Empowering Self-Sovereign Digital Identities: A Blockchain and AI-Driven Approach for Enhanced Privacy and Security

1st Dr. A. Karunamurthy, 2rd S. Kaushik
[1]Associate Professor, Department of computer Applications, Sri Manakula Vinayagar Engineering college (Autonomous)Puducherry 605107, India
Karunamurthy26@gmail.com
[2]Post Graduate student, Department of computer Applications, Sri Manakula Vinayagar Engineering college (Autonomous)Puducherry 605107, India
sakthikaushik082@gmail.com

ABSTRACT:

In today's hyper-digital landscape, secure and private digital identity management is paramount. With increasing reliance on online platforms for services, transactions, and communications, individuals require robust systems that offer privacy, control, and trust. Traditional centralized and federated identity systems—like those offered by major tech companies—have become inadequate, exposing users to significant risks such as data breaches, identity theft, and surveillance. These systems fail to provide users with autonomy over their personal information, as identities are controlled and stored by third-party providers.

To overcome these limitations, this project proposes an advanced self-sovereign digital identity (SSI) framework that integrates decentralized blockchain architecture with machine learning- driven trust evaluation. By leveraging blockchain's immutable and tamper-resistant ledger, digital identities are stored securely and directly under user control. A Unique Personal Identifier (UPI) is used to authenticate users, eliminating the need for centralized identity providers. To further enhance security, a Logistic Regression model is utilized to assess the trustworthiness of service providers. In scenarios where trust is not established, the system generates masked credentials using a Lookup Substitution Algorithm, ensuring that user privacy is preserved during verification processes.

This blockchain and AI-powered approach provides a user-centric, transparent, and secure alternative to traditional identity systems—empowering individuals to take ownership of their digital identities while mitigating the risks associated with data misuse and centralized control.

Keywords:
Self-Sovereign Identity, Blockchain Identity Management, Privacy-Preserving Authentication, Trust Prediction, Machine Learning, Logistic Regression, Data Security, Identity Decentralization, Cryptographic Identity, Digital Privacy, Decentralized Access Control, SSI, UPI, Credential Masking, Cybersecurity, Lookup Substitution Algorithm.

1. INTRODUCTION

The rise of digital services, remote access, and interconnected platforms has fundamentally reshaped how individuals interact with online systems and share personal data. The widespread adoption of mobile devices, cloud services, and AI-enabled applications has enabled seamless digital experiences but has also introduced significant privacy and security vulnerabilities. Centralized identity management systems—often operated by tech conglomerates—pose major risks, including unauthorized data access, surveillance, misuse, and large-scale breaches of personal information. These systems offer limited user control and have failed to provide robust protection against identity theft and data exploitation.

In response to these growing challenges, Self-Sovereign Identity (SSI) has emerged as a transformative solution, granting users full control over their digital identities through decentralized technologies. By leveraging blockchain, digital credentials can be stored in a tamper-proof, transparent, and secure manner—free from the constraints of third-party identity providers. When combined with machine learning models, such as Logistic Regression, the system can dynamically assess the trustworthiness of service providers before allowing access to sensitive identity data. This duallayered approach ensures that identity verification is both privacy-respecting and context-aware.

Furthermore, this decentralized identity framework aligns with regulatory standards such as the General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), and Payment Card Industry Data Security Standard (PCI-DSS) by minimizing data exposure and enforcing user-centric control. Through features like fine-grained verification, credential masking, and trust scoring, this approach enhances security and privacy in digital identity interactions, ultimately empowering users in an increasingly connected digital world.

2. LITERATURE SURVEY

Self-sovereign identity (SSI) has gained significant traction as a secure and privacy-focused alternative to conventional digital identity frameworks. The growing concern over centralized identity management systems—where user data is stored and controlled by third-party providers—has led researchers to explore decentralized approaches. Several studies highlight the limitations of federated identity systems in terms of data breaches, privacy violations, and lack of user control (Allen, 2016; Tobin & Reed, 2017).

Blockchain technology has emerged as a powerful enabler of decentralized identity due to its tamper-resistant, transparent, and distributed ledger structure. Projects like Sovrin, uPort, and Microsoft's ION have demonstrated how decentralized identifiers (DIDs) and verifiable credentials can be managed independently by users while maintaining interoperability and security (Preukschat & Reed, 2021). However, one of the persistent challenges in the domain is the integration of trust verification in a decentralized environment. Most blockchain-based identity systems lack real-time mechanisms to assess the credibility of service providers or requesting entities.

To address this gap, recent research has investigated the incorporation of machine learning algorithms, particularly supervised learning models like Logistic Regression, to enhance trust prediction in identity verification systems (Kumar et al., 2020; Singh & Aggarwal, 2021). These models use historical data and behavioral indicators to assess the trustworthiness of third-party service providers, reducing the chances of identity misuse or credential fraud.

In conclusion, the literature reveals a growing body of work focused on blockchain-based identity solutions and trust prediction using machine learning. While these technologies offer a promising foundation for self-sovereign identity, there remains a need for unified frameworks that combine decentralized storage, cryptographic authentication, and intelligent trust evaluation to build a truly secure and user-centric identity management system.

3. METHODOLOGY

PROPOSED SYSTEM:

The proposed system introduces a robust, blockchain-powered Self-Sovereign Identity (SSI) framework that empowers users with full ownership and control over their digital identities. Unlike traditional identity systems reliant on centralized authorities or third-party providers, this model eliminates single points of failure and data vulnerability by distributing identity verification and storage across a blockchain network.

At the core of this architecture lies a Decentralized Identity Vault, where users securely store their verifiable credentials. These credentials are associated with cryptographic key pairs, ensuring that only the rightful identity owner can access and share their data. Each interaction with a service provider is authenticated using a Unique Personal Identifier (UPI), which references the required credentials without exposing the underlying sensitive information.

To address the challenge of trust in decentralized systems, the platform incorporates a Machine Learning Trust Assessment Module. A Logistic Regression algorithm is trained on historical service provider behavior, credential usage patterns, and community feedback to predict whether the requesting service provider is trustworthy. This prediction score directly influences how user credentials are shared.

In scenarios where the trust score falls below a secure threshold, the system activates a Lookup Substitution Algorithm, which generates masked credentials—limited versions of the original credentials that preserve anonymity while maintaining the necessary level of verification.

