ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024

Review of Using Machine Learning and Blockchain to Protect Data Privacy

[1] Shivam Raj, [2] Divya Thakur

[1] Lovely Professional University
[2] Assistant professor, Lovely Professional University
Corresponding Author Email: [1] rajshivamms@gmail.com, [2] divyathakurdt73@gmail.com

Abstract*— Today, information is collected for no specific reason; every machine or individual action is recorded, and the data can be analyses if required in the future. However, as data becomes more reliable, the issue of trust will surface in periods for inspection by several parties. The data could include confidential or sensitive data the organizations that participate in the process of analysis could misuse. As a result, it makes sense to address data privacy challenges now. Data privacy involves with how to restrict the usage of data according to its importance. For example, someone will tell strangers his or her name but will not reveal his or her mobile phone contact through the outsider seems at comfortable. In up-to-date digital age, data pertaining to privacy is usually applied to essential personal information. Data privacy extends beyond the crucial data of employees and customers from a business standpoint. Because ML utilizes massive data sets to train and evaluate, there is widespread recognition that Intelligence and Machine Learning-powered technology are impeded by concerns about data security. Can this barrier block, however, be transformed into an entry stone Yes, it is conceivable. Consider the scenario that follows: no one can be believed; in this case, Blockchain enter the picture. Blockchain makes use of data, however it does it securely. So, in this work, we're presenting a way for ensuring data security through combining the block chain and Machine Learning.

Index Terms*— Blockchain, integration, data protection, encryption, machine learning.

I. INTRODUCTION

Advances in computer training and artificial intelligence are being hindered with a major impediment: data privacy. In principle, confidentiality of information indicates information is only to be utilized for its intended purpose. People possess the right to control when their data is gathered and utilized, a concept known as data privacy. Data privacy is an aspect of security for information that deals with data handling in line given consent, notification, and regulatory obligations. If the data is put to use If such wouldn't be the case, we might conclude they has been a violation of privacy or data privacy. So, how do we overcome the greatest limitation posed by data privacy? It's acceptable. It is well understood that a vast amount of data is needed for the training and evaluation of Computer Intelligence; the greater the amount it is learned, the greater the precision it will perform. As an outcome, when huge quantities of information are utilized, it is impossible to discern between general and private information. Following gathering data, language proficiency will go through numerous stages of analysis, increasing the likelihood that anyone here/some organization will abuse the unique/private data. There may be a simple answer to the aforementioned security of information problem: don't collect it in any form. Is this, however, even possible? No, the system will give it should be unambiguous that in lack of data, everything is conceivable. In the lack of data, however would an investigation be carried out? This is the point at which the technology and algorithmic learning are going to be used to protect data while without compromising affecting statistical effectiveness. The exactness of the information evaluation procedure will be reduced if private data is not available. So, rather than completely banning the utilization of private data, how about sanitizing it prior to analysis subsequently is simple to accomplish with the help of the digital currency blockchain. Another challenge is establishing whose information is was not intended. the greater the amount of data utilized to train and assess a machine learning system, the more reliable and accurate the findings. This is widely known. As a result, distinguishing between data that may be used broadly and data that cannot be used broadly can be difficult. When vast quantities of data are used, private data is compromised. Following the gathering of data, it will go through several stages of analysis, increasing the likelihood that a person or an organization will take advantage of any confidential or personal information. A simple answer to the quandary of data privacy is to not gather any data at all. Is this, however, a realistic possibility Obviously, nothing is possible in the absence of evidence. At this stage, machine learning, blockchain, and other technologies will be employed to ensure data privacy without compromising data analysis efficacy. Evaluation of data will be less accurate without private information. So, instead of completely eliminating the utilization of private data, how about anonymizing it before it is collected and analysed. It is simple to succeed with the help of blockchain technology. Furthermore, because each information distributor is oblivious of how-to categories private data, determining which information has become secure and what data is not is difficult. in such an instance,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024

