Novel identity management system using smart blockchain technology

Article *in* International Journal of Systems Assurance Engineering and Management · November 2021

DOI: 10.1007/s13198-021-01494-0

6 authors, including:

Sumegh Shrikant Tharewal
Symbiosis Institute of Computer Studies and Research (SICSR), Symbiosis Interna…
43 PUBLICATIONS **249** CITATIONS
SEE PROFILE

Mukesh Soni
Smt. S. R. Patel Engineering College
70 PUBLICATIONS **1,641** CITATIONS
SEE PROFILE

Ehtiram Raza Khan
Jamia Hamdard University
92 PUBLICATIONS **244** CITATIONS
SEE PROFILE

Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-021-01494-0

ORIGINAL ARTICLE

Novel identity management system using smart blockchain technology

A. Shobanadevi[1] · Sumegh Tharewal[2] · Mukesh Soni[3] · D. Dinesh Kumar[4] · Ihtiram Raza Khan[5] · Pankaj Kumar[6]

Received: 12 October 2021 / Revised: 30 October 2021 / Accepted: 3 November 2021

Abstract Blockchain is a system of recording information in a way that makes it difficult or impossible to change, hack, or cheat the system. A blockchain is essentially a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems on the blockchain. Blockchain is an especially promising and revolutionary technology because it helps reduce risk, stamps out fraud, and brings transparency in a scalable way for myriad uses. Blockchain offers decentralised credentials, making it a global system. It also promotes user-to-user data transmission. However, blockchain-based identity management is crucial for data security. The user owns the identification. Blockchain-based IDMs solve conventional IDM flaws. As technology advances, so does the need for user data protection. Advance payment to safeguard data is unacceptable. So, identity management solutions assist overcome the flaws. Blockchain consists of three important concepts: blocks, nodes, and miners. Blockchain makes life simpler by changing the way personal data is stored are made available. While the government and other third parties are addressing the problem, the users are still exposed to identity management. Personal identity management takes increasing security and productivity. Blockchain eliminates the third party by transferring data between two entities. This work develops a categorization model for Ethereum addresses based on a few criteria. Voting categorization outperforms the other models with an accuracy of 88.89% and an AUC Score of 97%.

Keywords Blockchain · Distributed Technology · Identity Management · Distributed Ledger Technology · Decentralised Database

1 Introduction

A block chain is essentially a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems on the blockchain. Each block in the chain contains a number of transactions, and every time a new transaction occurs on the blockchain, a record of that transaction is added to every participant's ledger. The

✉ Sumegh Tharewal
sumeghtharewal@gmail.com

A. Shobanadevi
shobanak07@gmail.com

Mukesh Soni
mukeshsoni@ieee.org

D. Dinesh Kumar
dineshkumard@stjosephs.ac.in

Ihtiram Raza Khan
iraza@jamiahamdara.ac.in

Pankaj Kumar
unpankaj@gmail.com

[1] Department of Data Science and Business Systems, SRM Institute of Science and Technology, Chennai, India

[2] Department of Computer Application, Manipal University Jaipur, Jaipur, India

[3] Seniro IEEE Member, Bhopal, India

[4] Department of Electronics and Instrumentation Engineering, St. Joseph's College of Engineering, Chennai, Tamil Nadu, India

[5] Academician, Jamia Hamdard, Delhi, India

[6] Lloyd Institute of Engineering & Technology, Greater Noida, Uttar Pradesh, India

Published online: 27 November 2021

decentralised database managed by multiple participants is known as Distributed Ledger Technology (DLT). Blockchain technology is most simply defined as a decentralized, distributed ledger that records the provenance of a digital asset. By inherent design, the data on a blockchain is unable to be modified, which makes it a legitimate disruptor for industries like payments, cybersecurity and healthcare (Stephen and Alex 2018). Cryptocurrencies such as Bitcoin, Ethereum, Doge are performing their transactions through their own cryptocurrencies that are trustworthy due to the advantages of blockchain; it provides a better solution for the authentication process. As of today, we are still facing many challenges and difficulties in maintaining users' identity and authenticating it and is prone to be theft by attackers. Authenticating is the process of checking whether the user is the owner of the sensitive data which he/she has provided. It influences trust among service providers and users. Personal identity and passwords are stored centralized servers, which are prone to attackers. Therefore, it is required to provide security. Hashing in blockchain makes it easy to secure data. Blockchain uses distributed ledger technology and provides proper identity management.

2 Literature review

Blockchain provides more security in transactions. Decentralization provides more flexibility than centralized application. It is peer to peer connected. Blockchain provides more scalability. Ledger records each and every transaction in a blockchain. A ledger is assigned to every user which is a decentralized application. Public and private Blockchain provides more security.

2.1 Blockchain based Identity management

Self-sovereign identity allows the control over the user's identity. Identity management system consists of user, service provider and ID provider as components. The Merkle tree contains all transactions in a block. Self-sovereign identity that is decentralized technology. uPort is of self-sovereign identity which depends on Ethereum. Shocard is digital identity card and authentication platform built on blockchain (Liu et al. 2020).

2.2 Sora identity

Uses blockchain technology to create a protocol to store personal information which are encrypted. The main components are the User, Mobile Device, a central server, and a blockchain platform. Users of the system have different identifier—DID. Private data is stored on server in an encrypted form which need not to be decrypted by the server. Hashing process in blockchain uses SHA-35 as algorithm (Sharma et al. 2021).