The methodology of the proposed system follows this operational flow:

- User Registration and Identity Generation: The user registers and is issued a decentralized identifier (DID), paired with cryptographic keys.
- Secure Identity Storage on Blockchain: The user's identity metadata is recorded immutably on the blockchain.
- Service Provider Trust Evaluation: When a service requests identity verification, the Logistic Regression model predicts its trustworthiness in real time.
- Credential Response Generation:
- If the provider is trusted, full credentials are shared.
- If untrusted, a masked version is generated using the Lookup Substitution Algorithm.
- Access Granted Based on Context: The service provider receives the credentials for authentication without the system compromising user privacy or revealing sensitive data.

By decentralizing identity control and intelligently regulating access based on provider trustworthiness, the proposed system offers a scalable, privacy-preserving, and user-centric alternative to conventional identity verification systems.

4. IMPLEMENTATION

The implementation of the proposed self-sovereign identity (SSI) system involves the seamless integration of blockchain technology, machine learning algorithms, and cryptographic protocols to enable secure, decentralized, and intelligent identity management.

At the foundation, a Hyperledger Indy or Ethereum-based blockchain network is deployed to register decentralized identifiers (DIDs) and maintain immutable records of credential issuances. Users generate their identities via a Decentralized Identity Wallet, which securely stores verifiable credentials locally while referencing identity metadata on-chain.

To authenticate identity requests, the system uses Zero-Knowledge Proofs (ZKPs), allowing users to verify credentials without revealing the underlying data. For example, a user can prove they are over 18 or possess a valid driving license without exposing their actual date of birth or license number. This ensures privacy-by-design.

The Machine Learning Trust Assessment Module, implemented using Python's scikit-learn, employs a Logistic Regression model trained on datasets comprising service provider history, credential usage patterns, and peer trust ratings. When a service provider requests access, this module calculates a real-time trust score.

Based on this trust score, the Lookup Substitution Algorithm dynamically adjusts the granularity of data shared:

- High-trust providers receive full verifiable credentials.
- Low-trust providers receive masked credentials that satisfy minimum verification requirements without revealing sensitive identity elements.

Integration with external services is achieved through DIDComm protocols that enable secure and interoperable communication between identity holders, issuers, and verifiers.

To ensure usability and adoption, the system provides:

- A Web Dashboard for users to manage credentials, consent, and view request histories.
- A Developer API for third-party services to integrate identity verification mechanisms.
- Mobile app support using Flutter or React Native for cross-platform credential management.

This layered, modular implementation ensures secure, user-controlled identity management, bridging the gap between privacy and trust in digital interactions.

5. ARCHITECTECTURE DIAGRAM

The architectural design of the proposed system focuses on a decentralized, scalable, and secure approach to managing self-sovereign digital identities. The system follows a layered architecture, encompassing the presentation layer, application layer, blockchain layer, and data layer, each playing a critical role in ensuring privacy, trust, and secure communication. At the core of this architecture is a client-server model in which user devices communicate securely with backend services through encrypted channels, facilitating operations such as identity creation, credential management, and trust evaluation.

The presentation layer comprises a user interface delivered through a mobile or web application. This interface enables users to create, store, and share their digital identities using decentralized identifiers (DIDs) and verifiable credentials. Users can configure privacy preferences, view activity logs, and check trust scores of verifiers before consenting to share their data. All interactions from this layer are securely transmitted to the backend using decentralized identity communication protocols such as DIDComm.

The application layer serves as the central logic unit of the system. It manages processes such as credential issuance, selective disclosure using zero-knowledge proofs, and dynamic trust evaluation of third-party verifiers. This trust evaluation is powered by an artificial intelligence engine, which analyzes behavioral indicators and historical interactions to generate trust scores. These scores influence the level of data disclosure, ensuring minimal exposure of user information based on context and risk.