algorithms trained with machine learning can help determine which at first data is considered personal and what information is not. As a result, we expected the combined effect of blockchain technology and machine learning to be effective. to maintain the private nature of material without impeding analysis. The remaining sections the document's contents is organised as follows. Part II discusses both the literature to be reviewed review and the specifications for the proposed assignment. Part III provides a comprehensive review of the present situation state of data privacy. Section IV explains why ML and Blockchain should be combined. Section V discusses the way algorithmic learning can be used. used, as well as the numerous ways. Machine learning algorithms can be used to identify whether or not a piece of information is private. As a result, we reasoned that merging Machine Learning with Blockchain would preserve private information while facilitating analysis.. In all industries, the fight for prospective markets and larger portions of the pie is ramping up. With the use of machine learning and artificial intelligence. In any case, it appears that continued growth of predictive modelling and Artificial Intelligence technologies is delayed by a key impediment: Data Privacy. In principle, data privacy signifies information is best utilized or utilized for its intended purpose. Consumers having the right to control where their data is gathered and utilized, a concept known as data privacy. Data privacy are the portion to the privacy of information involving with data handling in line given consent, announcement, and regulatory obligations. A compromise in security or an infraction of data privacy occurs when data is utilized for reasons for which it was not intended. So, how Is it possible to overcome the fundamental hurdle provided by data privacy. It is commonly understood that a huge amount of personal information needs to be collected for the training and evaluation of Computer Training the greater the amount it is learned, the better it will perform. As a result, it becomes challenging when huge quantities of data are used. to determine the information that can be utilized over wider uses and which data must be kept private. It will go through numerous rounds of analysis after data collecting, making it quite complex. It is probable that someone/someone Will misappropriate personal/private information over time. There might have an easy approach to an existing privacy problem don't collect any kind of data completely. However, is this feasible, given that everything can be achieved without data. In the lack of data, exactly can a research project be carried out This is the area where the combination Artificial learning alongside blockchain technology are going to be utilised for shielding personal information. whereas preventing interference alongside data analysis performance. The reliability of an information analysis technique will be lowered if private data is not available. As a result, without interfering with the entire utilization of personal information What if the personal information is anonymized before being analyzed subsequently is simple to accomplish with the help of blockchain. Furthermore, because each information provider is not conscious of directions categories private It is difficult to discern exactly what information is protected and what data is not. In this case, artificial intelligence can be used to decide which information is private and which is not. As a result, we thought that the combined power of Blockchain technology and predictive modelling would be capable of keeping information private while also allowing for analysis. Section VI discusses whether blockchain technology and machine learning might collaborate. The study finishes with suggested recommendations for future research in Section VII.

II. REVIEW OF LITERATURE

Machines learning-based technological advances is developing at a pace that is accelerating. As the outcome, the concentrate is exclusively on the application of an essential part of the processing of Data. systematically and effectively, they are able to utilise the analysis, the better the accuracy and the lesser the variance.and the ability to provide an accurate forecast. As a result, effective data and machine learning use is critical. the authors of this paper [1] reviewed ways to enhance the overall integrity of privately collected data, as improved quality contributes to greater accuracy. They emphasise the issue of confidence in utilising on a data collection obtained by an entity or organisationwith no prior experience collecting theme- based data. Additionally, they proposed ways a values-driven society might help technicians gain confidence with the data that they acquire. In this study, the paper's authors [2] propose a mechanism for establishing consistency throughout an untrusted Personal/Private Databases. This can be useful whenever you've got multiple phases and participants engaged in a lengthy procedure. Their role is to make specific that no one attempts to change the personal information throughout the transmission process.