2.3 Identity management using permission blockchain

Identity management using permissioned blockchain. Permissioned Blockchain: It make identity management possible only among registered entities and authorized users or entities. Certificate is provided by authorized agency (government, organizations etc.).

2.4 First look at identity management

DLT ensures trust and integrity of the transactions that contains much more benefits of underlying DLT in IDM (Prasanalakshmi et al. 2011). The user is provided with a ownership service that performs proofing of users identity (Takemiya and Vanieiev 2018). DLT has identity schemes such as uport, shocard, sovrin.

3 Technology

3.1 Identity management (IDM)

IDM plays an important role in data security; IDM refers to the building block of technologies and policy to ensure that only authorized users can access the resources. The main components of Identity management, as shown in the Fig. 1 are user, Service Provider ID provider.

3.1.1 User

As the name suggests users are end users who take the services provides by the service provider. All the users

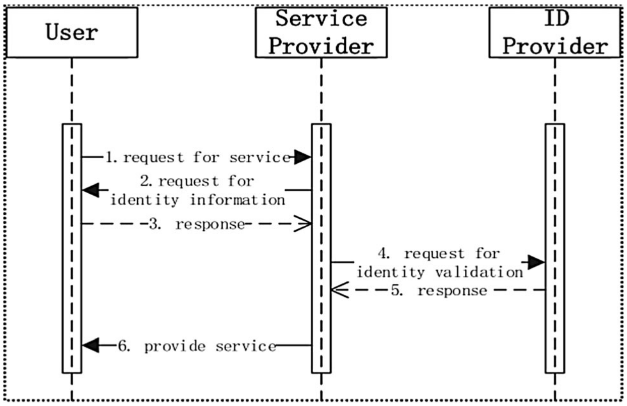


Fig. 1 A typical operation of an identity management system

don't have the same privilege. The user must request for an identity from ID provider (Gururaj 2020).

3.1.2 *Service provider*

Service Providers are connected to user and ID provider which provides services. Service provider tells id manger to check users. The ID provider provides the authentication results, service provider provides service based on the authentication result (Chakraborty et al. 2021).

3.1.3 *ID provider*

This is the core of the system which provides services such as authentication, registration and management.

3.1.4 *Independent IDM*

Here the service provider owns the identity. It is not possible to exchange identities of different service providers (MukeshSoni and Singh 2021).

3.1.5 *Centralized IDM*

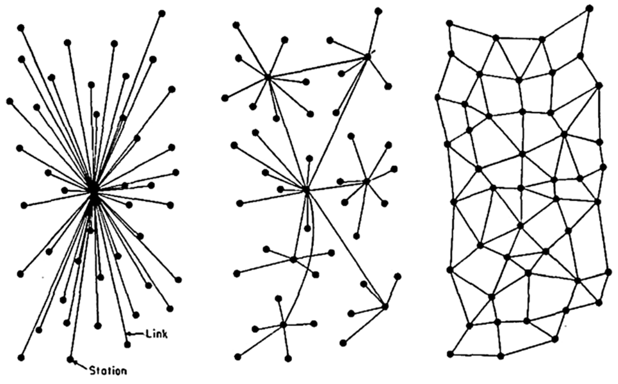
It has one identity provider and identifier. Identity management happens in one environment. The identifier should be cautiously chosen and unique identity should have distinctive qualities. It has only one identity provider and identifier

3.1.6 *Federated IDM*

It is a trusted domain and has many identity providers. A recognized domain has many service providers which distinguishes identity from service providers (Bharti et al. 2021).

3.1.7 *Decentralization*

Decentralization plays a prominent role in blockchain which provides more security compared to centralized network as shown in Figure 2. Centralized networks are more exposed to attackers as they are in a fixed position (Zaman et al. 2018). Decentralization transfers the control to managers. The networks are made up of computers which interact on the basis of peer to peer without the need for third party. Rather than following the instructions of central authority the nodes have common rules. The disadvantage is that if as scalability increases it can lead to reduced stability (Gomathi et al. 2021; Sanober et al. 2021).



(a) Centralized. **(b)** Decentralized. **(c)** Distributed networks.

Fig. 2 Network comparison

3.1.8 *Hashing*

A technique which maps hashes and values with hash function into the hash table. All the value pairs are accessed through index which is created by hash table. In blockchain previous hash value is connected to present block to provide more security (Anurag Shrivastava et al. 2021). If an attacker changes a hash value it breaks the chain of block. A blockchain contains data and hash pointer of the previous block of transactions (HedgeTrade, 2021). It is the hashing that make decentralized more secure. For a given input value the obtained output value should be same. i.e., deterministic. Shown in Figure 3.

Every block contains list of transactions which are linked to each other cryptographically as list. The transactions are collected by miners (Prasanalakshmi and Kannammal 2013) (Figure 4–5).

3.1.9 *Merkle tree*

Merkle tree also called as hash tree is a data structure shown in figure 6. Merkle tree encodes the blockchain data more effectively and securely. The Merkle tree contains all transactions in a block. The merkle tree root nodes leaf nodes and intermediate nodes. It flows back propagation i.e.; the value of the internal node is computed by calculate the hash of child nodes. Hence the root nodes represent all the transactions. In blockchain transactions are run through an algorithm to generate a hash (Soni and Singh 2021a; Tang and Shabaz 2021). It contains a group of numbers and letters that are in turn used to verify with the original data which are transactions, but original set of transactions are not obtained. Every transaction is hashed. One hash from the whole network is obtained and each pair of transactions are concatenated. As mentioned in the figure 4, T represents transactions, 'H' a hash.