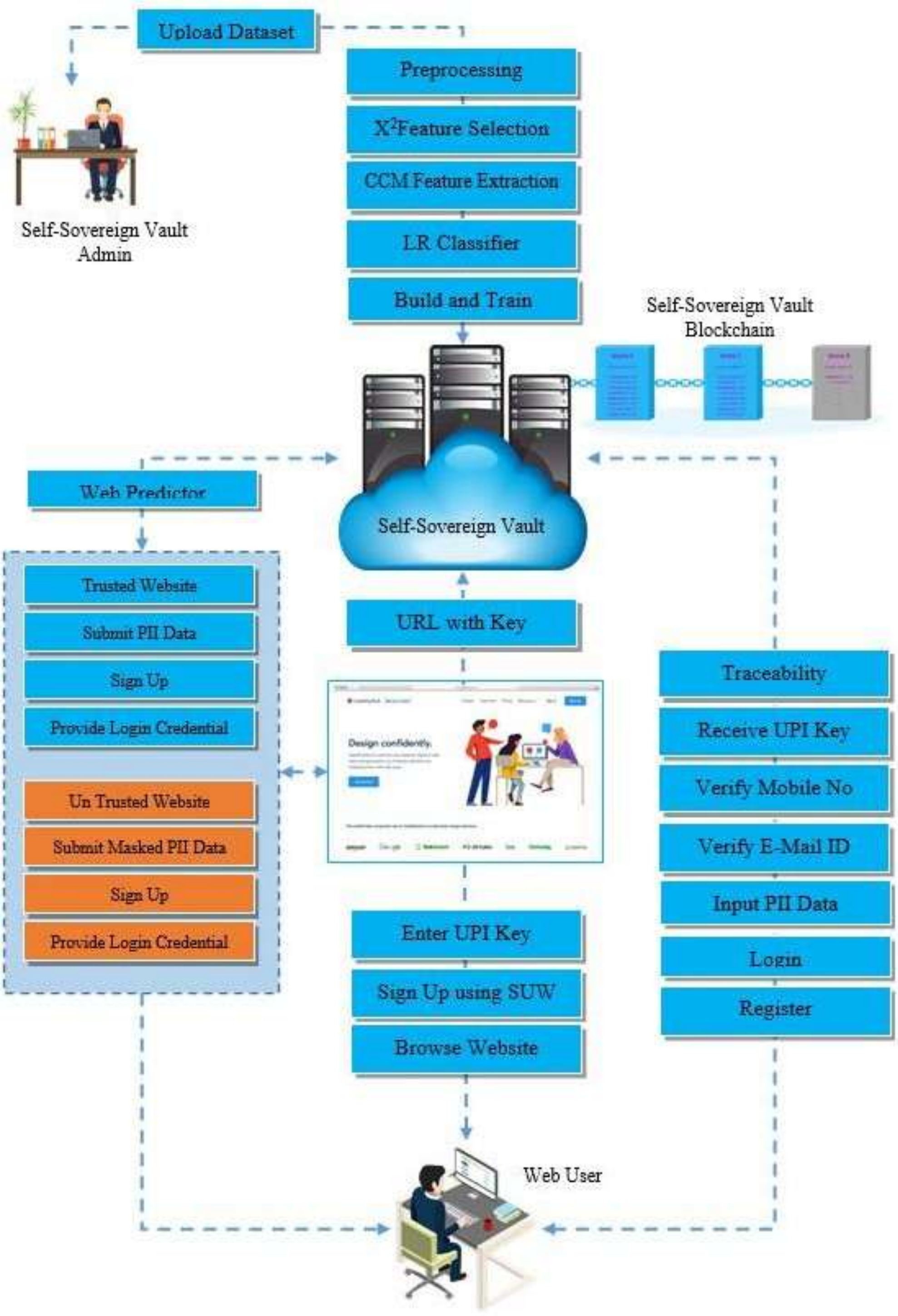


Fig. 1 Arichtecture Diagram

Beneath this lies the blockchain layer, which acts as the backbone for decentralized identity anchoring. Built on platforms like Hyperledger Indy or Ethereum, this layer records public DID documents and credential metadata in a tamper-evident manner. It does not store sensitive personal data but instead maintains cryptographic proofs that validate ownership and integrity of credentials. This immutable ledger enables verifiability without compromising user privacy. Supporting all other components, the data layer provides secure storage for actual credentials and encrypted user data. Rather than relying entirely on the blockchain, which is inherently public, the system stores sensitive information off-chain in local secure storage or encrypted cloud vaults. This separation ensures high performance, compliance with data protection regulations, and scalability for real-world deployment.

Together, these layers interact to form a robust system where users can control their identities, share only what is necessary, and rely on AI-powered decision-making to ensure trust and safety. The architecture enables a shift from centralized data management to a more user- centric model that prioritizes privacy and consent at every level.

6. ACTIVITY DIAGRAM

An activity diagram is a type of behavioral diagram that illustrates the flow of operations within a system or process. In the context of the Self-Sovereign Digital Identity system, the activity diagram models the sequence of interactions and verifications that occur when a user attempts to authenticate or share their digital identity with a service provider. The diagram helps to visualize the core components and decision points of the system, enabling the identification of potential vulnerabilities and the optimization of identity verification flows.

The activity flow of the Self-Sovereign Identity System presents a detailed view of how the system manages user identities in a secure, decentralized manner. The process begins with user registration, where a digital identity is created and encrypted using cryptographic keys and stored securely on the blockchain. This is followed by the generation of a Unique Personal Identifier (UPI), which acts as a reference for future identity verifications.

When the user interacts with a service provider, the UPI is submitted to initiate the verification process. The system employs a Logistic Regression-based trust evaluation module to assess the reliability of the requesting service provider. If the provider is verified as trustworthy, the user's credential is decrypted and securely shared. In contrast, if the provider is deemed untrustworthy, a masked credential is generated using the Lookup Substitution Algorithm to safeguard user privacy while ensuring the interaction remains functional.

Throughout the entire sequence, various modules such as the Blockchain Identity Ledger, Machine Learning Trust Engine, and Identity Masking Layer play critical roles in maintaining the integrity, confidentiality, and control of user data. The activity diagram thus reflects a comprehensive identity management workflow that blends decentralization, cryptography, and artificial intelligence to deliver a secure and privacy-centric digital identity ecosystem.

7. USE CASE DIAGRAM

A Self-Sovereign Identity (SSI) Use Case Diagram is a visual representation of the interactions between users, system components, and external entities involved in managing decentralized digital identities. This diagram is instrumental in modeling and analyzing the core functionalities of the blockchain-based identity management system and helps identify critical interactions and potential points of vulnerability. It provides a high-level view of how various stakeholders interact with the system to ensure secure and privacy-preserving identity management.

The primary actors in the SSI use case diagram include users, service providers, and the blockchain network. Users are individuals who create, own, and control their digital identities. Service providers are entities requesting identity verification during online interactions. The blockchain network functions as the decentralized ledger that securely stores identity data and manages verification logic using smart contracts.

Key use cases modeled in the diagram include user registration, identity creation, generation of the Unique Personal Identifier (UPI), submission of the UPI for verification, trust score evaluation using the Logistic Regression-based prediction model, and issuance of credentials. If the trust score indicates a risk, the system triggers the Lookup Substitution Algorithm to generate a masked version of the user's credentials. Additional use cases include updating identity attributes, revoking credentials, and viewing access history.

These interactions collectively represent the decentralized, AI-driven approach of the system. The use case diagram thus plays a crucial role in visualizing the system's functionality and ensuring that all security, privacy, and usability requirements are effectively captured and implemented.

