International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 12 | December - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Integrating Blockchain and Machine Learning for Healthcare Systems

Pasham Narayan Rohith Reddy, Dipankar Maiti, Kommu Pavan Sai[1]

Ms. P.Sandhya Priyanka [2]

[1] Department of Information Technology, Malla Reddy Engineering College, Secunderabad, Telangana-500100

[2]Assistant Professor, Department of Information Technology, Malla Reddy Engineering College, Secunderabad, Telangana-500100

Abstract: Health care systems are that systems which help the patients reach directly the concerned specialist. These are the systems which are having lots of advantages during the pandemic situation and in high emergency situations.In the proposed work user will search for the disease summary (disease and treatment related information) by giving symptoms as a query in the search engine. Initially when a pdf is downloaded and saved in the system it first performs per processing on the data in the document and the extracted relevant data is stored in the database. The symptoms entered by the user are further classified using SVM classifier to make the further process easier to find the semantic keyword which helps to identify the disease easily and quickly. Then the semantic keyword found is matched with the stored medical input database to identify the exact disease related to that keyword present. Once the disease related to the symptom is identified, it is sent to medical database to extract the articles pertaining to that disease. The preprocessing process involves tokenization, removal of stop words and stemming. Followed by that, relevant information is extracted using the keyword searching algorithm. In our implementation of our proposed system, we have used SVM classifier which gives us an improved result. The decentralized database where transactions get recorded in append only shared ledger has many advantages in healthcare industry. In medical treatment, the complete history of patient is very important and value is added when same information is accessed by different parties. The convergence of these two technologies can give highly accurate results in terms of machine learning with the security and reliability of Blockchain Technology.

Keywords: Machine Learning, Tokenization, Stemming, SVM classifier, Naïve Bayes, Decision Trees, Blockchain.

1. Introduction

Blockchain Technology got recognition from Bitcoin cryptocurrency which was described in Satoshi Nakamoto's whitepaper in 2008. This can be explained as, if someone in the Blockchain network initiates a transaction. The transaction will be broadcasted in the P2P network called as nodes. The nodes will validate the transaction. If the transaction gets verified, it will be combined with other valid transaction to form a block. The block is marked with previous block's hash and timestamp and will be added to the existing Blockchain. The block is permanent and unaltered. It has no problem of single point of failure and there is no central server to manage the network. Blockchain Technology has gone far beyond Bitcoins. Healthcare is one of its application areas. The proposed system was based on bitcoins approach was meeting information customer's need and protecting patient's privacy.

In healthcare, many parties have to access same information. Latest treatment information is available to prevent use of outdated information in a transparent way. It also shows previous interaction between patent and all physicians. In this, the patient's health data like heartbeat, blood pressure, sleep patterns etc. can be taken from various health apps available or from wearables or physician visit. This data was stored in Blockchain. In medical treatment, the complete history of patient is very important and value is added when same information is accessed by different parties.

The permission blockchain network will have following components: 'Shared Ledger', It is the append only distributed system which keeps the track of all the transactions that take place in a blockchain network. It is maintained by each peer in the blockchain network. Peer Network Peers will validate any transactions that are initiated by any node in the network. The validated transaction will be further added in the block. 'Membership/Certificate Authority', The user needs permission to join the network. The Certificate Authority will authenticate the users and ensures that they gets right access to ledger for the transaction they are performing. Smart Contract The business logic or the software running on the ledger.

2. Literature Survey

A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution. Digital signatures provide part of the solution, but the main benefits are lost if a trusted third party is still required to