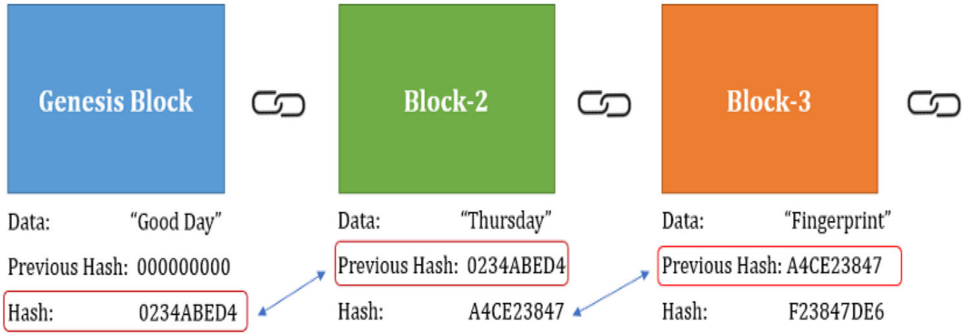


Fig. 3 Hashing process

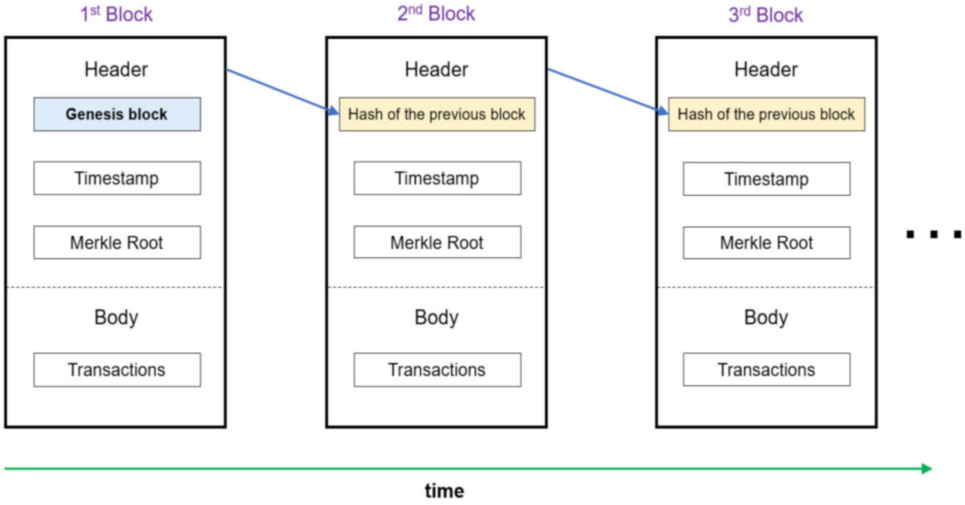


Fig. 4 A genesis block pointed by hash pointer

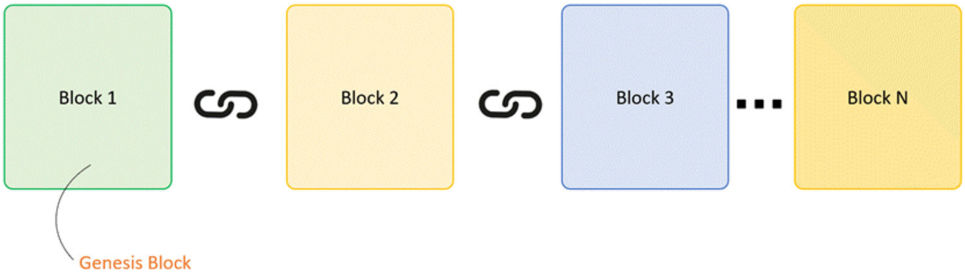


Fig. 5 Sample of a single block

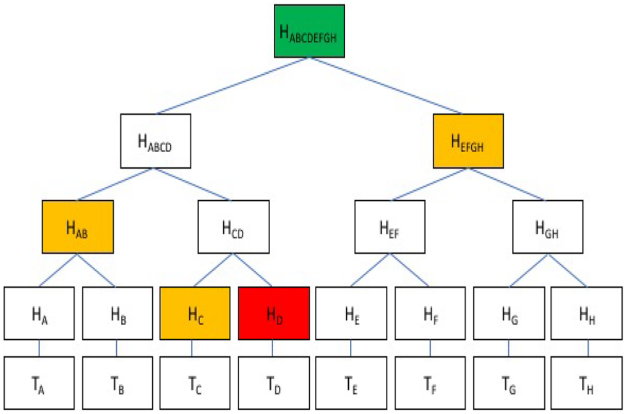


Fig. 6 An overview of Merkle tree

4 Blockchain-based identity management systems

4.1 Uport

Uport is a secure system for self-sovereign identity shown in figure 7. It used smart contract-based identities which provide sovereign identity to organizations users and others, it is used for digital representation of a user. The addressing used is 160-bit hexadecimal identifiers in smart contracts. They are run by Ethereum Virtual Machine. There are two uport identities designed by creators: controller and proxy. A new controller is initiated which stores a reference to public key. The mobile app creates asymmetric key pair. A new proxy contains to immediate created controller contract, functions are invoked only by the