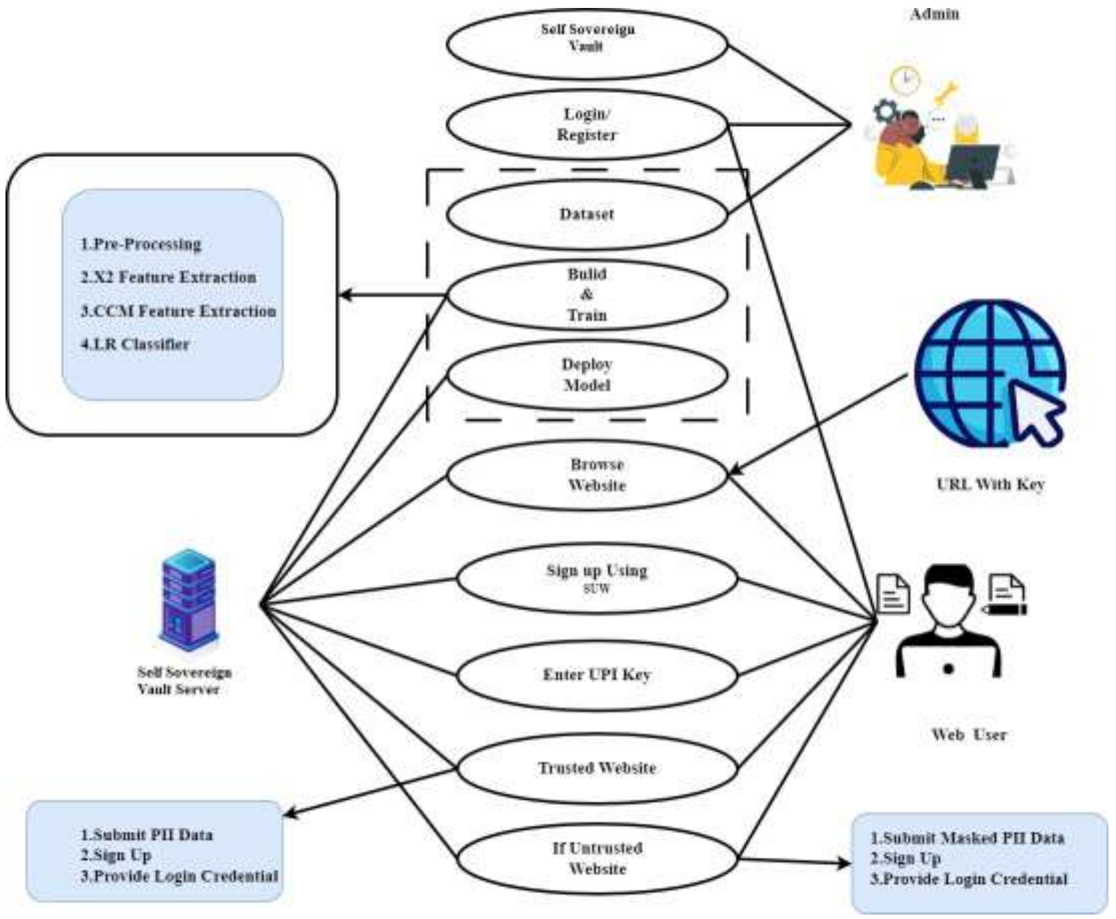


Fig. 2 Use Case Diagram

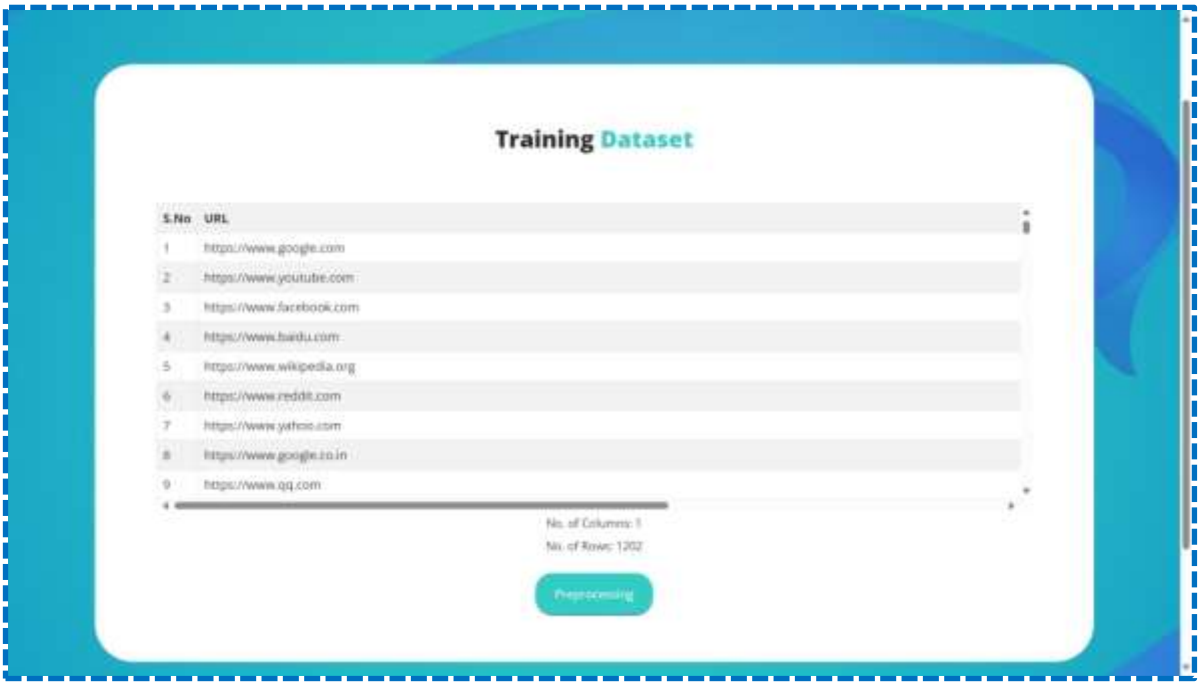
8. DISCUSSIONS AND RESULT

The discussion on the proposed Self-Sovereign Identity (SSI) system emphasizes the transformative potential of integrating blockchain and artificial intelligence to enhance the security, privacy, and autonomy of digital identity management. Traditional centralized systems pose significant risks due to their reliance on third-party providers and susceptibility to data breaches. In contrast, the proposed system empowers users with full ownership of their digital identities, eliminating the need for intermediaries and reducing exposure to external threats.

The system combines blockchain-based credential storage with AI-driven trust evaluation, using logistic regression to assess the reliability of service providers. This trust prediction capability allows the system to dynamically adapt its response, such as by triggering the Lookup Substitution Algorithm to generate masked credentials in untrusted scenarios, thereby preserving privacy without compromising functionality. The introduction of a Unique Personal Identifier (UPI) streamlines interactions while ensuring identity verification remains secure and decentralized.

Experimental results validate the system's capability to maintain security and user privacy even in untrusted environments. The use of cryptographic techniques, tamper-proof blockchain storage, and AI-enhanced decision-making demonstrates a robust framework that mitigates the risk of identity theft, unauthorized access, and surveillance. Overall, the results indicate that the system significantly improves digital identity management by offering a decentralized, transparent, and user-centric approach that aligns with modern security and privacy expectations.