prevent double-spending. We propose a solution to the double-spending problem using a peer-to-peer network [1]. The protocol has been designed to adhere to fundamental e-voting properties as well as offer a degree of decentralization and allow for the voter to change/update their vote (within the permissible voting period). This paper highlights the pros and cons of using blockchain for such a proposal from a practical point view in both development/deployment and usage contexts [2]. Methods for using blockchain to provide proof of pre-specified endpoints in clinical trial protocols were first reported by Carlisle. We wished to empirically test such an approach using a clinical trial protocol where outcome switching has previously been reported. Here we confirm the use of blockchain as a low cost, independently verifiable method to audit and confirm the reliability of scientific studies [3]. In the proposed work user will search for the disease summary (disease and treatment related information) by giving symptoms as a query in the search engine. Initially when a pdf is downloaded and saved in the system it first performs per processing on the data in the document and the extracted relevant data is stored in the database. The symptoms entered by the user are further classified using SVM classifier to make the further process easier to find the semantic keyword which helps to identify the disease easily and quickly. Then the semantic keyword found is matched with the stored medical input database to identify the exact disease related to that keyword present. Once the disease related to the symptom is identified, it is sent to medical database to extract the articles pertaining to that disease. The preprocessing process involves tokenization, removal of stop words and stemming. Till now the best result obtain is 98.51% F-measure by Oana Frunza, for the extraction of cure and prevents relations. In our implementation of our proposed system, we have used SVM classifier which gives us an improved result. The problem statement of the existing system was, it didn't identify the best disease treatment. So, the proposed solution used data mining concepts using voting algorithm to resolve the problem and find the best treatment for disease out of the treatment identified by the system [4]. Telemedicine uses message and information technology to deliver health care when distance keeps the patients far from doctors. It plays a noteworthy title role to modernize the health care systems mainly in the developing countries. Telemedicine is popularly classified as tele-cardiology, tele-dermatology, tele-radiology, tele-ophthalmology, etc. based on the nature of medical services delivered in this manner. The Government of India has launched a number of projects to explore the feasibility of application of telemedicine to extend better health care services to the people in the rural and remote areas of the country. As a result, India is now leading the developing world in the installation and support of telemedicine services to remote regions through various types of telecom networks ranging from satellite to optical fiber to mobile [5].

3. Materials and Methods

A. Datasets

In this dataset we have a set of symptoms through which the model tells the disease of the patient. The patient can lifestyle advice. On the basis of patient's symptoms, the trained model can give clinical suggestions to doctors. If the symptom mentioned by the patient is not present in the dataset then it prints a message saying – "Good news! You're going to have a disease named after you!". If the mentioned symptom is present in any of the dataset then it is going to print a message saying – "You may have these diseases".

Disease	Symptoms
Cold	Sore throat, runny nose, congestion, cough, aches
Flu	Fever, headache, muscle aches, returning fever
Ebola	Tiredness, death, bruising over 90% of body, black blood
Spondylosis	Tingling, numbness, weakness, Abnormal reflexes, muscle spasms
alcohol misuse	Antisocial behavior, impulsivity, self-harm, loneliness
Stroke	Numbness arm, Confusion, Difficulty speaking, difficulty walking, slurred speech
Lower respiratory infections	Phlegm, fever, difficulty breathing, a blue tint to the skin, chest pain, wheezing
Pulmonary	Dyspnea, Fatigue, fainting spells, Chest pressure, Swelling
Bronchus	Cough with blood, Wheezing, Shortness of breath, Chest pain, Flushing
Diabetes	thirst and hunger, urination, Weight loss or gain, Fatigue, Nausea, Blurred vision
Alzheimer	Memory loss, Vision loss, Misplacing items, Difficulty making decisions, meaningless repetition
Dehydration	Vomiting, sweating, Individuals, dry mouth, lethargy, dizziness
Tuberculosis	Coughing, Chest pain, weight loss, Fatigue, Fever, Night

	sweats, Chills
Cirrhosis	Jaundice, Weakness, Loss of appetite, Itching, Easy bruising, dark urine
Plague	diarrhoea, nausea, malaise, delirium, shortness of breath, tender lymph node

Table- Symptoms dataset

B. Machine Learning Algorithms

1) SVM Classifier

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

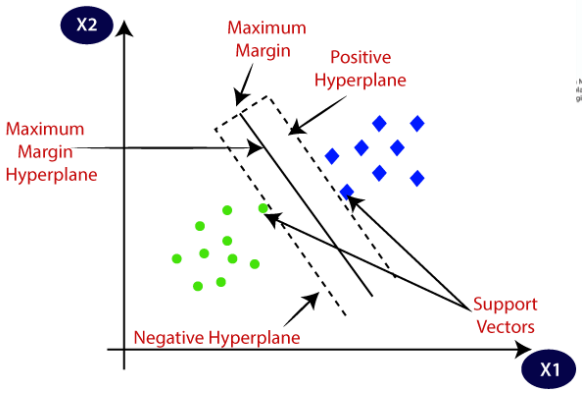

Fig-1: SVM Classifier using Hyperplane

2) Naïve Bayes Classifier

Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

Bayes' Theorem:

Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
The formula for Bayes' theorem is given as:

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

where,
P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.
P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.
P(A) is Prior Probability: Probability of hypothesis before observing the evidence.
P(B) is Marginal Probability: Probability of Evidence.