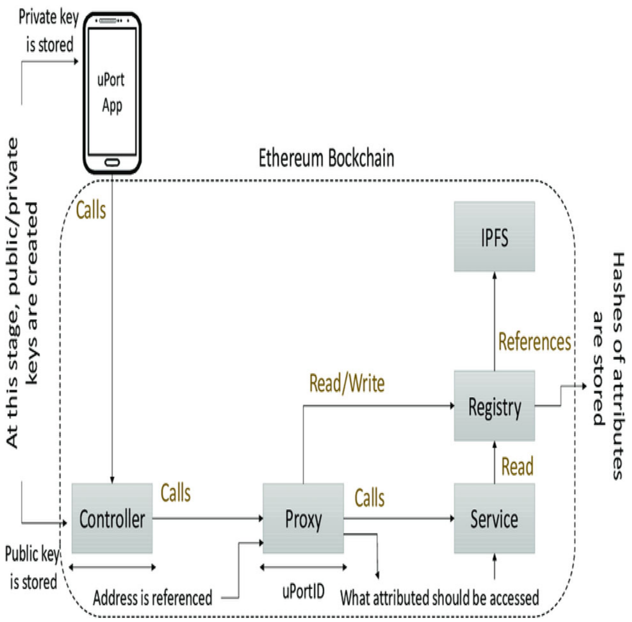


Fig. 7 Uport architecture

controller (Soni and Singh 2021b). A private key tagged with the uport id is stored with the user's mobile device. Uport doesn't have a central server to authenticate. The uport is compromised permanently if an attacker exploits the uport application and replaces the trustees via controller (MukeshSoni and Gomathi 2020).

4.2 Sovrin

Sovrin is an open source identity which is built on distributed ledger technology that stores record of identity shown in Figure 8. It is public and is trusted by institutions such as stewards (ex. banks, governments, universities etc). Sovrin gives access users to create many ids'. The users are in charge of sovrin or an appointed guardian. The sovrin ledger that has transactions linked to a particular identifier. Users communicate to sovrin using a mobile application. Agents are the end points that are accessible and addressable. The mobile app helps the user to manage the cryptographic key. Sovrin provides full control of all the aspects of their identity Although the transactions could be stored on ledgers but mobile is more secured as it is not exposed to third parties (Ahmad et al. 2021; Soni et al. 2021).

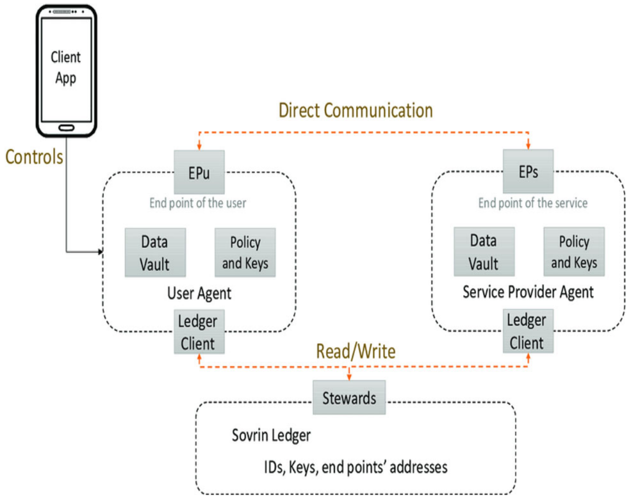


Fig. 8 Sovrin architecture

4.3 ShoCard

Shocard is a digital identity that binds identifier a trusted credential such as driving license, passport etc - together via cryptographic hashes (Ratta et al. 2021). Its use is verification of identity in face-to-face online interactions shown in Figure 9. Server act as an essential part of its scheme; server is intermediate between a user and a depended party to exchange information (Duraisamy et al. 2019) (Figure 10).

The shocard's central server function which acts as intermediate is to manage the distribution of encrypted messages between shocard and encrypted users. ShoCard is risk free when stored and distributed as plain text. Storage of information and sharing with parties is user specific i.e., sharing controlled by the end-users. Users don't have access with cryptographic key management which worries the overall deploy ability of ShoCard. On an average it takes 10 min to mine bitcoin and six (additional) blocks to be mined is advised. Bootstrap occurs at the creation of a new shocard and symmetric key pair for the user (Yuvaraj et al. 2021). Once bootstrapped. In certification process the user can interact with the service provider that rely on seal. Although researchers are proposing various protocols schemes for providing confidentiality, integrity among the critical information among the clients and server in the shared network (Kannan et al. 2021) (Figure 11).

