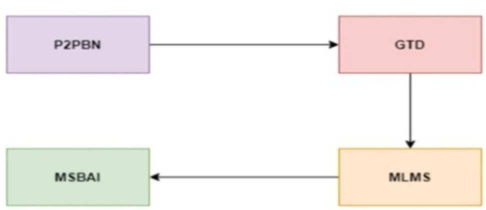


Figure 1.Topology of AI

3. **Running with servers/ stakers:** Execute each of these in different Terminals with following command
   python blockchain.py –port {portnumber},
   Now run multiple servers as to have more stakers :
   https://127.0.0.1:{portnumber}\newline
4. **Check the chain for respective Stakers/Servers:** This can be executed in postman or browser. The output would be like as shown below after executing https://127.0.0.1:{portnumber}\newline on browser.
   "chain": ["index": 1, "previoushash" : "1", "proof": 100 ,"timestamp": 1622716895.7398086, "transactions": []], "length": 1
5. **Finding the Validator :** The next step is to find a validator to execute the transaction. Right now, POS is running where a Staker is randomly picked for the first 50 Blocks. Then on, only the early Stakers who joined before 50 blocks will be added to K-means Algorithm to continue finding new Stakers. Data of a total number of blocks validated is saved in the blockchain . Important to note the port:5000 as that's where the transaction occurred.
   http://127.0.0.1:5000/validate this will search for the validator and allow them to validate the transaction with their block number.
6. **Checking chains of stakers/servers:** Now check the blockchain of selected servers in http://127.0.0.1:{portnumber}/chain, the server with longest chain is considered. Hence the consensus will have to role out the longer blockchain to others in the network.
7. **Storing Blockchain Data as a CSV file:** In the folder there is a new file with csv extension which will contain all Staker's details from the Blockchain system in http://127.0.0.1:5000/savetocsv.

C. Implementation of ML model for PoAI

This module implements the machine learning algorithms and test their accuracy in predicting a correct miner. In this module, three different machine algorithms run, namely Random Forest, KNN, and Logistic Regression. A separate function is made by the name of models , which creates a list of three models and trains them. Its function is named fit and score, which take five parameters to fit and evaluate the given machine learning models. Following are the parameters:

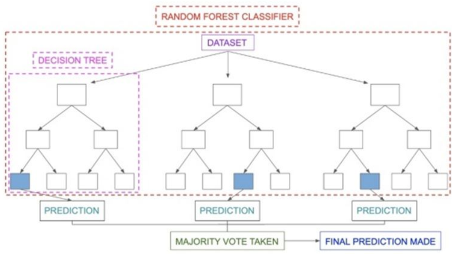
- models : a dict of different Scikit-Learn machine learning models.
- X$_$train : training data (no label)
- X$_$test : testing data(no labels)
- y$_$test : training data labels
- y$_$test : test labels

The model score function gives the accuracy score of all the three models implemented using the same set of five parameters.
Following are the machine learning models that were used to test and train the data. Since, the data needs to be classified as binary classification, there were three models that were trained.

i. **Random Forest:** It is a supervised machine learning algorithm widely used in classification and regression problem. It involves building of decision trees on different samples and takes their majority vote for classification and average in case of regression. One of the most important features of this algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems. In this case we have binary classification weather one will mine the block or not. Figure 2 shows the working of this algorithm.

ii. **Linear Regression:** Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression. Figure 3 shows the working of this algorithm.

iii. **KNN Classification:** It is one of the simplest Machine Learning algorithms based on Supervised Learning technique. The algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. It stores all the available data and classifies a new data point based on the similarity. It means when new data