Author Name	Proposed Technique	Drawback
M. Boukhlif, M. Hanine and N. Kharmoum	softwaretesting,artificial intelligence, biblioshiny application; VOSviewer.	Limited Scope, Lack of ContextualAnalysis,,Limited Generalizability
H.Taherdoost	Blockchain technology artificial intelligence AI;	ScalabilityIssues, DataPrivacyConcerns,,Interoperability

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024

S. Tanwar, Q. Bhatia, P. Patel, A. Kumari, P. K. Singh and W.-C. Hong,	Blockchain, Artificial intelligence, the energy efficiency, security of information, security, and statistics.	Outdated,Lack of ComprehensiveAnalysis ,Overemphasison Technical Solutions	
R. Wang, J. He, C. Liu, Q. Li, W. Tsai, and E. Deng	Artificial intelligence (AI), analysing information, blockchain technology, insertion, and data security.	Scalability Issues,Data PrivacyTrade-offs ,LimitedMachine Learning Capabilities, Security Risks	
B.K.Mohanta,A.Sahoo,S.Patel, S.S. Panda, D. Jena, and D. Gountia		Authentication, Decentralized, Blockchain, IoT Device, Ethereum.	Latencyand Performance,Data Privacy,Security Risks,Dependency onBlockchain Infrastructure

The writer of this paper [3] database reprocessing. Oftentimes statistics is acquired for an initiative and subsequently abandoned, but almost identical contents is gathered again for a work that is almost identical to the first. The one preceding it. essentially obtaining an identical set of data again, a data collection can be specific to the demands of the task The author [4] This document discusses ways for getting rid of superfluous data. In general, data is usually gathered without a specific objective in consideration, however we will eventually realise throughout the evaluation procedure that the most the data used represents irrelevant. Furthermore, as data grows exponentially every day, storing all of it is a major issue Where the conclusions of the data might be maintained having keeping the complete more extensive information been explored.The authors and editors of the present paper [5] explore ethical collecting information practices. Information must be accurate and accessible. The difficulties encountered during information extraction are also addressed. The paper's authors [6] provide an overview of the data's oldest known history and investigate the classification of data origination. This study [7] contains an extensive evaluation of safeguarding privacy data analytics. a fair comparison of various elements of a survey, such as Security publication, analytical activities, security mechanisms, and so on. The authors [8] described whether artificial intelligence can be applied to forecast application determination with the method known as Random Forest to categories The information provided was divided into six groups utilising the characteristic recovery method. The authors of [9] applied machine learning techniques to evaluate the reliability of location. maps data in the current investigation. Blockchain-based technologies is being viewed as an extremely exciting innovation in our current context since it solves the reliability challenge. Whenever multiple parties are A breakdown of security develops as a result of the procedure of analysis and the transmission of details between them. Blockchain technologyaddresses this issue by establishing an autonomous, trustfree environment. The authors of this paper [10] talked about how to guarantee data privacy, they proposed a permissioned blockchain network, as well as a privacy aware PKI system. This paper [11] [12] discusses how blockchain can be used to improve trustworthiness. The authors of the paper of this paper [13] [14] briefly covered how to utilise a system based on blockchain when there is a possibility of data modification. Additionally, they proposed a decentralised blockchain technology system. that does not require real-time analysisthe authors of this paper [15] [16] provided an outline of how to use the blockchain architecture in an unbreakable and accessible manner while adhering to all data privacy rules. The authors of the paper of this writing [17] indicate how to combine artificial intelligence and Blockchain to communicate data in an interconnected system. They additionally supply a comprehensive discussion of the applications of Blockchain technologies and Machine Learning and built a collection of intersections Wherever two forms of technology meet. The study [18] hints at upcoming advances in the field of blockchain technology. Even though AI/ML simplifies today's challenges, worries concerning privacy and the safety of data have been since the dawn of data application. Information stored in databases with data concerning personalities can be classified into the following categories, based on security and privacy concerns:

- Personally Identifying Specific data (PII) refers to material that can be used to locate someone personally (e.g., Aadhaar number, phone number, email address, and so on).
- Quasi-Identifiers (QI) comprises elements the fact that require further information to authenticate an identity. (e.g., PIN code, age, sex, etc.).

