ISSN 2319 – 7595

Volume 8, No.2, March - April 2019

International Journal of Information Systems and Computer Sciences

Available Online at http://warse.org/IJISCS/static/pdf/file/ijiscs10822019.pdf

https://doi.org/10.30534/ijiscs/2019/10822019

An Overview on Integrating Machine Learning with Blockchain

Kevin Shaju Varghese[1], Kezia Ann Mathew[2], Riya Rachel Samuel[3], Srijith Sasikumar[4], Fabeela Ali Rawther[5]

[1]Amal Jyothi College of Engineering, Kottayam, India, kevinshajuvarghese@cs.ajce.in

[2]Amal Jyothi College of Engineering, Kottayam, India, keziaannmathew@cs.ajce.in

[3]Amal Jyothi College of Engineering, Kottayam, India, riyarachelsamuel@cs.ajce.in

[4]Amal Jyothi College of Engineering, Kottayam, India, srijithsasikumar@cs.ajce.in

[5]Amal Jyothi College of Engineering, Kottayam, India, fabeelaalirawther@amaljyothi.ac.in

ABSTRACT

Machine learning and blockchain are among the most disruptive technologies and will fundamentally reshape how we live, work, and interact. Also, with smart contracts, blockchain can govern interactions among participants with no intermediary or a trusted third party. Blockchain can power decentralized marketplaces and coordination platforms for various components of machine learning, including data, algorithms, and computing power. Blockchain will also help machine learning's decisions be more transparent, explainable, and trustworthy. As all data on blockchain is publicly available, artificial intelligence is the key to providing users with confidentiality and privacy. The alliance of machine learning and blockchain is expected to create numerous possibilities. The design and operation of a blockchain involve thousands of parameters and tradeoffs between security, performance, decentralization, and many others. Artificial intelligence can ease those decisions and automate and optimize blockchain for higher performance and better governance. Moreover, as all data on blockchain is publicly available, AI plays a key role in providing users confidentiality and privacy.

Key words: Artificial Intelligence, Blockchain, Machine Learning, Smart Contract

1. INTRODUCTION

Combining two disruptive technologies such as Machine Learning and Blockchain have paved the way for innovations. Blockchain mainly deals with the invention of the cryptocurrencies and its trading. Machine Learning deals with the identification of patterns and learning data with the help of predictive and descriptive algorithms. Both have the potential to hasten data exploration and analysis as well as intensify transactions security. Machine Learning requires huge data sets to make accurate predictions and so to initiate this process, distributed blockchains can be the input for machine learning.

1.1 Blockchain

In layman terms, blockchain is a distributed ledger of records that is immutable and verifiable. The blockchain is a mechanism that allows transactions to be verified by a group of unreliable actors. It provides a distributed, immutable, transparent, secure and auditable ledger. The blockchain can be consulted openly and fully, allowing access to all transactions that have occurred since the first transaction of the system, and can be verified and collated by any entity at any time [1]. The blockchain protocol structures information in a chain of blocks, where each block stores a set of Bitcoin transactions performed at a given time [2]. Blocks are linked together by a reference to the previous block, forming a chain. [12]The most common application of blockchain is in finance and banking domain.

The following are the characteristics of blockchain:

•*Decentralization*: When we consider the conventional centralized transaction systems, each transaction is validated through the central trusted agency (e.g., the central bank). This results in cost and performance bottle-necks at the central servers. On comparing with the centralized mode, the requirement for a third party is no longer needed in a blockchain. To maintain data consistency in a distributed network, consensus algorithms are used [3].

•*Persistency*. The transactions that take place in a blockchain can be validated quickly. If the miners are honest, invalid transactions would not be admitted. Once data is included in the blockchain, it is nearly impossible to delete or rollback the transactions. It is very quick to identify the blocks that contain invalid transactions [3].

•*Anonymity*. When a user interacts with the blockchain with a generated address, blockchain maintains the anonymity of the user. Due to the intrinsic constraint of the blockchain, one must also know that blockchain does not guarantee perfect privacy preservation [3].

1.2 Machine Learning

Machine learning is defined as a computational process that is used to discover the underlying models of system behavior. Machine learning takes datasets, processes them, and attempts to discover causal variables. Machine learning algorithms can be used to gather understanding from the data under study, abstract the understanding of underlying phenomena in the form of a model, predict the future values using the generated model and detect any form of anomalous behavior shown by a phenomenon under observation.[13]

Machine learning can be categorized into two categories. The first category involves the process by which it learns, and it is thereby classified into, supervised learning, semi-supervised learning, and unsupervised learning. The second category involves the type of output or problem it attempts to solve, and it is thereby classified into classification, clustering, regression, and anomaly detection. [13]

2. INTEGRATING MACHINE LEARNING WITH BLOCKCHAIN

2.1 Sentiment Analysis

Let us consider an E-Com website or app that gets a lot of feedback, suggestions, and complaints by users. This E-Com website runs on the blockchain. In this system, the transactions take place with the help of smart contracts. Suppose there is a provision for customers to give reviews about a seller. Here we can use sentiment analysis to analyse the online conversations happening on digital media about a few products that the seller provides. With the help of sentiments of user reviews, it can give you an idea about overall brand perceptions. It also promotes a wholesome customer-seller interaction.

