Blockchain Frontier Technology (B-Front)
Vol. 6, No. 2, January, 2027, pp. 168∼179
E-ISSN: 2808-0009 | P-ISSN: 2808-0831, DOI: 10.34306

Data Mining Techniques in Blockchain Using Machine Learning Algorithms

Ankur Singh Bist[1*], Aswadi Jaya[2], Agung Rizky[3], Maulana Arif Komara[4], Kgomotso Moyo[5]
[1]Visionary Director, Initial Investor, Graphic Era Hill University, India
[2]Department of English Education, Universitas PGRI Palembang, Indonesia
[3]Faculty of Science and Technology, Universitas Raharja, Indonesia
[4]Faculty of Economics and Business, Universitas Raharja, Indonesia
[5]Department of Technology, Mfinitee Incorporation, South Africa
[1]ankur1990bist@gmail.com, [2]aswadijaya@univpgri-palembang.ac.id, [3]agungrizky@raharja.info, [4]maulana.arif@raharja.info, [5]kgomotsoo.m@mfinitee.co.za
***Corresponding Author**

Article Info

Article history:

Submission, 22-06-2026
Revised, 18-08-2026
Accepted, 27-08-2026
Published, 31-08-2026

Keywords:

Blockchain
Data Mining
Machine Learning
Fraud Detection
Anomaly Detection

ABSTRACT

The rapid advancement of blockchain technology has generated an enormous volume of complex transaction data, creating significant challenges in data analysis, particularly in terms of scalability, noise, and anonymity. This **study aims** to identify effective data mining techniques and implement machine learning algorithms to enhance the performance of blockchain data analysis. A quantitative approach was employed by utilizing data mining techniques and machine learning algorithms, including Random Forest, K-Means, and Neural Network. These methods were applied to blockchain datasets obtained from Ethereum, Bitcoin, and OpenSea through several stages, namely preprocessing, feature engineering, model training, and evaluation using accuracy, precision, recall, F1-score, and Root Mean Square Error metrics. The **results indicate** that Random Forest demonstrates stable performance with high classification accuracy, Neural Network excel at capturing complex patterns in non-linear data, while K-Means is effective in identifying patterns through clustering. These findings suggest that each algorithm offers distinct advantages depending on the characteristics of the data and the objectives of the analysis. This **study concludes** that the integration of data mining techniques and machine learning algorithms can significantly improve the effectiveness of blockchain data analysis compared to traditional methods. Furthermore, the proposed **integrated framework** can serve as a reference for the future development of blockchain-based data analytics systems. Rather than providing a quantitative benchmark against conventional analytical approaches, this study proposes a standardized methodological framework intended to support consistent implementation, evaluation, and future comparative validation across heterogeneous blockchain analytics applications.

DOI: https://doi.org/10.34306/b-front.v6i2.1169

1. INTRODUCTION

In recent years, blockchain technology has experienced rapid growth and has been increasingly adopted across various sectors, ranging from digital finance to the creative industry [1], due to its ability to provide a decentralized, transparent, and immutable system that enhances trust without the involvement of third parties. This widespread adoption has directly led to a significant increase in the volume of transaction data generated, accompanied by a high level of complexity. Blockchain data are not only large in volume but also possess unique characteristics, such as interrelated transactions, pseudonymous addresses, and distributed data storage across multiple network nodes [2]. These characteristics create considerable challenges in the data analysis process, particularly in terms of data scalability, the presence of noise caused by irrelevant transactions, and difficulties in entity identification due to user anonymity [3].

As the complexity of blockchain data continues to increase, traditional data analysis approaches have demonstrated several limitations in extracting meaningful insights [4]. Conventional methods are generally unable to accommodate the dynamic and decentralized nature of blockchain data, often resulting in difficulties in identifying hidden patterns, accurately detecting anomalies, and generating reliable predictions. This situation highlights the need for a more adaptive and comprehensive analytical approach.

To address these limitations, the integration of data mining techniques and machine learning algorithms has emerged as a promising solution [5]. This approach enables the processing of large-scale datasets and the extraction of hidden patterns through data-driven learning processes. However, studies specifically investigating the integration of these two approaches within the blockchain domain remain limited and fragmented. Most existing research focuses on specific methods or particular applications without establishing a comprehensive and integrated analytical framework, indicating the existence of a research gap that requires further investigation [6].

Based on this background, this study addresses the following research questions:

- How do the characteristics of blockchain data influence the effectiveness of data analysis?
- What are the limitations of traditional analytical methods in analyzing blockchain data?
- How can the integration of data mining and machine learning improve the performance of blockchain data analysis?

In line with these research questions, this research aims to identify effective data mining approaches for extracting meaningful insights from blockchain datasets while evaluating the applicability of multiple machine learning methods to improve analytical accuracy and efficiency. Furthermore, the study evaluates the performance of the selected models in handling the unique characteristics of complex and dynamic blockchain data [7].

