International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Secured Document Storage using Blockchain with Integrated Data Processing and Analysis

Kalaiselvan M P S[1] Kirthka P[1] Sridhar M[1] Edees Narsullah Dar[1]
Sowbarnika A[1] Ms.R.Rajalakshmi[2]

[1]UG Student, Department of Computer Science and Engineering, Coimbatore Institute of Technology
[2]Assitant Professor, Department of Computer Science and Engineering, Coimbatore Institute of Technology

Abstract

Traditional methods of storing, retrieving, and processing legal, financial, and personal documents often lack sufficient security measures, accessibility, and efficient analysis capabilities. Privacy issues arise from the redundant nature of storage systems, and manual operations are excessively time-consuming and prone to errors. The merging of numerous texts and languages increases connectivity and accessibility challenges. This solution addresses these challenges by designing innovative systems based on advanced technologies such as the InterPlanetary File System (IPFS) for decentralized storage and blockchain technology for encrypted asset sharing. End-to-end encryption ensures confidentiality, while blockchain guarantees privacy and immutability, preventing fraud. Additionally, the system processes legal documents through features like editable text translation and summarization in English and Tamil. For financial data, it supports formats like CSV and XLSX for accounting integration and uses machine learning techniques to analyze data and predict future revenue. This comprehensive solution enhances security, access, and analysis for legal and financial document management.

Keywords: IPFS, Blockchain, End-to-End Encryption, Machine Learning, Legal and Financial Analytics

1.Introduction

The management of legal and financial documents has seen a significant transformation with the rise of digital solutions, replacing traditional paper-based systems and promising enhanced efficiency and convenience. However, these advancements also introduce several challenges, including data security concerns, potential hardware failures, and inefficiencies in managing large datasets. Furthermore, traditional OCR technologies often fail to process handwritten texts commonly found in legal documents, necessitating better solutions for document processing and analysis. Legal professionals require systems to locate specific provisions in extensive contracts, and financial professionals depend on accurate document management systems for data analysis, trend prediction, and risk management.

Key technologies in this system include the InterPlanetary File System (IPFS) for decentralized storage, blockchain for immutable record-keeping, and end-to-end encryption for enhanced security. Text processing techniques like summarization, translation, and OCR are employed, while machine learning models provide predictive financial analysis. Together, these components create an integrated ecosystem designed to improve the efficiency, security, and adaptability of legal and financial document management.

2.Related Work

In [1], Distributed Document Storage in Blockchain-based B2B Applications Hammoud et al. (2021) proposed a blockchain-based architecture to enhance distributed electronic document storage in B2B applications. The system integrates blockchain's immutability with distributed storage networks to mitigate issues like data tampering, operational costs, and reliance on centralized servers. It utilizes cryptographic

hashing for data integrity and smart contracts for automating document access and sharing. This work sets a foundation for secure document management systems in decentralized environments.

In [2], Secured Data Storage in Decentralized Cloud Using Ethereum
Khan et al. (2022) designed a decentralized cloud storage model utilizing Ethereum blockchain to address privacy and data integrity challenges. The solution incorporates cryptographic encryption for secure data storage and smart contracts for access control automation. This architecture highlights Ethereum's potential to provide cost-effective, scalable, and transparent solutions for decentralized storage, contributing to advancements in secure cloud infrastructure.

In [3], Blockchain for Chain of Custody in Physical Evidence
Batista et al. (2023) explored blockchain's application in managing the chain of custody for physical evidence. Through a systematic review, the authors analyzed blockchain features like immutability, traceability, and transparency in preserving evidence integrity. Smart contracts were proposed for automating access permissions and process tracking. The work highlights blockchain's role in improving accountability and reliability in evidence management for sectors such as law enforcement and forensic science.

In [4], Survey on Blockchain Technology: Evolution, Architecture, and Security
Bhutta et al. (2021) conducted a survey on blockchain's development, architecture, and security features, emphasizing its applicability across diverse industries. The paper detailed blockchain's core components, such as distributed ledgers and consensus protocols, and explored security measures against tampering and cyberattacks. Challenges like scalability, energy inefficiency, and privacy were discussed, along with potential solutions to enhance blockchain adoption.

In [5], Blockchain Interoperability and Smart Contracts
Khan et al. (2021) investigated blockchain interoperability challenges and the role of smart contracts in enabling seamless integration between multiple blockchain systems. The study highlighted the importance of interoperability for achieving cohesive blockchain ecosystems. Smart contracts were identified as critical enablers for automating cross-chain transactions and maintaining security. This work underscores the necessity of developing standardized frameworks to promote blockchain interoperability.

3.Methodology

3.1 Overview of Data Protection Mechanism

The methodology begins by exploring critical data protection mechanisms within cloud environments, focusing on data storage, sharing, encryption, access control, and privacy preservation. A comparative analysis of these mechanisms is carried out to determine their effectiveness in real-world applications, such as healthcare

3.2 Blockchain Integration for secure Document Management

Blockchain's distributed ledger technology ensures data integrity and immutability, while smart contracts and eVaults automate compliance tasks. Blockchain's integration guarantees transparency and security in document management, addressing challenges like fraud prevention and scalability.