Confidential Columns of data These refer to qualities that do not fit into the that preceded two classes yet comprise details regarding the individual's identity the fact that must be maintained for multiple purposes. (e.g., pay, HIV detection,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024

Bank account subtleties, live geo-area, and so on

Non-controversial Columns — these are the remaining characteristics that do not fit into the first three (e.g., nation, college, etc.). certainly QIs believe that eliminating PII parts within a collection of data is unacceptable for safety assurance. For example, if key quantitative information (which qualifies as QI) can be obtained in a the database, it may be merged with additional accessible sources, such as voter registration. List to specifically identify those individuals. So, ahead of we go into the details with the personal information security scenario, let's list to specifically identify those individuals. So, ahead of we go into the details with the personal information security scenario, let's appear into the way data moves through the various stages. (See Figure 2). The initial period, Identifying the Objective, can potentially not be regarded a stage. because it does not directly contribute to the analysis process, but it is the most important because it determines the primary goal of data gathering. The personal information collection procedure begins in its subsequent step, Data Collection. The details can be acquired in a variety of ways, including online/offline questionnaires, random data collecting from various sites, for instance vehicles going through a toll the square, and monitoring employees' entrance and exit timings at operate. All of the information gathered is irrelevant. Whenever the full package of data is analysed using insignificant information, it'll be take a long period without producing the desired results. As a result, in the third step, the irrelevant data is removed. Data cleaning. The final phase is Data Preparation, which entails preparing the data to be consistent with the study's objectives. Considering the subsequent (5th) step is Analytics & Visualisation, data must be meticulously prepared to avoid inaccurate visualisations. The third level is the machine learning process, which conducts practical research and predicts O/P. The five and sixth steps can be paired together, but for simplicity, we keep them distinct. allow us immediately discuss the security of data.. Analyses Data Responsibly requires that technological advancement not jeopardize people's privacy security. Additionally, there has been of course, a very simple approach to ensure 100% protection do not gather the data in any way. Because such a technique contradicts the major premise of the data-driven basic technique for data analysis, confidentiality security and safety techniques have attempted to compensate for data utilisation. Three broad approaches have arisen, including.

1) Regulation of accessibility:-This approach explains whoever is providing the information into a database, how, and the purpose for which she or he needs the information being entered.
2) An anonymous collection of the information:- This approach alters expertise for the purpose to protect the true identity of persons and data.
3) Information sharing and Privacy Protection:- The technique centres on securing Multiple parties Computation (SMC), which provides distant utilisation of encrypted data while maintaining characterised and controllable security.

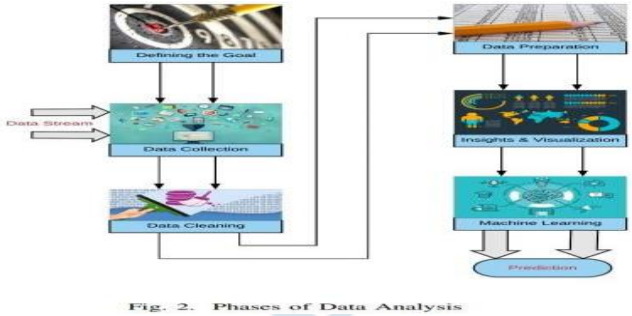


Fig. 2. Phases of Data Analysis

Fig. 1. Phases of Data Analysis

III. PROPOSED METHODOLOGY

A. Integrating With Blockchain

Previously, in the present scenario part, we discussed how data is processed from the point of collection to the point of prediction. Many parties and stages are involved in the flow of the dataprocess to perform the processing. As a result, there is a high risk that data will be disclosed, stolen, or misused. The main and critical issue is the misuse of private data. To address this issue, we propose modifying the previous data analysis process's framework. To guarantee data privacy, the new proposed framework combines Machine Learning and Blockchain. Figure 3 depicts the entire process as well as how the Blockchain module was integrated into the data analysis process. The second level of the data analysis process follows the definition of the goal.