Sentiment analysis or Opinion mining is a common tool which is used for classification and it analyses the incoming data as positive, negative and neutral. It is a field of Natural Language Processing (NLP). More than identifying the opinion, the system extracts the attributes of the opinion. As a first step it deals with the polarity that is dividing the opinion as positive or negative opinion and then it deals with the subject that is the thing being talked about and at last, it deals with the opinion holder or the person expressing the opinion. Text can be broadly categorized into facts and opinion. Facts are the objective expressions about something. Opinions are usually subjective expressions that describe the sentiments and feeling towards a topic.

Sentiment analysis [7]can be applied at different levels of scope that is at the document, sentence and sub-sentence level. Document- level analyses the complete document. Sentence level obtains the emotion of a single sentence while the sub-sentence level analyses the sub-expression. In order to analyze the customer reviews, we need to mainly focus on sentence and sub-sentence level.

Generally, the type of sentiment analysis being followed is the faint grained sentiment analysis. The customer review is often classified into different categories (see Table I).

Table I: Categories for Customer Review

Category	Rating
Very Positive	5 stars
Positive	4 stars
Neutral	3 stars
Negative	2 stars
Very Negative	1 star

A. Sentiment Analysis Algorithms

The methods to implement sentiment analysis can be classified as:

Rule-based systems perform sentiment analysis based on a set of manually crafted rules. These rules may use different type of inputs such as classic NLP techniques like stemming, tokenization etc and other resources such as lexicons. [14]

Automatic systems that rely on machine learning techniques and make the machines learn. This system does not rely on manual crafted rules unlike rule based system. The problem of sentiment analysis is often modelled as a classification problem in which the classifier is fed with the input text and returns the corresponding category. [14]

Hybrid systems that combine the best of both rule based and automatic approaches in order to improve accuracy and precision of the system. [14]

2.2 Price Prediction

Machine learning and deep learning have a very crucial role in financial institutions for their power in predicting data that varies with time, with a high degree of accuracy and there are still researches going on to increase the capabilities of the model. Stock market price prediction is one of the most difficult tasks. This prediction includes a variety of factors physical factors, physiological, rational and irrational behaviour, etc. These factors combined make share prices very difficult to predict with accuracy.

The analysis of the stock market involves two processes: fundamental analysis and technical analysis. Fundamental analysis analyzes the company's future profitability. This is done based on the current business environment and financial performance. Technical analysis focuses on reading the charts and using the statistical figures to identify the trends in the stock market.

The stock market prediction involves using various machine learning algorithms and methodologies. The various methods involve support vector machine using RBF kernel algorithm [5], deep neural networks [6] etc. The stages of stock price prediction involve data collection, deciding the input parameters which are done using historical stock data, deciding the parameters like open high, moving averages etc., combining the features, using SVM classifier and the last step involves accuracy check [5].

On questioning the aspect of interoperability, trust and transparency that arises in the market systems, blockchain can be the best answer. The traders, brokers, regulators and stock exchange, who are referred to as the stock market participants, are required to go through a cumbersome process (which takes more than 3 days to complete transactions, because of the role of intermediaries and regulatory processes). Automation and decentralization of blockchain make stock exchanges optimal by

reducing the cost on the customer in terms of commission while speeding up the process for fast transaction settlements. The technology can have a viable use in clearing and settlement, while securely automating the post-trade process, easing paperwork of trade and legal ownership transfer of the security.[4]

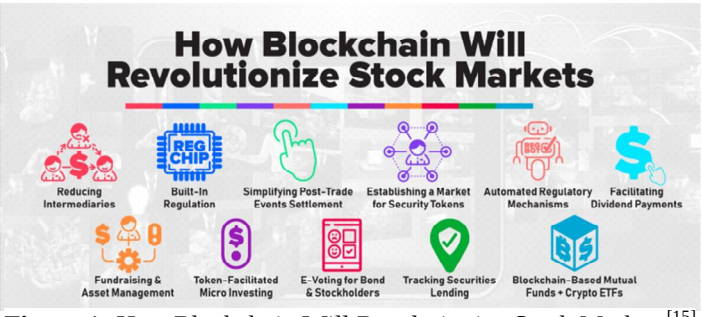

Figure 1: How Blockchain Will Revolutionize Stock Markets[15]

For a transaction, the blockchain technology can act as an online automated surveillance system. An exchange empowered by blockchain will have characteristics to track, block and report an illegitimate attempt by anyone in the network and thus provide a robust platform. Blockchain transactions are faster, as smart contracts by peers are used for trade confirmation other than an intermediary. The inefficiencies, cost and entry barriers can be reduced through automation in the blockchain. For people, who could not access the markets due to cost barriers will be able to participate, ultimately increasing liquidity and investment [4].