9. RESULT

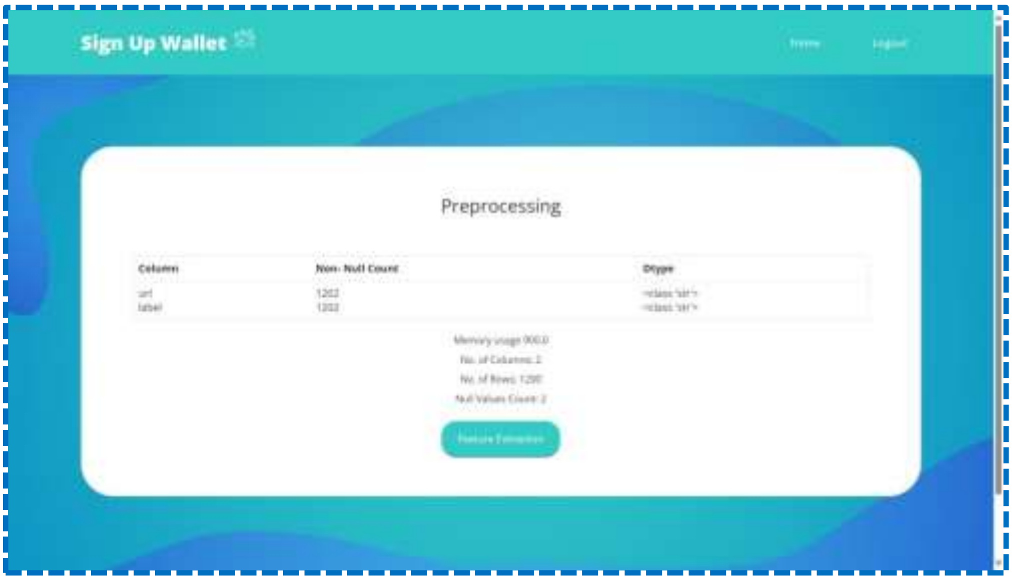


Training Dataset

S.No	URL
1	https://www.google.com
2	https://www.youtube.com
3	https://www.facebook.com
4	https://www.baidu.com
5	https://www.wikipedia.org
6	https://www.reddit.com
7	https://www.yahoo.com
8	https://www.google.co.in
9	https://www.qq.com

No. of Columns: 1

No. of Rows: 1202



Preprocessing

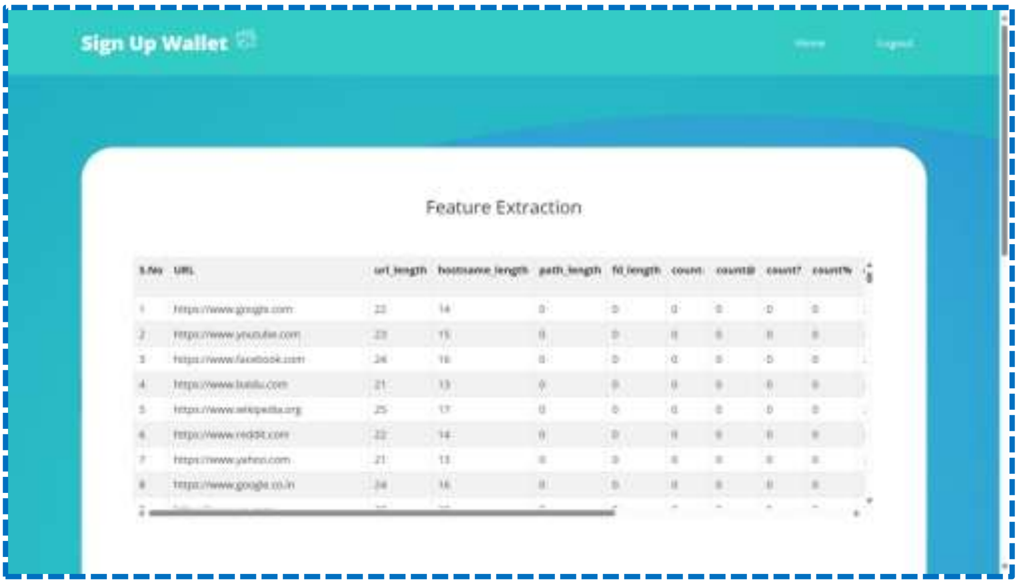
Column	Non- Null Count	Dtype
url	1202	<class 'str'>
label	1202	<class 'str'>