. Figure 2. Working of Random Forest Algorithm

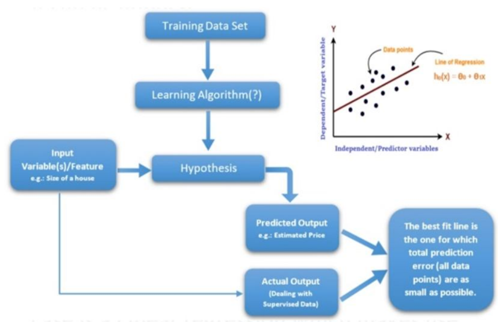


Figure 3. Working of Linear Regression Algorithm

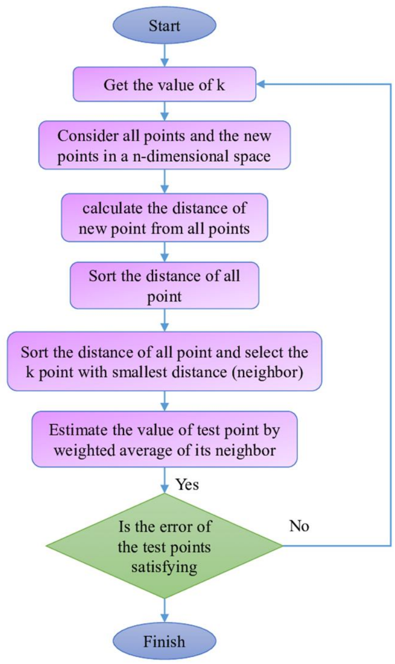


Figure 4. Working of KNN Algorithm

appears then it can be easily classified into a well suite category by using K NN algorithm. It is a non-parametric algorithm, which means it does not make any assumption on underlying data.
Figure 4 shows the working of this algorithm.

IV. Results and Discussions

In this section ,discussion about the accuracy score of each of the models is done. On analyzing the number of miner to those who were not able to mine we can see that the miner are selected at random from the model. The results of each of the models trained is shown in Figure 5. Here 0 is showing who has not mined and 1 shows who mined. The significant difference shows that consensus algorithm chooses a random miner and number of miner is less in comparison to number of miners who did not mined. Accuracy of models is given in Table I.

TABLE I. ML Models

ML Models	**Accuracy in %**
Random Forest	96.249
Logistic Regression	94.946
KNN classification	87.705

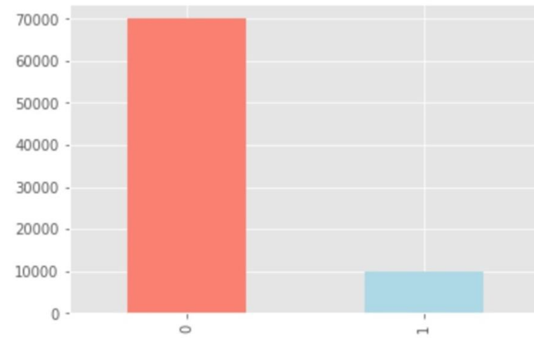


Figure 5 Miner vs. non-miner

The random forest model gave the best accuracy score of 96.249\%. The reason is that it is an ensemble classifier that averages out the results of different trees to maximize the results. Figure 6 compares the accuracy score of each of the three model used. All the three models used here are Random Forest, Logistic Regression, and KNN. Random forest yields the best accuracy because it is an ensemble machine learning algorithm that creates multiple trees comprising the forest. The results of each tree combine to give a better model score using the law of averages. While the other two machine learning algorithms like, KNN and Logistic Regression, also gave good results but less than Random Forest.

The next step is to analyze the parameters used by the model with the best accuracy score of 96.249\%. This will give us insight into which factor plays an essential role in the selection of miners. And how it can be used to develop a PoAI consensus mechanism that would be more energy efficient and free from monopoly. And increase the speed. That is how it has contributed to improving scalability. This is done by extracting the model's features to see their individual contributions in determining the miner.