Third party regulator can be eliminated to a large extent with the advent of blockchain, which is because the rules and regulations are in-built within the smart contract and are to be enforced with each trade to register transactions with the blockchain network as a regulator. Applying blockchain and smart contracts to post-trade activities can eliminate the intermediaries reduce counter-parties and operational risk while providing the infrastructure for faster trade settlement. Financial institutions can settle securities in minutes [4].

Through blockchain technology, margining system and payment of margin can be done instantly and the frequency of valuation of securities deposited as capital can be done daily compared with the weekly process prevalent now, minimizing the risk [4].

2.3 Social Media

It is no doubt that the popular online social media sites are governed by logically centralized services. There are two marketing strategies which focus on data collection and on learning as much as possible about potential customers, their tastes, their habits, their spending patterns, or even their feelings and states of mind. Based on such information, potential customers can be smartly targeted, or retargeted by presenting the right product to the right person at ideally the right time, making the advertisement more profitable. This collected data does not only cover the information that social media users willingly upload and share with their contacts, but encompasses it to implicitly disclosed information, such as the times when users are online, the locations from where they connect, the type of activities they perform based on different locations and times, etc. [8]

A. Decentralized Social Network

The control of data pertaining to the users remains in the hands of social media and thus they have too much power to dictate terms. Traditional social media sites do not promote security and privacy of information. The blockchain is a technology that gives prime importance to these aspects. There is a more equitable balance of power between the platform provider and its users because of the decentralized infrastructure of blockchain-based social media. Traditional social media sites are centralized in nature. Due to this, they are easy targets for those who wish to misuse the data. Harvesting or otherwise misappropriating data is far more difficult without a central point of weakness to attack, thus ensuring the privacy of data.

B. Control over Self-Created Content

The content that the users post on social media is not patented. As a result of this, anybody from any part of the world can obtain access to the content that different users post on their social media profile. One can also further copy content and take credit for another user's creation. Due to this, people are at high risk of losing out on credits that they deserve. In order to have total control over the content that a user has created and earn credits for it, users should consider using blockchain-enabled social media platform.

Integrating with machine learning enhances the possibility of having a privacy protected and secure social media browsing. Nowadays, social media sites have started to use machine learning technology to their advantage. With it, the site can detect the spam content and the users, recommend the content and predict the likelihood that a person will use the content, can monetize ad performance.

2.4 Healthcare Industry

Healthcare interoperability describes the ability for heterogeneous information technology systems and software applications, such as the EHR (Electronic Health Record) system, to communicate, exchange data, and use the exchanged data. Interoperability enables providers to securely share patient medical records with one another (given patient permissions to do so), regardless of provider location and trust relationships between them [9].

Secure and scalable data sharing is essential to provide an effective collaborative treatment and care decisions for patients. Data sharing helps improve diagnostic accuracy by gathering confirmations or recommendations from a group of medical experts, as well as preventing inadequacies and errors in the treatment plan and medication. This helps the doctors to analyze the patient's need and thus enables them to make an effective diagnosis [9].

Let us consider the scenario of a cancer research centre. Here, groups of physicians with different specialties in cancer care

form tumor boards that meet regularly to discuss cancer cases, share knowledge, and determine the effective cancer treatment and care plans for patients. Consider another scenario if a cancer patient under treatment is admitted to the emergency room (ER) in a different hospital, then it would be critical for the ER provider to access the patient's medical records to identify potential drug interactions; while the patient's primary cancer care provider should be notified that the patient is being treated in the ER[9].

Nowadays, patients who are diagnosed with a disease say cancer prefer to obtain a second opinion on their medical condition, diagnosis and treatment plan. Once they reach out to a doctor or board of doctors, they are required to submit the essential documents such as their family history, a record of their previous diagnosis, treatment plan. All these reports are to be physically submitted to the doctors. This is a tedious process to start off with as it would require a lot of time to obtain these documents.

Now, blockchain technology offers the opportunity for trustless exchange and disintermediation that allow existing trust relationships to be aggregated and propagated across various organizations and providers. This approach is like the patient referral process, except the referrals are not limited to a single provider's network. Instead, they could be expanded across different regions, states, and even countries. A Blockchain-based system can also capture existing trust relationships between patients and providers, allowing patients to decide which provider(s) can share their data [9].