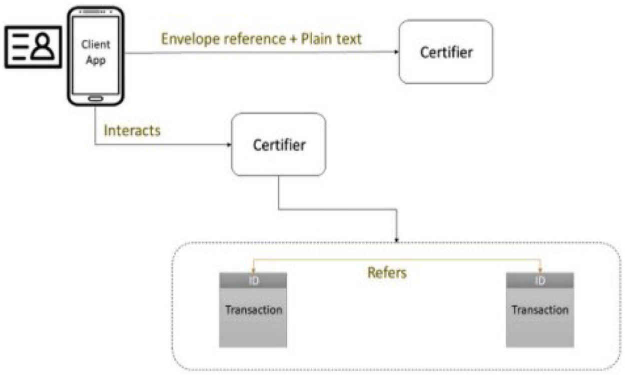


Fig. 9 Sho card architecture

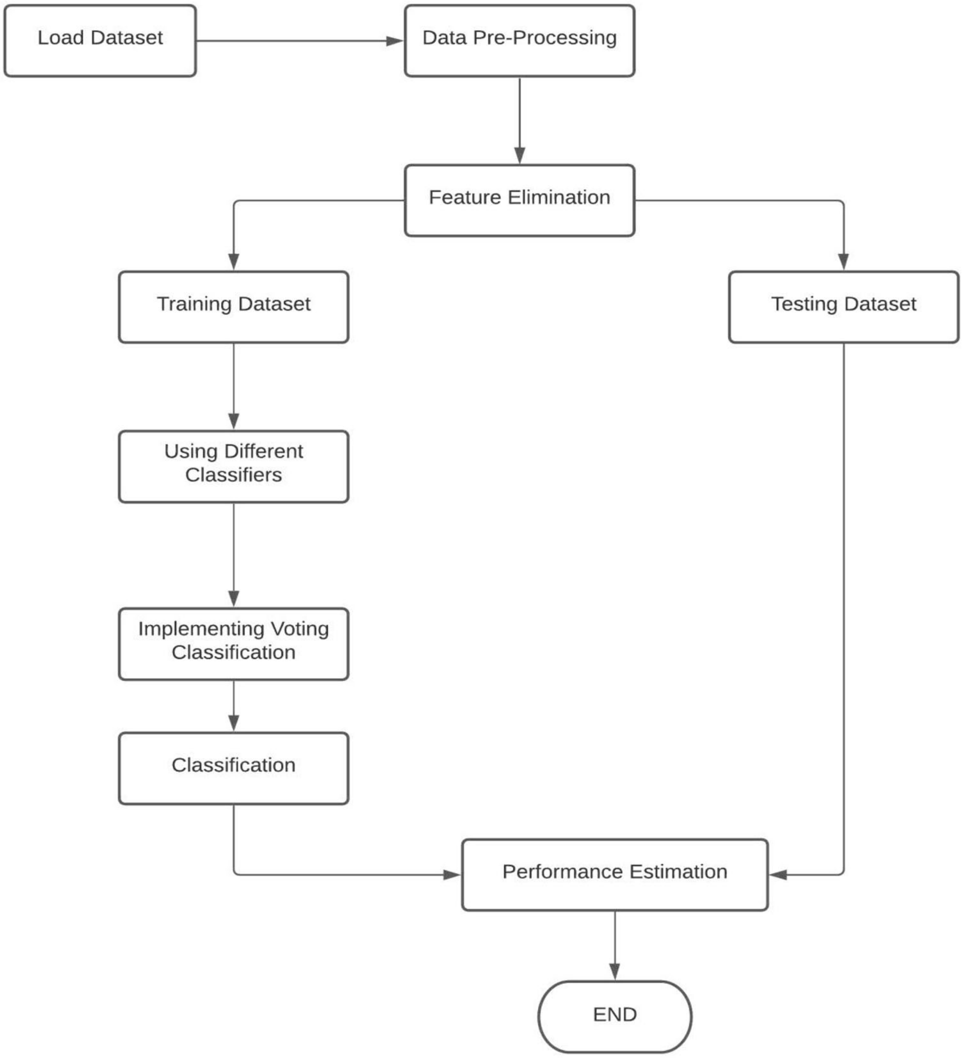


Fig. 10 Flow Diagram for Ethereum Address Classification

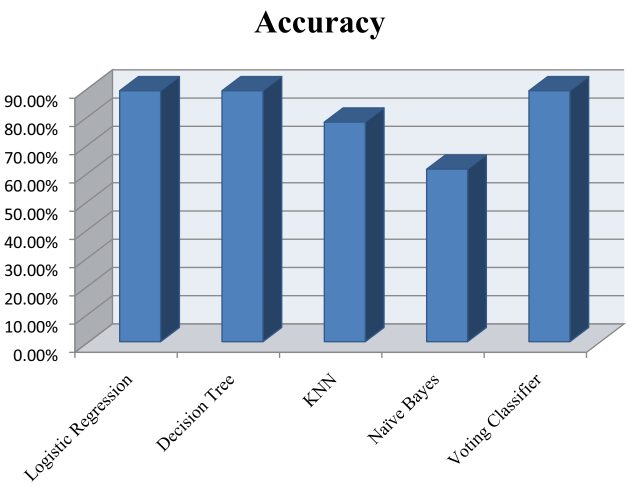


Fig. 11 Accuracy for Ethereum Address Classification using ML

5 Applications

5.1 Real-time data analysis

Real time data analysis is used for making decisions quickly which is an added advantage for any organizations with assets in data. It is one of the most effective ways of safeguarding the against theft in data driven industry. It helps to reduce areas of vulnerability (Jairath et al. 2021) (Figure 12).

5.2 Data sharing

Agencies such as governments and organizations collect large data on individuals and companies. It is stored in department and agencies. Different sectors won't have complete access to it in public sectors, different

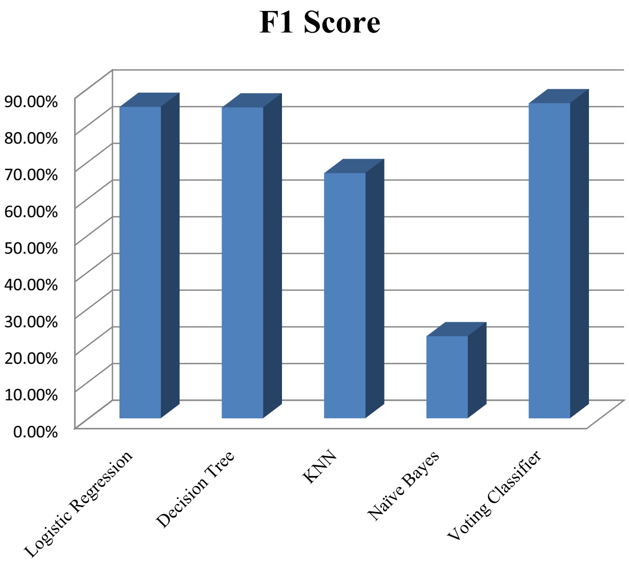


Fig. 12 F1-Score for Ethereum Address Classification using ML

departments don't have access to information on individuals as they don't have required permission. This result in more amount of time spent to trace the information from government bodies (Mushtaq and Rizwan 2018) (Figure 13).