Memory usage 000.0

No. of Columns: 2

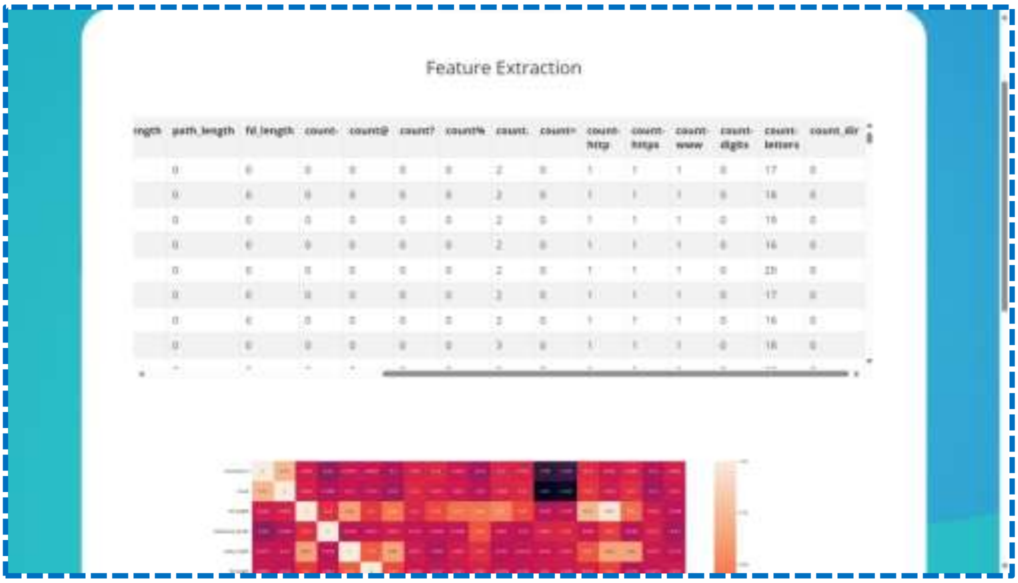
No. of Rows: 1200

Null Values Count: 2



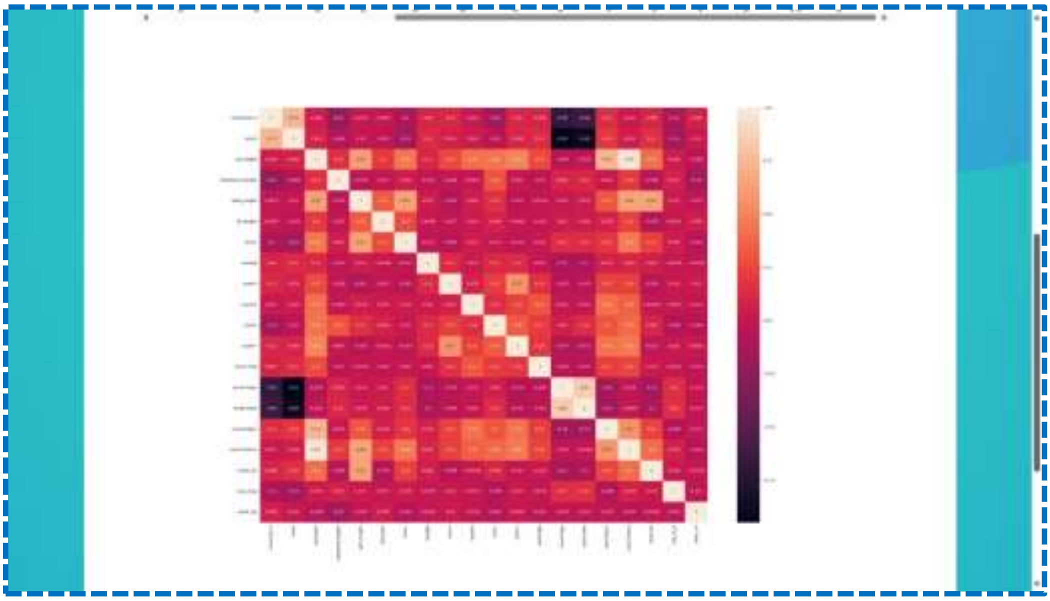
Feature Extraction

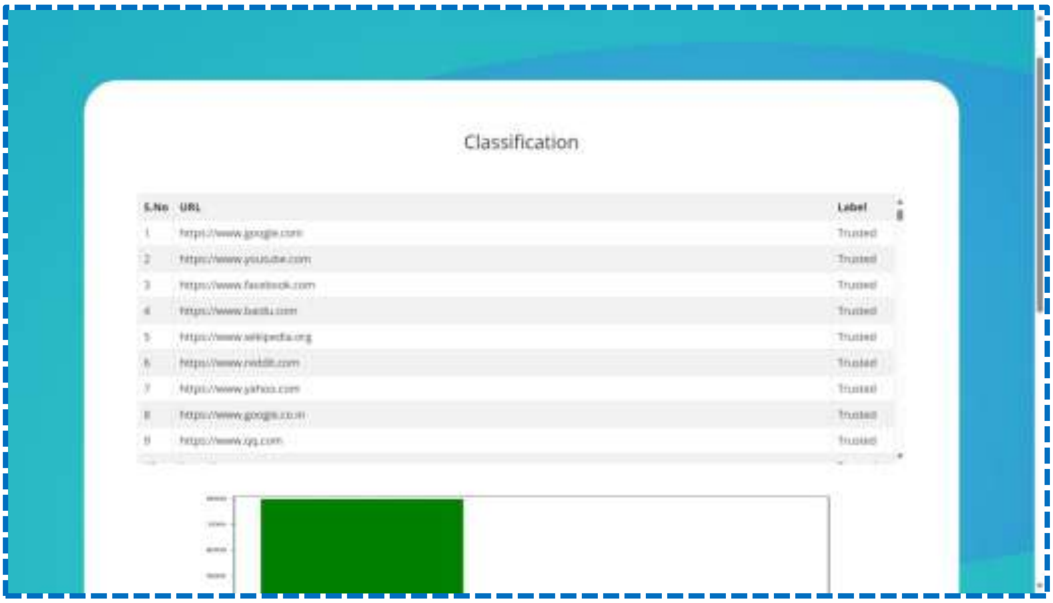
S.No	URL	url_length	hostname_length	path_length	fd_length	count-	count@	count?	count%
1	https://www.google.com	22	14	0	0	0	0	0	0
2	https://www.youtube.com	23	15	0	0	0	0	0	0
3	https://www.facebook.com	24	16	0	0	0	0	0	0
4	https://www.baidu.com	21	13	0	0	0	0	0	0
5	https://www.wikipedia.org	25	17	0	0	0	0	0	0
6	https://www.reddit.com	22	14	0	0	0	0	0	0
7	https://www.yahoo.com	21	13	0	0	0	0	0	0
8	https://www.google.co.in	24	16	0	0	0	0	0	0