Machine learning is widely used to diagnose various types of cancer such as breast cancer, lung cancer. Medical image examination is the most the effective method for diagnosing breast cancer. Such methods include digital mammogram, magnetic resonance imaging (MRI), ultrasound etc. Images interpretation is operator dependent and it requires a great deal of expertise. Using CAD (Computer-aided Detection and Diagnosis) which uses machine learning classification algorithms can be very useful for the doctors to make an effective diagnosis[10].

Computer aided detection entails the use of a computer output that only yields the location of suspect lesions. Characterization and diagnosis of the abnormality as well as patient management are left for the radiologist. A complete CAD system involves the use of segmented structure, the detection of abnormalities and the extraction of their characteristics for the classification of the problem. There are mainly four stages. The first stage is the data preprocessing stage. In this stage, the images are prepared for subsequent stages such as cleaning the medical image and removing noise for it through a set of preprocessing operations. The second stage is the segmentation of the region of interest in the image. This procedure consists of dividing the input image into several regions. The third stage is the feature extraction and selection stage. Here the features are extracted from the cleaned images. From these cleaned features, the best and the most discriminative features are chosen. The selected features help in determining the normal and cancerous regions. These selected features are collected and stored in a database. These features become input to the classification stage. In the final stage, classification takes place. This is a data mining stage that involves the assignment of labels or classes to different groups. It extracts hidden patterns using different machine learning techniques such as K. Nearest Neighbors, Artificial neural network, Support Vector Machine. By selecting the apt machine learning technique, it builds a classifier responsible for classifying different breast lesions [11].

3. CONCLUSION

In this paper, we present a combination or collaboration of two technologies, machine learning and blockchain, with the faith on the longevity of both the technologies. As the blockchain technology enhances the security and integrity of data, machine learning can be integrated to predict or make the machines learn many things. Blockchain stores the data while machine learning uses this data to identify various factors of the entity involved. Machine learning is advancing day by day but the scope of blockchain still need to be explored when compared. Blockchain has various scopes in the future other than its use for crypto-currencies. Blockchain, which provides very secure transactions, can be integrated with machine learning for sentiment analysis in E-commerce, medical fields, can be used for stock market price prediction, and social media. The integration of machine learning with blockchain has a wide range of scope for researches, which can prove to be a boon and further improve the innovation in these research areas.

REFERENCES

[1] Aras et al. 2017. ***Blockchain and Its Applications – A Detailed Survey***. 2017 International Journal of Computer Applications

[2] Satoshi Nakamoto. 2008, ***Bitcoin: A Peer-to-Peer Electronic Cash System***, [Online] http://www.bitcoin.org

[3] Dataflair team. 2018. ***6 Major Features of Blockchain***, Jun 2018, https://data-flair.training/blogs/features-of-blockchain

[4] economictimes.indiatimes.com:***How blockchain will change the trade in stock markets way you***

[5] Raut Sushrut Deepak et al. 2017. ***Machine learning approach in stock market prediction.***

[6] Sotirios P.Chatzis et al. 2018. ***Fore casting stock market crisis events Using Deep And Machine Learningtechniques.*** https://doi.org/10.1016/j.eswa.2018.06.032

[7] Shashank Gupta, ***Sentiment Analysis: Concept, Analysis and Applications***, www.towardsdatascience.com

[8] Leila Bahri et al.2018 ***Decentralized privacy preserving services for Online Social Network*** https://doi.org/10.1016/j.osnem.2018.02.001

[9] Peng Zhang et al.2018: ***Blockchain Technology Use Cases in Healthcare*** https://doi.org/10.1016/bs.adcom.2018.03.006

[10] NisreenI.R.Yassin et.al 2018: ***Machine learning techniques for breastcancer computer-aided diagnosis using different image modalities: A systematic review*** https://doi.org/10.1016/j.cmpb.2017.12.012

[11] Maryellen L. Giger et.al 2008: ***Anniversary Paper: History and status of CAD and quantitative image analysis: The role of MedicalPhysics and AAPM***

[12] AnaReyna et.al 2017: ***On blockchain and its integration with IoT. Challenges and opportunities***

[13] Thomas W Edgar et.al 2017,Pages 153-173: ***Research Methods in Cyber Security Chapter 6: Machine Learning.*** https://doi.org/10.1016/B978-0-12-805349-2.00006-6

[14] MonkeyLearn: ***Sentiment Analysis Nearly Everything You Need To Know,*** https://monkeylearn.com/sentiment-analysis

[15] Sam Mire. 2018. ***Blockchain for Stock Markets: 11 Possible Use Cases***, https://www.disruptordaily.com/blockchain-use-cases-stock-trading/