3.3 Advanced Encryption Techniques

The system incorporates end-to-end encryption to safeguard sensitive data during storage and transmission. Searchable encryption allows users to search for documents without exposing sensitive content, while fuzzy search enhances usability by accommodating variations in query terms.

3.4 Use of Inter Planetary File System

IPFS is employed for decentralized storage, ensuring data availability and resilience. By distributing files across a peer-to-peer network, the system increases data access while reducing storage costs, making it ideal for storing legal and financial documents.

3.5 Machine Learning for Financial Analysis

Machine learning models are integrated to analyze financial data, identifying trends and forecasting future revenues. These models support financial professionals in decision-making by providing insights into risk assessment and investment planning.

3.6 Validation through case studies and testing

The system's effectiveness is validated through case studies in education, healthcare, and legal sectors. Testing ensures that the encryption, blockchain, and machine learning functionalities are robust and perform well in real-world scenarios.

4.System Components and Workflow

4.1 Overview

SecuSto is a comprehensive platform designed to manage personal, financial, and legal documents. It ensures secure storage, efficient retrieval, and advanced processing capabilities for a variety of document types. The platform employs role-based access control to allow users to perform their tasks securely and efficiently, ensuring that each user interacts with the system according to their designated role.

4.2 Layout

The architecture of SecuSto is organized into several layers, each serving a distinct function to enhance the overall performance of the system. The User Interface Layer provides intuitive portals for users to manage their documents, whether personal, financial, or legal. The Application Layer integrates core functionalities such as access control and document processing, enabling seamless interactions with the platform. The Data Security Layer is responsible for implementing encryption and blockchain technologies to protect the integrity and confidentiality of the data. Finally, the Data Layer uses IPFS and hybrid databases to store and manage data efficiently, ensuring high availability and fault tolerance.

4.3 System Components

SecuSto consists of several critical components that work in tandem to provide its functionalities. The User Management System is responsible for handling authentication and role-based access control, ensuring that only authorized users can access specific documents or features. The Document Storage System leverages IPFS for decentralized storage, ensuring that data is distributed and secure across multiple nodes.

The Data Encryption System uses encryption techniques to safeguard documents during storage and transmission. The Blockchain Ledger ensures an immutable record of all document transactions, providing transparency and accountability. The Document Processing System includes features such as text extraction, summarization, and translation, enabling efficient document management. For financial data, the Financial Analytics Engine analyzes the stored data, identifies trends, and makes predictions to assist users in financial planning and decision-making.

4.4 Workflow Design

The workflows within SecuSto are customized to meet the distinct needs of users interacting with different types of documents. The workflow for personal documents emphasizes secure storage and organization, ensuring that individuals can easily manage and retrieve their files. The workflow for financial documents integrates advanced analytics for trend analysis and forecasting, enabling financial professionals to analyze data and make informed decisions. The legal document workflow includes specialized features like text summarization, translation, and secure updates, allowing legal professionals to handle documents more efficiently and securely.

4.5 Security Features

SecuSto incorporates cutting-edge security measures to ensure that all data and documents remain safe. Advanced encryption techniques, such as AES and RSA, protect documents during storage and transmission. Role-based access control ensures that only authorized users can access sensitive data, while blockchain integration guarantees transparency and

immutability of the records. These security features collectively ensure that the SecuSto system provides a secure and reliable environment for document management.

5.Conclusion

The proposed SecuSto system revolutionizes document management by integrating blockchain, IPFS, encryption, and machine learning to address the challenges of security, accessibility, and analysis. By providing a comprehensive, secure, and efficient platform for managing legal and financial documents, SecuSto paves the way for enhanced trust, scalability, and automation in these critical sectors. It offers a robust solution that ensures the safe storage, retrieval, and processing of documents, helping organizations maintain efficiency and security in an increasingly digital world.

References

[1] O. Hammoud, I. Tarkhanov, and A. Kosmarski, "An architecture for distributed electronic documents storage in decentralized blockchain B2B applications," Computers, vol. 10, no. 11, p. 142, Nov. 2021.

[2] N. Khan, H. Aljoaey, M. Tabassum, A. Farzamnia, T. Sharma, and Y. H. Tung, "Proposed model for secured data storage in decentralized cloud by blockchain Ethereum," Electronics, vol. 11, no. 22, p. 3686, Nov. 2022.

[3] D. Batista, A. L. Mangeth, I. Frajhof, P. H. Alves, R. Nasser, G. Robichez, G. M. Silva, and F. P. de Miranda, "Exploring blockchain technology for chain of custody control in physical evidence: A systematic literature review," Journal of Risk and Financial Management, vol. 16, no. 8, p. 360, Aug. 2023.

[4] M. N. M. Bhutta, M. R. Malik, R. Ghaffar, F. Ullah, and Z. Rehman, "A survey on blockchain technology: Evolution, architecture and security," IEEE Access, vol. 9, pp. 61050–61078, Apr. 2021.

[5] S. Khan, M. B. Amin, A. T. Azar, and S. Aslam, "Towards interoperable blockchains: A survey on the role of smart contracts in blockchain interoperability," IEEE Access, vol. 9, pp. 116672–116691, Aug. 2021.