B. Cryptographic Approaches

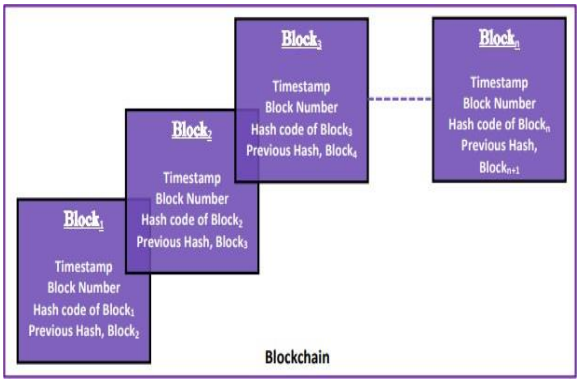
Basically, homomorphic encryption, garbled circuits, and secret sharing are the three kinds of cryptographic methods used to achieve privacy in ML:

1) Homomorphic Encryption: This method is used to make calculations involving encoded or encrypted information more difficult.
2) Garbled Circuits: This protocol makes it possible for two parties to safely communicate while evaluating each other's functions and work over a private channel without the involvement of a third party.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024



3) Secret Sharing: This encryption technique divides the data into multiple parts and distributes them to various parties. Thus, the parties can only combine their individual shares to recover the data. On occasion, the bearer of the Data specifically referred to as the dealer generates n & specifies the thresholds t for various shares that call for rebuilding the entire data.

C. Machine Learning in Data Security

Various input was required collectively to prepare machine learning models without missing out on applying the private data information in their unique structure [25], which was achieved by using cryptographic methodologies, or differentially-private utilizations of data (annoyance systems). Differential security is particularly useful for anticipating positive results [15].

D. Approaches of Perturbation

This method uses mathematical techniques to determine an approximation of the answer to a problem by starting from the precise solution of a problem that belongs to the same category as the one being solved. Differential privacy (DP) techniques in privacy-preserving machine learning employ perturbation methods (PPML). There are two divisions of the perturbation approach:

1) Differential Privacy (DP): DP is a framework for freely sharing dataset data by showing examples of gathering within the dataset while maintaining the data of people in the dataset.
2) Local Differential Privacy: A subclass of differential privacy, local differential privacy has some additional limitations. Even if someone is able to access a person's private information, they will not be able to obtain anything the complete information of that person. The stage is the level of data collection. Data is gathered in a variety of ways, including randomly or directly from users. When data is gathered at random, it is impossible to determine which data should be classified as private and which should not be.

The data that is collected directly from the users, where the users can classify which data is deemed private and which is not. So, following the data collection step, two streams of data are generated:

1) Private and general records are classified and classified.
2) There is no private or universal classification for random data. The second data stream is fed into the machine learning1 module, which is designed and taught to classify and categories Private and General data from the initial random data stream. The first category of data in the stream is immediately fed into the blockchain module, which is the next stage of the Machine Learning-1 stage.

The Blockchain module's specifics will be covered eventually in this section.

When data is collected from different users, if the user understands what is private data and what is general data, the user can determine which data is private and which is general. As a result, the data can be immediately fed into the blockchain module. If the user is not sufficiently aware to Differentiate between private and public data before feeding it to the Machine Learning-1 tool. Machine Learning-1 is intended to classify and categories private and public data. We can train the machine using the first category of the data stream, which is gathered directly from responsible and knowledgeable users..

1. Preliminaries

We started this section by describing collaborative filtering and the characteristics of our platform (for example, privacy and management). We also discussed about ways for integrating collaborative filtering and privacy. We describe the cryptographic structure and cryptographic tools that make our platform. Then (For example, browsing information and purchase history) is recorded in order to produce future recommendations. Both approaches are employed in collaborative filtering, though the implicit voting method has expected privacy difficulties. usually, users are not told that the system will maintain a log of their activity This kind of behavior presents a number of concerns, such as the invasion of the user's privacy.