Working of Naïve Bayes' Classifier:

Working of Naïve Bayes' Classifier can be understood with the help of below steps :

1. Convert the given dataset into frequency tables.
2. Generate Likelihood table by finding the probabilities of given features.
3. Now, use Bayes theorem to calculate the posterior probability.

3) Decision Tree Classification Algorithm

Decision Tree is a **Supervised learning technique** that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

Below diagram explains the general structure of a decision tree:

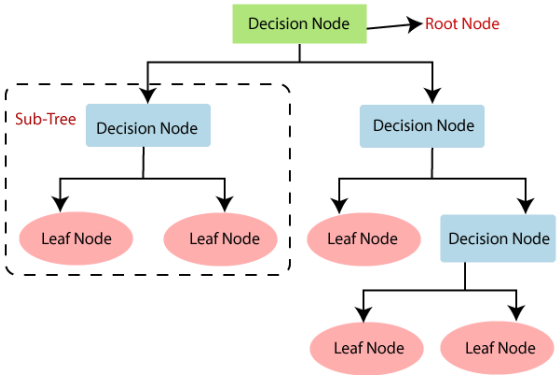

Figure-2 Decision Tree nodes

4. Methodology

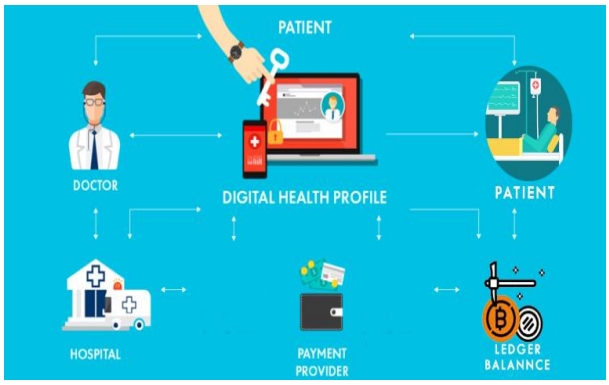

Figure-3 System Architecture

Description of Proposed Method

Doctor: On the basis of symptoms and other laboratory reports, the train model can predict the outbreak and the diseases related to it.

Patient: On the basis of patient's symptoms/problems it can give suggestions and lifestyle advices. And since the trained model deals with real data, this increases the accuracy and efficiency.

Hospital: Expert doctors from different hospitals will login and help the patients on the basis of symptoms and reports. And thus, doctors will suggest the medicines.

Payment Provider: After diagnosis process of the patient then the hospital will provide a payment slip online and the patient will further pay it by online transaction.

Ledger Balance: Ledger report shows the history of patient services, service charges and descriptions, applied payments and adjustments and also shows remaining balances. Undisbursed patient payments also appear on this report. Applications of this method is

- Secure storage and integrity protection.
- Privacy and ownership of data.
- It increases accuracy and efficiency.
- It provides data reliability and data sharing.
- It provides transparency .and accountability of data.

5. Results and Discussion

From fig 4(a)-4(d) Doctor gave suggestion soon the basis of symptoms and other laboratory reports, the train model can predict the outbreak and the diseases related to it. On the basis of patient's symptoms/problems it can give suggestions and lifestyle advices. And since the trained model deals with real data, this increases the accuracy and efficiency. Expert doctors from different hospitals will login and help the patients on the basis of symptoms and reports. And thus, doctors will suggest the medicines. Finally, after diagnosis process of the patient then the hospital will provide a payment slip online and the patient will further pay it by online transaction.