Feature Extraction

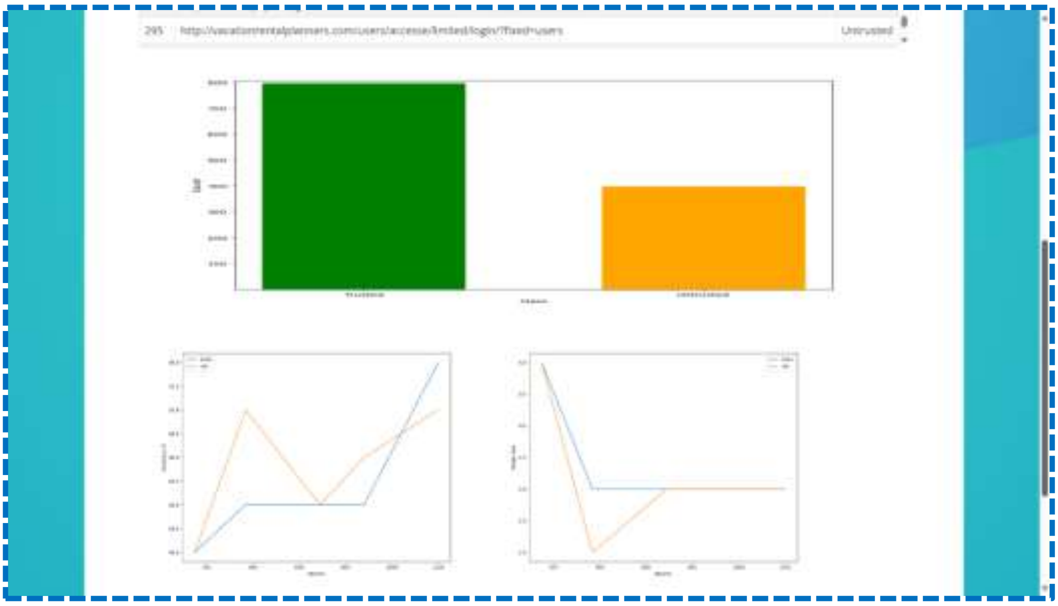
ngth	path_length	fd_length	count-	count@	count?	count%	count.	count=	count-http	count-https	count-www	count-digits	count-letters	count_dir
	0	0	0	0	0	0	2	0	1	1	1	0	17	0
	0	0	0	0	0	0	2	0	1	1	1	0	18	0
	0	0	0	0	0	0	2	0	1	1	1	0	19	0
	0	0	0	0	0	0	2	0	1	1	1	0	16	0
	0	0	0	0	0	0	2	0	1	1	1	0	21	0
	0	0	0	0	0	0	2	0	1	1	1	0	17	0
	0	0	0	0	0	0	2	0	1	1	1	0	16	0
	0	0	0	0	0	0	3	0	1	1	1	0	18	0



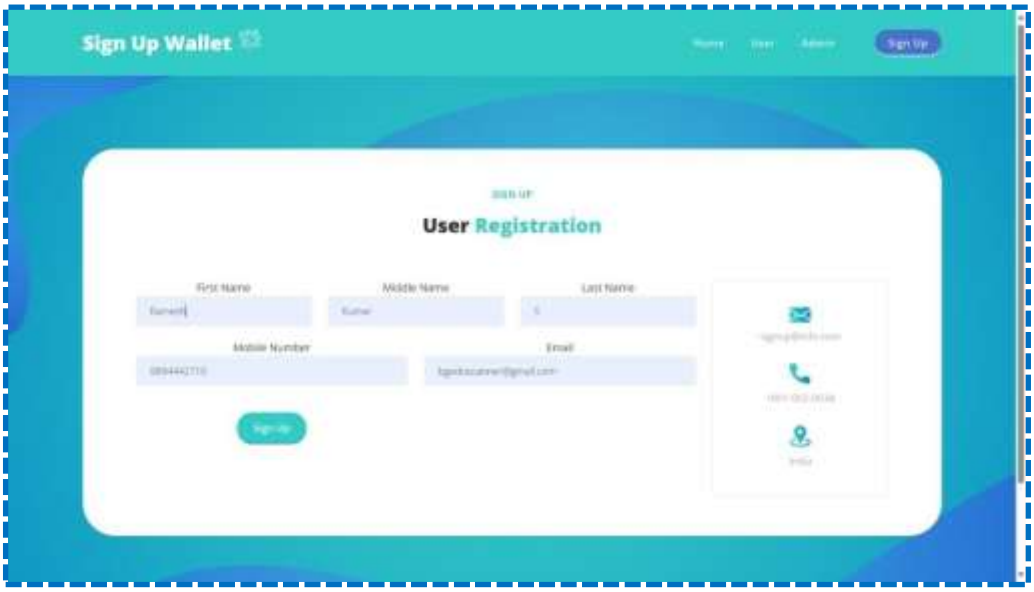


Classification

S.No	URL	Label
1	https://www.google.com	Trusted
2	https://www.youtube.com	Trusted
3	https://www.facebook.com	Trusted
4	https://www.baidu.com	Trusted
5	https://www.wikipedia.org	Trusted
6	https://www.reddit.com	Trusted
7	https://www.yahoo.com	Trusted
8	https://www.google.co.in	Trusted
9	https://www.qq.com	Trusted



295	http://vacationrentalplanners.com/users/access/limited/login/?fixed=users	Untrusted



User Registration

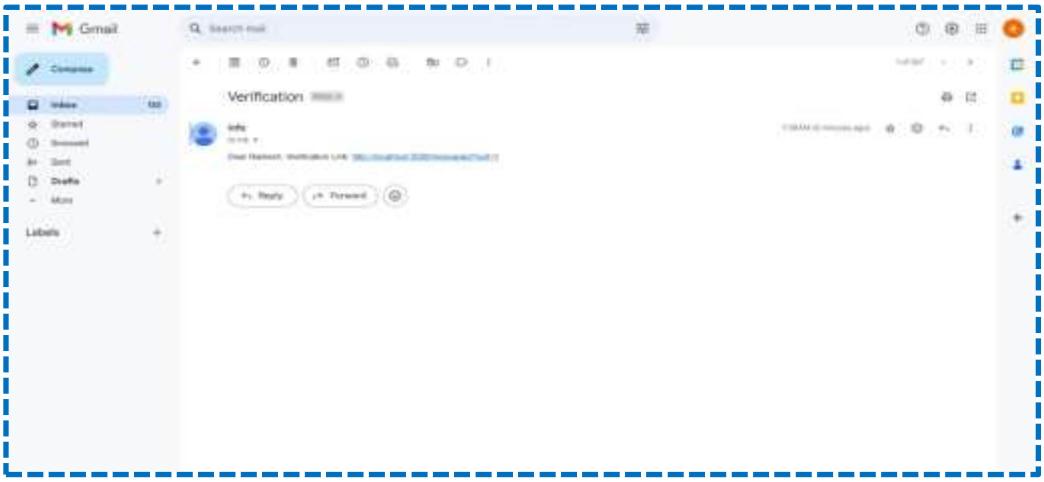
First Name: [illegible]

Middle Name: Kumar

Last Name: [illegible]

Mobile Number: [illegible]