A lot of companies are also doing business with this confidential information without paying for customers' personal info. Apart from voting techniques, collaborative filtering provides two algorithmic approaches. Finally, we go over the assumptions that we utilized to build our protocols Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

2. Current Scenario

As the supply of AI and machine learning-related possibilities for businesses increases and becomes increasingly inescapable. Data protection and security experts from various backgrounds are likely to face many challenges. The issues are drawing a line of social criteria,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024

holding candidate/private information, and being truthful and forthcoming about its use. It was never a black or white thing in the first place, but now there are going to be a lot more shades of grey [19]. Before collecting data for purposes of analysis, the question of why the data is being collected must be resolved. While AI/ML decreases the complexity of current issues, concerns about

data security and privacy have been present since the very beginning of data utilisation. Databases keeping information about individuals can be classified into the categories that follow, according to security and privacy queries:

- **Personally Identifiable Information (PII)** – these are parts that may be directly related to or identified an individual (e.g., Aadhaar number, phone number, email address, and so on).
- **Quasi-Identifiers (QI)** – these are components that may not be helpful without the input of others, but can be combined with other Quasi-Identifiers, query results, and some outside data to identify an individual (e.g., PIN code, gender, age, etc.).
- **Sensitive Columns** — These are qualities that do not fit into the previous two categories but contain data regarding the individual that should be protected for various reasons (e.g., pay, HIV identification, Bank account intricacies, live geo-area, and so on).

Non-controversial Columns – these are the remainder traits that do not fit into the first three (e.g., nation, college, etc.).

All of the data that was collected is worthless. If the entire data set is processed with irrelevant information, it will take a long time without giving the intended results.

In the third stage of Data Cleaning, the irrelevant information is removed. The fourth stage is Data Preparation, in which the data is generated in accordance with the analysis's objective. Although the next (fifth) stage is Insights & Visualization, data must be carefully prepared to avoid any incorrect graphical representations. The next stage is Machine Learning, which performs the actual analysis and predicts the O/P. The fifth and sixth stages can be combined, but for clarity, we maintain them separate. Let us now consider the data security scenario analyses Data appropriately necessitates that advances in technology not jeopardize people's privacy protection. There is, of obviously, an incredibly simple solution to guaranteeing the ultimate security: not gathering the information. Considering this methodology defeats the primary desired of the data-driven essential approach to analyzing information.

In general, such as information security assurance and protection approaches have attempted to achieve an offset with data implementation. Three broad approaches have come about, including:

1) **Regulation of access**:- This method explains who is entering the database and how, as well as why she or he requires the data.
2) **Anonymization of data**:- This methodology is used for altering information in attempt to minimize people's and data's identities.
3) **Sharing information with Privacy Security:-** The technique depends on the use of Secure Multiparty Computation (SMC), which ensures remote access to private information with specified and controlled security.

Hybrid blockchain solutions: In the literature, hybrid blockchain solutions are being studied as an approach for preserving data privacy. These ideas entail mixing a public and a private blockchain to form a hybrid network. It is possible to ensure that sensitive data is retained on a private blockchain while nevertheless benefiting from the safety and openness of a public blockchain by employing a hybrid network. Zhang et

al. (2021), for example, suggested a hybrid blockchain-based framework for secure and private machine learning, which has been shown to be effective in preserving data privacy. Secure multi-party computation is a mechanism for doing computations on the data of numerous parties without revealing private data. It can be done to use secure multi-party computing. It is possible to run machine learning algorithms on blockchain without releasing sensitive data. Zhang et al. (2020), for example, proposed a secure multi-party computation-based system for secure machine learning on blockchain. The framework has been determined to be effective in protecting data privacy while yet allowing for the production of accurate machine learning models.