With the advent of cryptocurrencies, the digital payment system has improved a lot over time, posing many advantages such as no dependence on central authorities to process the transaction, minimum transaction fees, and many more as such. Besides, cryptocurrencies are also used for many malicious activities, such as asking for a ransom amount in the form of virtual currencies, asking to transfer funds in the name of fake donations, etc. Using Machine learning algorithms, addresses are classified as Legit or Dodgy based on few features like the type of the account, total transactions, outgoing and incoming balances, total records in the Ethereum blockchain (Al-Turkistani and AlSa'awi 2020) (Figure 14).

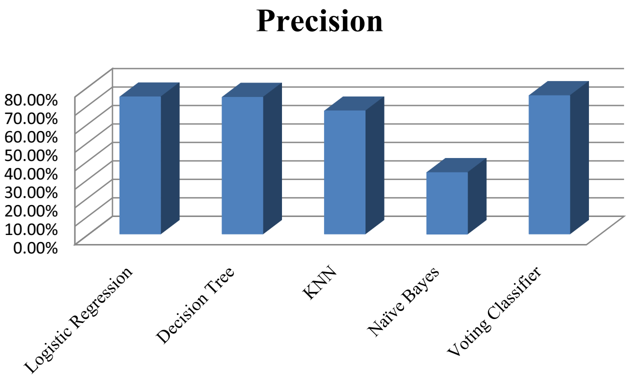


Fig. 13 Precision for Ethereum Address Classification using ML

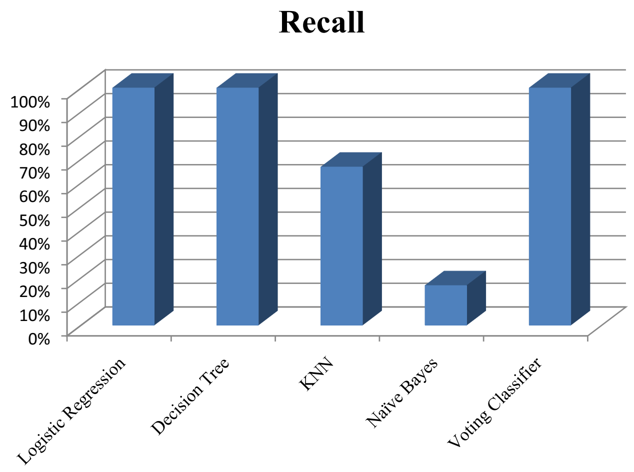


Fig. 14 Recall for Ethereum Address Classification using ML

Using various machine learning algorithms, the binary classification was performed and then the performances of the model were compared based on the following factors: (i) Accuracy Score (shown in Fig. 11), (ii) Recall (shown in Fig. 14) (ii) Precision (shown in Fig. 13) (iv) F1 Score (shown in Fig. 12), and (v) AUC Score (shown in Fig. 15). In Table 1, the comparison of different metrics is displayed (Nagothu et al. 2018). (Figure 16).

Based on accuracy score, the Logistic Regression, Decision Tree, and Voting Classification perform better than the other classifiers for which the AUC Score is considering for model selection and Voting Classifier (VC) is selected for the Binary Classification. In Fig. 3, it is shown that VC will perform better than the rest of the models (Thuraisingham 2020).

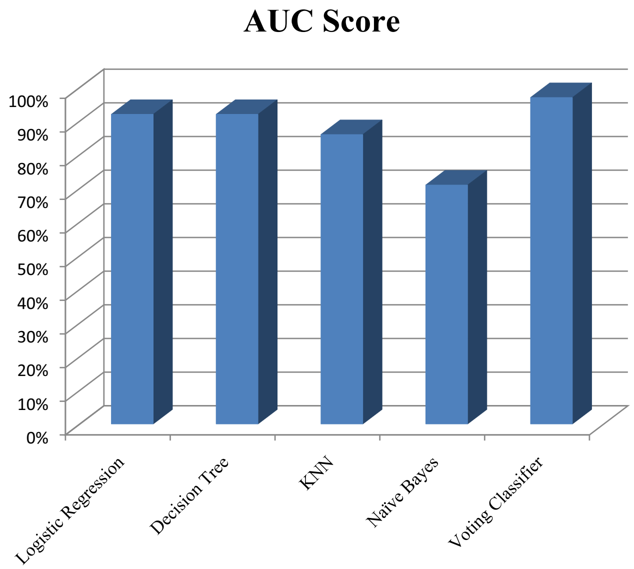


Fig. 15 AUC Score for Ethereum Address Classification using ML

Table 1 Performance Metrics Comparison of ML models

fMetrics	Logistic regression	Decision tree	KNN	Naïve bayes	Voting classifier
Accuracy	88.89%	88.89%	77.78%	61.11%	88.89%
F1 Score	84.71%	84.55%	66.67%	22.22%	85.71%
Precision	74.29%	74.11%	66.67%	33.33%	75.0%
Recall	100%	100%	66.67%	16.67%	100.0%
AUC score	92%	92%	86%	71%	97%