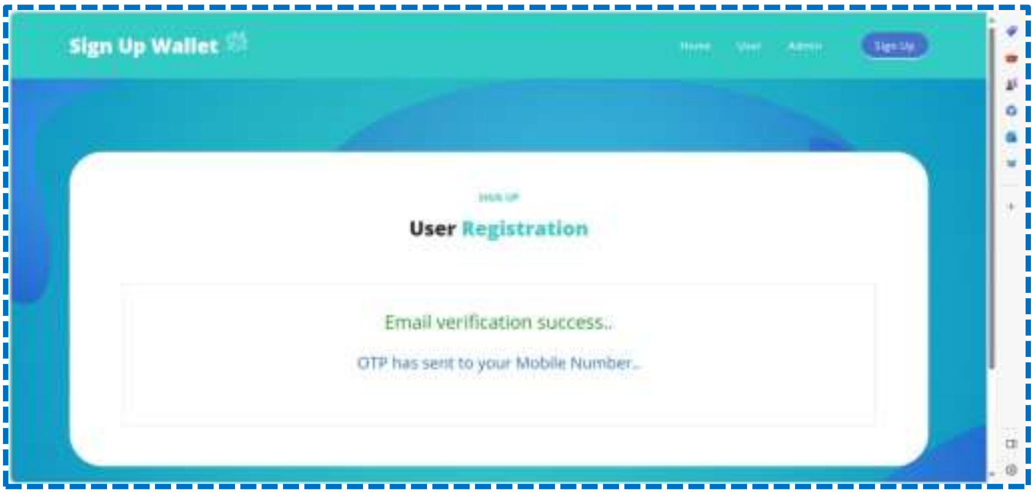
Email: [illegible]



Verification

info

Dear [illegible], Verification Link: [illegible]



User Registration

Email verification success..

OTP has sent to your Mobile Number..

10. CONCLUSION

This study proposed a novel framework for secure digital identity management by combining blockchain technology with machine learning to develop a self-sovereign identity (SSI) model. The system successfully eliminates the need for centralized identity providers, allowing users full ownership and control over their digital identities. Through blockchain's decentralized, tamper-proof infrastructure and logistic regression-based trust evaluation, the model ensures privacy, transparency, and trust in digital interactions.

The implementation of a Unique Personal Identifier (UPI) code for streamlined credential sharing, combined with the Lookup Substitution Algorithm for masking identities in untrusted environments, proved effective in enhancing user privacy while maintaining service compatibility. Despite the promising results, challenges such as scalability, interoperability across different platforms, and model generalization across a wider user base remain.

Overall, the proposed SSI framework lays a strong foundation for decentralized identity ecosystems, offering a secure, user-centric alternative to traditional identity systems. It marks a significant advancement toward greater privacy, data sovereignty, and digital trust in today's increasingly connected world.

11. FUTURE ENCHCEMENT

- Decentralized Trust Networks – Establishing peer-to-peer trust frameworks using blockchain smart contracts for collaborative identity verification without centralized authorities.
- Biometric-Based Identity Anchoring – Integrating facial recognition or fingerprint data to securely bind physical identity with digital credentials.
- Interoperability Across Platforms – Designing APIs and cross-platform modules to support seamless SSI usage across various web and mobile applications.
- Real-Time Risk Assessment using AI – Enhancing trust prediction models with real-time behavioral analysis and anomaly detection using deep learning.
- Quantum-Resistant Cryptography – Researching and implementing post-quantum cryptographic algorithms to future-proof identity security.
- Integration with IoT Ecosystems – Extending SSI to connected devices, enabling secure, authenticated interactions across smart environments.

12. REFERENCE

1. Allen, C., "The Path to Self-Sovereign Identity," Life with Alacrity Blog, 2016. [Online]. Available: https://www.lifewithalacrity.com/2016/04/the-path-to-self-sovereign- identity.html

(Pioneering work outlining the foundational principles of decentralized identity models.)

2. Tobin, A. and Reed, D., "The Inevitable Rise of Self-Sovereign Identity," Sovrin Foundation White Paper, 2016.

(Introduces the conceptual and architectural framework for self-sovereign identity systems.)

3. Zyskind, G., Nathan, O., and Pentland, A., "Decentralizing Privacy: Using Blockchain to Protect Personal Data," IEEE Security and Privacy Workshops, pp. 180-184, 2015.

(A landmark paper presenting the integration of blockchain and privacy-preserving identity systems.)

4. Dunphy, P. and Petitcolas, F. A. P., "A First Look at Identity Management Schemes on the Blockchain," IEEE Security & Privacy, vol. 16, no. 4, pp. 20–29, Jul./Aug. 2018.

(Analyzes various blockchain-based identity schemes in terms of privacy, decentralization, and security.)

5. Liu, B., Lin, Y., and Wen, H., "Blockchain-Enabled Privacy-Preserving Authentication with Logistic Regression in IoT," IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9896–9907, Jun. 2021.

(Discusses the use of machine learning models integrated with blockchain for trust-based authentication.)

6. Wörner, D., and von Bomhard, T., "When Your Sensor Earns Money: Exchanging Data for Cash with Bitcoin," Proceedings of the ACM Conference on Embedded Networked Sensor Systems, 2014.

(Explores secure decentralized IoT data exchange models relevant for digital identity verification.)

7. Bhargav, M., and Manogaran, G., "Secure Blockchain-Based Self-Sovereign Identity Model Using Artificial Intelligence for Personal Data Protection," Computer Communications, vol. 182, pp. 74–83, 2022.

(Proposes a hybrid AI-blockchain approach to self-sovereign identity with a focus on predictive security.)

8. Kshetri, N., "Blockchain's Roles in Strengthening Cybersecurity and Protecting Privacy," Telecommunications Policy, vol. 41, no. 10, pp. 1027–1038, 2017.

(Highlights the applicability of blockchain technology in reinforcing data privacy and access control.)

9. Hammi, B., Khatoun, R., and Zeadally, S., "IoT Technologies for Smart Cities," IET Networks, vol. 7, no. 1, pp. 1–13, 2018.

(Offers context for deploying blockchain-based identity solutions in large-scale IoT and urban systems.)

10. Hardjono, T., and Smith, N., "Decentralized Trust: The Trust over IP Stack," MIT Connection Science & Engineering, 2020.

(Introduces the Trust over IP (ToIP) architecture for verifiable digital credentials.)