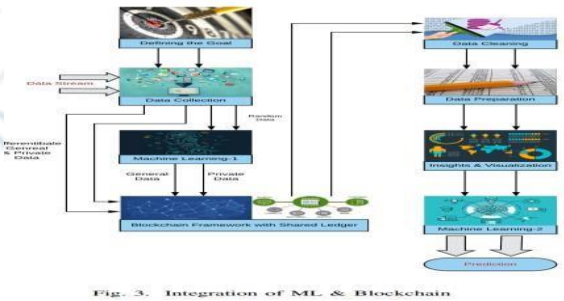


Fig. 3. Integration of ML & Blockchain

Fig. 2. Phases of Data Analysis significant benefits in ensuring data privacy and security:

1. **Immutably Data Storage:** Blockchain technology provides a decentralised and immutable database for data storage that is visible and tamper-proof. This characteristic is critical for ensuring the consistency of the training information used in supervised learning methods. The organizations may verify that the training data remains unmodified and auditable by storing it on the blockchain, avoiding unauthorised modifications and protecting data privacy.
2. **The Encryption of Data and Control over Access:** For training and inference, prediction models require access to enormous amounts of data. Not all data in a blockchain, however, must be exposed to all participants. Highly sensitive information can be

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024

secured and kept on the blockchain using algorithms for encryption and access control systems, guaranteeing that only authorised persons with the right decryption credentials can access and use the data. This method allows for extremely fine supervision of data privacy while limiting exposure to unauthorised users.

3. **Decentralised Machine Learning (ML):** Conventional machine learning models frequently process and analyse data on centralised servers. However, because the data must be transported and held in a single location, this centralised technique poses questions concerning data privacy and security. Organisations may train and deploy machine learning models in a decentralised manner by merging machine learning and blockchain. Privacy remained onto the decentralised ledger and is handled within a distributed manner in this configuration, limiting the probability of security breaches and unauthorised access.
4. **Sharing of Data using Smart Contracts:** Intelligent contracts, essentially executing itself agreements contained inside the blockchain system, can be used to establish the conditions and limitations of data sharing. Smart contracts can be used by businesses to implement data privacy laws, determine whoever is granted permission to the information, and manage how it is used. This method addresses data privacy concerns in a secure and automated manner, decreasing the need to depend on confidence between parties.
5. **Learning through Machine Learning Techniques that Protect Your Privacy**: To improve data privacy, various privacy-preserving machine learning algorithms can be combined with blockchain. Techniques like as unique confidentiality, federated analysis, and homomorphic cryptography allow data to be handled and analysed avoiding compromising the foundational sensitive information. Organisations can gain increased privacy by integrating these strategies with blockchain predictive models while taking advantage of the advantages of networked consent and data integrity.
6. **Accessible Auditing and Compliance:** The transparency and auditability of blockchain can help with fulfilment of data privacy requirements. Because of the blockchain's immutability, organisations may observe and verify when data is collected and processed, ensuring compliance with privacy rules that include the General Privacy Regulation (GDPR). Organisations can verify conformance through supplying accountable documentation of all data transactions and blockchain processing stages.
7. **Credibility and Verification:** Combining machine learning and blockchain can boost participant trust by creating a decentralised and verified platform. Participants can ensure the accuracy of the data, techniques, and predictions utilised in predictive processes by verifying their integrity. Transparency and verifiability enhance system trust and boost optimism about data security and protection.

IV. CONCLUSION

With the massive increase in data in recent years, it is critical to address the problem of responsible data use. As the use of data increases, so does the value of private data. This paper explains why there is a need to combine the most cutting-edge technologies, Machine Learning and Blockchain. This paper explains the data flow throughout the entire analysis process, from goal setting to prediction. There was discussion about how data can be manipulated and how confidential data can be misused. It also briefly describes how machine learning can be used to address the issue of data privacy on its own, before moving on to the paper's main goal, the integration of Machine Learning & Blockchain. The integration will allow for a necessary solution to the most sensitive data issue. Privacy. It will secure, efficient, intelligent, robust, and decentralized the entire data processing process. In We have only suggested a framework for integrating Machine Learning and Blockchain in this paper; real-time analysis and implementation remain to be completed. Because this is a prototype, it can be used in any industry such as finance, healthcare, defense, e-voting, banking, and so on. In summary, the integration of Machine Learning and Blockchain is vast, and numerous challenges lie ahead. The suggested model's limitations are that when every data provider is added as a node in the blockchain framework, the processing time will be reduced. To address this problem, a parallel processing strategy was developed can be used, as well as a high-performance device. We also anticipate that some sophisticated algorithm will be developed in the near future to handle the limitations. This paper attempts to investigate integration at a very preliminary level, as well as to address future research directions that may profit from our work.