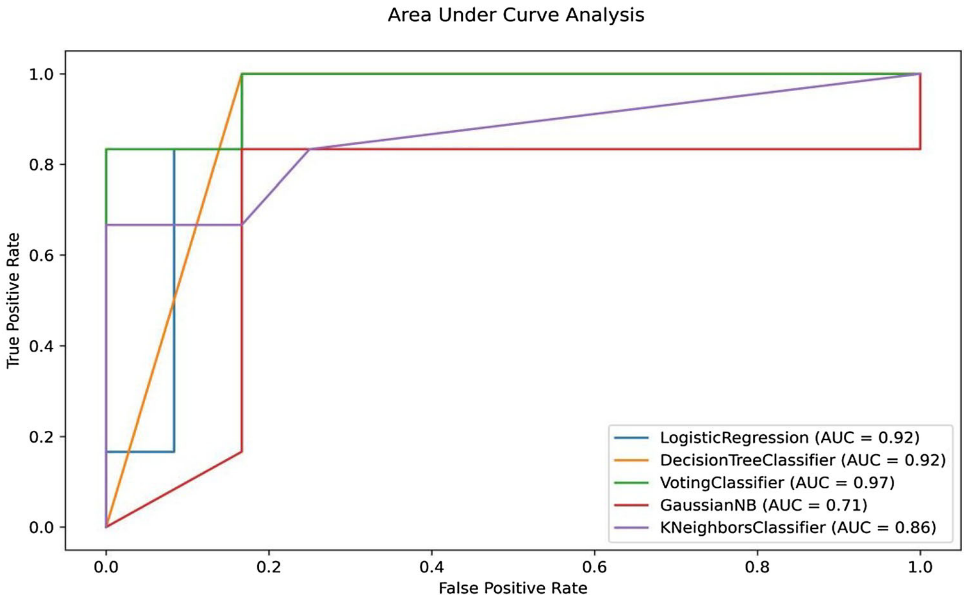


Fig. 16 Comparison of AUC Scores

As the trend of cryptocurrencies grows exponentially, the new digital payment gets a huge boost among people for buying and selling goods, trading the crypto assets, and in many other ways, but the disadvantage of tracking down the transactions gives a head start for all the scammers to receive payments in the form of cryptocurrency (Liang et al. 2018; Chakraborty and Dr. Subhendu Khan, Mohammad NOORWALI, MOIRANGTHEM Kaur, Manpreet Gupta, Shikha. 2021). Several phishing sites are created to mimic crypto exchanges, and then, in turn, the people lose their crypto assets (Kumar et al. 2021). In this study, a classification model is designed which will classify an Ethereum address on the basis of few parameters. On further analysis, it is found that voting classification performs better than the other models, which give an accuracy of 88.89% and an AUC Score of 97%.

6 Conclusion

Blockchain provides decentralization of credentials which makes a universally acceptable protocol. It also facilitates peer-peer exchange of data among users. On the other hand, identity management which uses blockchain plays a prominent role in data security. The user is the owner of the identity. In addition to buying and selling items using cryptocurrencies, individuals are also exchanging crypto assets, but the inability to trace transactions offers a head start to fraudsters who want to obtain bitcoin payments. People lose their crypto holdings due to phishing sites built to seem like crypto exchanges. This work develops a categorization model for Ethereum addresses based on a few criteria. Voting categorization outperforms the other models with an accuracy of 88.89% and an AUC Score of 97%. Blockchain-based IDMs solve conventional IDM flaws. As technology advances, so does the need for user data protection. Advance payment to safeguard data is unacceptable. So, in future identity management solutions assist overcome the flaws.

Funding This research work is self-funded.

Declarations

Conflict of interest The authors declare that they have no conflict of interest and all ethical issues including human or animal participation has been done. No such consent is applicable.

References

Ahmad I, Serbaya SH, Rizwan A, Mehmood MS (2021) spectroscopic analysis for harnessing the quality and potential of gemstones for small and medium-sized enterprises (SMEs). J Spectrosc. https://doi.org/10.1155/2021/6629640

Al-Turkistani H F, AlSa'awi NK (2020) Poster: Combination of blockchains to secure smart home internet of things. In: 2020 1st international conference of smart systems and emerging technologies (SMARTTECH) pp 261–262, doi: https://doi.org/10.1109/SMART-TECH49988.2020.00069

Anurag Shrivastava K, Krishna M, Rinawa M L, MukeshSoni GR, Jaiswal S (2021) Inclusion of IoT, ML, and blockchain technologies in next generation industry 4. 0 Environment. Mater Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.07.273

Barot Y, Soni M, Gomathi S (2020) A review on privacy-preserving data preprocessing. J Cybersecur Inform Manag 4(2):16–30

Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P (2021) Prediction of heart disease using a combination of machine learning and deep learning. Comput Intell Neurosci. https://doi.org/10.1155/2021/8387680

Chakraborty C, Sarkar C, Sinha D. (2021) Design of a priority based local energy market using blockchain technology

Duraisamy S, Pugalendhi GK, Balaji P (2019) Reducing energy consumption of wireless sensor networks using rules and extreme learning machine algorithm. J Eng 2019(9):5443–5448

Gomathi S, Soni M, Dhiman G, Govindaraj R, Kumar P (2021) A survey on applications and security issues of blockchain technology in business sectors. Mater Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.02.088

Gururaj, Prabhanjan. (2020) Identity management using permissioned blockchain. In 2020 International conference on mainstreaming block chain implementation (ICOMBI), IEEE, pp 1–3