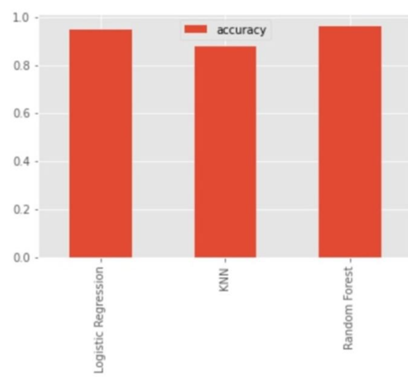


Figure 6. Accuracy Comparison

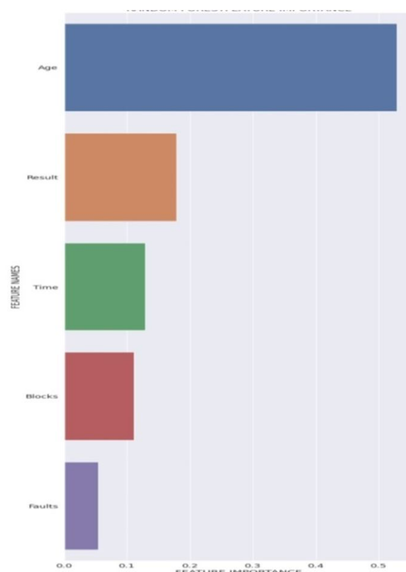


Figure 7. Feature importance of Random Forest

From the figure 7 graph it can be seen the feature importance of Random Forest model. From the graph it can be seen that age plays the massive factor in deciding the miner node which is approximately 52% and rest of the features played less than 15% role in selection of miners. The above result emphasize that the model is predicting the nodes which are relatively new and hence eliminating the 2 important disadvantage of power consumption and monopoly. The correlation matrix helps seeing the relation b/w those who mined and the input parameters.

Figure 8. Correlation matrix

The Figure 8 shows the matrix that verifies the age is the most important factor which decides the miner and the miner which has more age is not given preference in miner selection. From this section, it can be concluded that age is one of the most important factors that is determining the selection of miners by machine learning model, which suggests that miners with less age is preferred over the miner, which is the network for a longer period by our model.

V. CONCLUSIONS

Thus on, implementing different machine learning models and analysing the results can be concluded that Proof of AI (Po-AI) can become a great consensus algorithm with the positives of traditional consensus algorithm and diminishing the demerits of power consumption and monopoly. This consensus algorithm improves the scalability feature, thereby reducing the power consumption and improving the speed. The model which gave the best result with 96.4% accuracy was Random Forest with the most important parameter being age. Bitcoin and cryptocurrencies are not the only popular application of this technology.

This ingenious idea has led many scientists, businessmen, industries, and entrepreneurs from across the world to experiment and develop novel solutions and ideas to existing problems. It can be concluded that this technology has huge potential to be the future of humankind in terms of trading and currency exchange. It has the power to transform the world in a good way. There is a future scope for improving the performance with different AI-based algorithms in combination with different consensus algorithms.

REFERENCES

[1] M. Sharples and J. Domingue, “The blockchain and kudos: A distributed system for educational record, reputation and reward,” in European conference on technology enhanced learning, pp. 490–496, Springer, 2016.

[2] J. Mattila, “The blockchain phenomenon–the disruptive potential of distributed consensus architectures,” tech. rep., ETLA working papers, 2016.

[3] G.-T. Nguyen and K. Kim, “A survey about consensus algorithms used in blockchain.,” Journal of Information processing systems, vol. 14, no. 1, 2018.

[4] A. Gervais, G. O. Karame, K. Wust, V. Glykantzis, H. Ritzdorf, ¨ and S. Capkun, “On the security and performance of proof of work blockchains,” in Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, pp. 3–16, 2016.

[5] Kiayias, A. Russell, B. David, and R. Oliynykov, “Ouroboros: A provably secure proof-of-stake blockchain protocol,” in Annual International Cryptology Conference, pp.357–388, Springer, 2017

[6] E. Garcia Ribera, “Design and implementation of a proof-of-stake consensus algorithm for blockchain,” B.S. thesis, Universitat Politecnica ‘ de Catalunya, 2018.