REFERENCES

[1] M. Boukhlif, M. Hanine and N. Kharmoum, "A decade of intelligent software testing research: A bibliometric analysis", Electronics, vol. 12, no. 9, pp. 2109, May 2023.

[2] H. Taherdoost, "Blockchain technology and artificial intelligence together: A critical review on applications", Appl. Sci., vol. 12, no. 24, pp. 12948, Dec. 2022.

[3] B. Ji, Y. Zhao, J. Vymazal, Ü. Mander, R. Lust and C. Tang, "Mapping the field of constructed wetlandmicrobial fuel cell: A review and bibliometric analysis", Chemosphere, vol. 262, Jan. 2021.

[4] R. Wang, M. Luo, Y. Wen, L. Wang, K.-K. R. Choo and D. He, "The applications of blockchain in artificial intelligence", Secur. Commun. Netw., vol. 2021, pp. 1-16, Sep. 2021.

[5] B. K. Mohanta, D. Jena, U. Satapathy and S. Patnaik, "Survey on IoT security: Challenges and solution using machine

learning artificial intelligence and blockchain technology", Internet Things, vol. 11, Sep. 2020.

[6] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, ''Blockchain and federated learning for privacy-preserved data sharing in industrial IoT,'' IEEE Trans. Ind. Informat., vol. 16, no. 6, pp. 4177–4186, Jun. 2020.

[7] S. S. Gill, S. Tuli, M. Xu, I. Singh, K. V. Singh, D. Lindsay, et al., "Transformative effects of IoT blockchain and artificial intelligence on cloud computing: Evolution vision trends and open challenges", Internet Things, vol. 8, Dec. 2019.

[8] S. Tanwar, Q. Bhatia, P. Patel, A. Kumari, P. K. Singh and W.-C. Hong, "Machine learning adoption in blockchain-based smart applications: The challenges and a way forward", IEEE Access, vol. 8, pp. 474-488, 2020.

[9] A. Parthy, L. Silberstein, E. Kowalczyk, J.-P. High, A. Nagarajan, and A. Memon, "Using machine learning to recommend correctness checks for geographic map data," in Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice, ser. ICSE-SEIP '19. IEEE Press, 2019, p. 223–232. [Online]. Available: https://doi.org/10. 1109/ICSE-SEIP.2019.00032

[10] R. Wang, J. He, C. Liu, Q. Li, W. Tsai, and E. Deng, "A privacyaware pki system based on permissioned blockchains," in 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Nov 2018, pp. 928–931.

[11] Y. Yuan and F.-Y. Wang, "Blockchain and cryptocurrencies: Model techniques and applications", IEEE Trans. Syst. Man Cybern. Syst., vol. 48, no. 9, pp. 1421-1428, Sep. 2018.

[12] B. K. Mohanta, A. Sahoo, S. Patel, S. S. Panda, D. Jena, and D. Gountia, "Decauth: decentralized authentication scheme for iot device using ethereum blockchain," in TENCON 2019-2019 IEEE Region 10 Conference (TENCON). IEEE, 2019, pp. 558– 563.

[13] M. Sahoo, S. S. Singhar, B. Nayak, and B. K. Mohanta, "A blockchain based framework secured by ecdsa to curb drug counterfeiting," in 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1–6.

[14] M. Sahoo, S. Samanta Singhar, and S. Sahoo, A Blockchain Based Model to Eliminate Drug Counterfeiting, 03 2020, pp. 213–222.

[15] B. K. Mohanta, S. S. Panda, and D. Jena, "An overview of smart contract and use cases in blockchain technology," in 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018, pp. 1–4.

[16] U. Satapathy, B. K. Mohanta, D. Jena, and S. Sobhanayak, "An ecc based lightweight authentication protocol for mobile phone