Jairath K, Singh N, Jagota V, Shabaz M (2021) Compact ultrawide band metamaterial-inspired split ring resonator structure loaded band notched antenna. Math Probl Eng. https://doi.org/10.1155/2021/5174455

Kannan S, Dhiman G, Natarajan Y, Sharma A, Mohanty SN, Soni M, Easwaran U, Ghorbani H, Asheralieva A, Gheisari M (2021) Ubiquitous vehicular Ad-Hoc network computing using deep neural network with Iot-based bat agents for traffic management. Electronics 10:785. https://doi.org/10.3390/electronics10070785

Kaur M, Khan MZ, Gupta S, Noorwali A, Chakraborty C, Pani SK (2021). MBCP performance analysis of large scale mainstream blockchain consensus protocols.

Kumar A, Abhishek K, Bhushan B, Chakraborty C (2021) Secure access control for manufacturing sector with application of ethereum blockchain. Peer-to-Peer Netw Appl 14(5):3058–3074

Liang X, Shetty S, Tosh D, (2018) Exploring the attack surfaces in blockchain enabled smart cities In: *2018* IEEE international smart cities conference *(ISC2)*, pp 1–8, doi: https://doi.org/10.1109/ISC2.2018.8656852.

Liu Y, He D, Obaidat M S, Kumar N, Khan M K, Choo K-K R (2020) Blockchain-based identity management systems: a review. J Netw Comput Appl 166:102731

Mushtaq I, Rizwan A (2018) Obstacles to knowledge sharing in engineering organisations: a quantitative approach. Int J Knowl Manag Studies 9(3):293. https://doi.org/10.1504/ijkms.2018.094216

Nagothu D, Xu R, Nikouei S Y, Chen Y, (2018) A microservice-enabled architecture for smart surveillance using blockchain technolog. In: 2018 IEEE international smart cities conference (ISC2), pp 1–4, doi: https://doi.org/10.1109/ISC2.2018.8656968.

Prasanalakshmi B, Kannammal A (2013). ECC based biometric encryption of compressed image for security over network channels. In. Springer India, (pp 343–351), https://doi.org/10.1007/978-81-322-1000-9_33

Prasanalakshmi B, Kannammal A, &Sridevi R. (2011). Frequency domain combination for preserving data in space specified token with high security. In informatics engineering and information science, Springer Berlin Heidelberg. (pp 319–330), https://doi.org/10.1007/978-3-642-25327-0_28

Ratta P, Kaur A, Sharma S, Shabaz M, Dhiman G (2021) Application of blockchain and internet of things in healthcare and medical sector: applications, challenges, and future perspectives. J Food Qual. https://doi.org/10.1155/2021/7608296

Sanober S, Alam I, Pande S, Arslan F, Rane KP, Singh BK, Khamparia A, Shabaz M (2021) An enhanced secure deep learning algorithm for fraud detection in wireless communication. Wirel Commun Mob Comput. https://doi.org/10.1155/2021/6079582

Sharma C, Amandeep B, Sobti R, Lohani TK, Shabaz M (2021) A secured frame selection based video watermarking technique to address quality loss of data: combining graph based transform, singular valued decomposition, and hyperchaotic encryption. Secur Commun Netw. https://doi.org/10.1155/2021/5536170

Soni M, Singh DK (2021) Blockchain-based security & privacy for biomedical and healthcare information exchange systems. Mater Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.02.094

Soni M, Singh DK (2021b) LAKA: Lightweight authentication and key agreement protocol for internet of things based wireless body area network. Wirel Person Commun. https://doi.org/10.1007/s11277-021-08565-2

Soni M, Dhiman G, Rajput BS et al (2021) Energy-effective and secure data transfer scheme for mobile nodes in smart city applications. Wireless PersCommun. https://doi.org/10.1007/s11277-021-08767-8

M. Soni and D. K. Singh, (2021) Blockchain implementation for privacy preserving and securing the healthcare data. In: 2021 10th ieee international conference on communication systems and network technologies (CSNT), pp 729–734, doi: https://doi.org/10.1109/CSNT51715.2021.9509722

Stephen R, Alex A (2018) A review on blockchain security. IOP Conf series: Mater Sci Eng 396(1):012030

M Takemiya, B Vanieiev (2018) Sora identity: Secure, digital identity on the blockchain. In 2018 IEEE 42nd annual computer software and applications conference (compsac) IEE, 2: 582-587

Tang S, Shabaz M (2021) A new face image recognition algorithm based on cerebellum-basal ganglia mechanism. J Healthcare Eng. https://doi.org/10.1155/2021/3688881

Thuraisingham B, (2020) blockchain technologies and their applications in data science and cyber security.In: 2020 3rd international conference on smart blockchain (SmartBlock), pp 1–4, doi: https://doi.org/10.1109/SmartBlock52591.2020.00008.

Yuvaraj N, Srihari K, Dhiman G, Somasundaram K, Sharma A, Rajeskannan S, Soni M, Gaba GS, AlZain MA, Masud M (2021)

(2021) Nature-inspired-based approach for automated cyberbullying classification on multimedia social networking. Math Probl Eng 12:6644652. https://doi.org/10.1155/2021/6644652

Zaman S, Chakraborty C, Mehajabin N, Mamun-Or-Rashid M, Razzaque MA (2018). A deep learning based device authentication scheme using channel state information. In 2018 international conference on innovation in engineering and technology (ICIET), IEEE, (pp 1–5)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.