[7] L. Chen, L. Xu, N. Shah, Z. Gao, Y. Lu, and W. Shi, “On security analysis of proofof-elapsed-time (poet),” in International Symposium on Stabilization, Safety, and Security of Distributed Systems, pp. 282– 297, Springer, 2017.

[8] Pawar M.K., Patil P., Hiremath P.S. (2021) A Study on Blockchain Scalability. In: Tuba M., Akashe S., Joshi A. (eds) ICT Systems and Sustainability. Advances in Intelligent Systems and Computing, vol 1270. Springer, Singapore.

[9] Pawar M.K., Patil P., Hiremath P.S., Hegde V.S., Agarwal S., Naveenkumar P.B. (2021) “Scalable Blockchain Framework for a Food Supply Chain”. In: Thampi S.M., Gelenbe E., Atiquzzaman M., Chaudhary V., Li KC. (eds) Advances in Computing and Network Communications. Lecture Notes in Electrical Engineering, vol 735. Springer, Singapore. https://doi.org/10.1007/978-981-33-6977-1_35.

[10] M. K. Pawar, P. Patil, M. Sharma and M. Chalageri, "Secure and Scalable Decentralized Supply Chain Management Using Ethereum and the Interplanetary file system (IPFS) Platform," 2021 International Conference on Intelligent Technologies (CONIT), 2021, pp. 1-5, doi: 10.1109/CONIT51480.2021.9498537.

[11] Nakamoto, Satoshi. ”Bitcoin: A peer-to-peer electronic cash system.” Decentralized Business Review (2008): 21260

[12] AI and Blockchain: A Disruptive Integration: T. N. Dinh and M. T. Thai, ”AI and Blockchain: A Disruptive Integration,” in Computer, vol. 51, no. 9, pp. 48-53, September 2018, doi: 10.1109/MC.2018.3620971.

[13] N. Kumar, C. Parangjothi, S. Guru and M. Kiran, "Peer Consonance in Blockchain based Healthcare Application using AI-based Consensus Mechanism," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225550.

[14] S. Kim, Y. Kwon, S. Cho, A survey of scalability solutions on blockchain, in 2018 International Conference on Information and Communication Technology Convergence (ICTC) (Jeju, 2018),pp. 1204–1207.

[15] Y. Dai, D. Xu, K. Zhang, S. Maharjan and Y. Zhang, "Deep Reinforcement Learning and Permissioned Blockchain for Content Caching in Vehicular Edge Computing and Networks," in IEEE Transactions on Vehicular Technology, vol. 69, no. 4, pp. 4312-4324, April 2020, doi: 10.1109/TVT.2020.2973705.

[16] M. Liu, F. R. Yu, Y. Teng, V. C. M. Leung and M. Song, "Performance Optimization for Blockchain-Enabled Industrial Internet of Things (IIoT) Systems: A Deep Reinforcement Learning Approach," in IEEE Transactions on Industrial Informatics, vol. 15, no. 6, pp. 3559-3570, June 2019, doi: 10.1109/TII.2019.2897805.

[17] Pawar, M.K., Patil, P., Patel, A.S. (2023). Distributed and Scalable Healthcare Data Storage Using Blockchain and KNN Classification. In: Zhang, YD., Senjyu, T., So-In, C., Joshi, A. (eds) Smart Trends in Computing and Communications. Lecture Notes in Networks and Systems, vol 396. Springer, Singapore. https://doi.org/10.1007/978-981-16-9967-2_70.

[18] M. K. Pawar, P. Patil, R. Sawhney, P. Gumathanavar, S. Hegde and K. Maremmagol, "Performance Analysis of E-Certificate Generation and Verification using Blockchain and IPFS," 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 345-350, doi: 10.1109/ICICT54344.2022.9850830.