Figure-4(a) Home Page

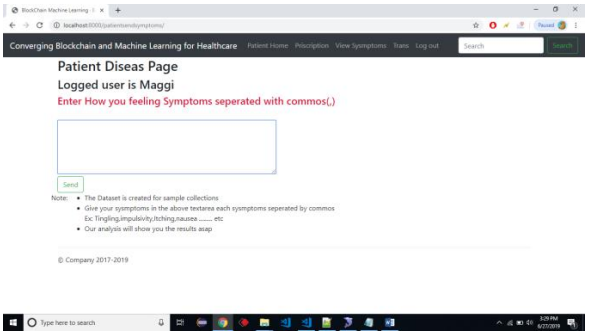

Figure-4(b) Patient - Prescription page

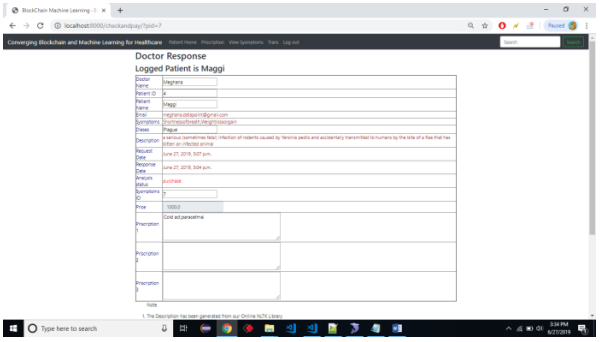

Figure-4(c) Doctor - Response page

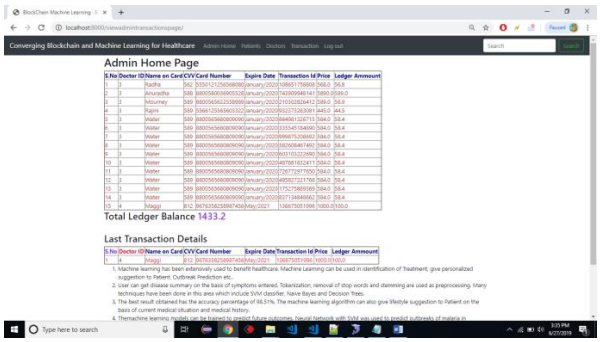

Figure-4(d) Admin - Transactions page

6. Conclusion

Blockchain Technology gives great number of opportunities if utilized properly and is seen beyond bitcoin. With blockchain, the dominance of central authority could be eliminated and so the commission. Machine Learning models can directly fed with data (however the rights will be managed by central authority). This will increase the accuracy and efficiency of machine learning models and so their usability. Healthcare industry directly correlates with the life of a person. This could help patients as well as doctors. In terms of future scope, the practical implementation of this model will be there. This model can be further extended for Inventory to prevent fraud. Like any new technology, the blockchain is an idea that initially disrupts, and over time it could promote the development of a larger ecosystem that includes both the old way and the new innovation.

References

[1] N. Satoshi (2008), Bitcoin: A Peer-to-Peer Electronic Cash System, [Online]. Available: https://bitcoin.org/bitcoin.pdf

[2] Freya Sheer Hardwick, Apostolos Gioulis, Raja Naeem Akram, and Konstantinos Markantonakis, E-Voting with Blockchain: An E-Voting Protocol with Decentralisation and Voter Privacy

[3] Kan Yang, Xiaohua Jia, Kui Ren, Bo Zhang, Ruitao Xie, How blockchain-timestamped protocols could improve the trustworthiness of medical science

[4] Bharti E. Nerkar and Sanjay S. Gharde, "Best Treatment Identification for Disease Using Machine Learning Approach in Relation to Short Text", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 22788727Volume 16, Issue 3, Ver. VII (May-Jun. 2014), PP 05-12.

[5] S. Som, R. Roy Som, R. Mahajan (2016) "Survey of e-Health Status and Services in Eastern India", Far East Journal of Electronics and Communication (International Journal), Vol. 3, Scopus Indexed, ISSN: 0973-7006, November 2016

[6] G. Dwyer. "The economics of Bitcoin and similar private digital currencies," Journal of Financial Stability, vol. 17, pp. 81-91, April 2015.

[7] G. Irving, H. John (2016, May), How blockchain-timestamped protocols could improve the trustworthiness of medical science. [Online].

[8] Keerrthega.M.C, Ms.D.Thenmozhi "Identifying Disease Treatment Relations using Machine Learning Approach", 2016 International Conference on Computational Science.

[9] G. Prisco (2016, April), The Blockchain for Healthcare: Gem Launches Gem Health Network With Philips Blockchain Lab, [Online].

[10] O. Williams-Grut (2016, March), Estonia is using the technology behind bitcoin to secure 1 million health records,[Online].

[11] P. Taylor (2016, April), Applying blockchain technology to medicine traceability, [Online].