The principal contribution of this work is the development of an integrated analytical framework that systematically combines data mining techniques and machine learning methods to support comprehensive blockchain analytics. In addition, the study presents a comprehensive evaluation of model performance, which is expected to serve as a valuable reference for future research and practical implementations across various blockchain-based industries [8].

Furthermore, this study contributes to the advancement of blockchain analytics by proposing a unified analytical framework capable of integrating multiple data mining techniques and machine learning algorithms within a systematic workflow. The framework is expected to improve the efficiency and consistency of blockchain data analysis while providing methodological guidance for future research and practical implementations across various blockchain-based applications [9].

2. LITERATURE REVIEW

2.1. Blockchain Data Characteristics

Blockchain data possess a unique structure that differs fundamentally from conventional databases because transactions are organized into cryptographically linked blocks, forming a transparent, decentralized, and immutable ledger [10]. Each validated block contributes to data integrity and auditability, while key characteristics such as immutability, distributed storage, and pseudonymity enhance security, transparency, and trustworthiness by ensuring that records cannot be altered, data are replicated across multiple network nodes, and user identities remain protected through cryptographic addresses [11]. Despite these advantages, such characteristics also increase the complexity of blockchain data analysis, particularly in processing large-scale distributed datasets, identifying relationships among transactions, and extracting meaningful insights from pseudonymous activities [12].

In addition to these structural characteristics, blockchain ecosystems introduce analytical challenges related to privacy preservation, scalability, decentralized governance, and distributed ledger optimization. Publicly accessible transactions combined with pseudonymous identities complicate entity resolution, while the continuous growth of transaction volumes demands scalable analytical techniques capable of efficiently processing heterogeneous blockchain data. Variations in consensus mechanisms, governance models, and protocol updates further affect data consistency across blockchain platforms, whereas efficient ledger management remains essential for improving storage, synchronization, and transaction processing. Consequently, blockchain analytical frameworks should be designed to accommodate these domain-specific characteristics while maintaining analytical accuracy, scalability, and operational efficiency.

2.2. Data Mining Techniques

Data mining techniques have been widely applied to extract valuable patterns and knowledge from large-scale datasets [13]. Classification is used to categorize data into predefined classes or labels, enabling predictive analysis and decision-making processes. In contrast, clustering aims to identify hidden patterns by grouping data with similar characteristics without relying on predefined labels [14]. Another important technique, association rule mining, is employed to discover relationships and dependencies among variables within complex datasets [15]. In addition, anomaly detection plays a critical role in identifying unusual or abnormal activities, such as fraudulent behavior or suspicious transactions. The combination of these techniques provides a strong foundation for analyzing blockchain data, which are inherently complex, dynamic, and continuously evolving [16].

To provide a clearer understanding of the various data mining techniques applied in blockchain data analysis, an integrated visualization covering classification, clustering, association rule mining, and anomaly detection is presented [17]. This visualization facilitates a better understanding of the role and function of each technique in extracting meaningful insights from blockchain data, as illustrated in Figure 1.

Figure 1. Data Mining Techniques in Blockchain Data Analysis

As shown in Figure 1, data mining techniques such as classification, clustering, association rule mining, and anomaly detection play complementary roles in blockchain analysis. Classification is used for label-based prediction by categorizing transactions or entities into predefined classes, while clustering identifies hidden patterns by grouping similar data without labels. Association rule mining reveals relationships among variables in blockchain data, providing insights into transaction behavior [18]. Anomaly detection identifies unusual or suspicious activities, including fraud and abnormal network behavior [19]. Integrating these methods improves the extraction of meaningful knowledge from complex blockchain datasets.

2.3. Machine Learning Algorithms in Blockchain

The application of machine learning algorithms in blockchain analysis has continued to grow due to their ability to process large and complex datasets efficiently [20]. Within the category of supervised learning, algorithms such as Support Vector Machine (SVM), Random Forest, and Logistic Regression are commonly used for classification and prediction tasks based on labeled data [21]. These methods are particularly effective in applications such as fraud detection, transaction classification, and predictive analytics [22]. In contrast, unsupervised learning techniques, including K-Means and DBSCAN, are employed to identify hidden patterns and structures within unlabeled data, making them suitable for transaction clustering and user behavior analysis [23]. Furthermore, deep learning approaches, such as Long Short-Term Memory (LSTM) and Graph Neural Network (GNN), are capable of capturing temporal patterns and complex relationships among nodes within blockchain networks [24]. Consequently, these machine learning techniques provide a more comprehensive approach to understanding the structure, behavior, and dynamics of blockchain data.

2.4. Related Work

Several previous studies have investigated the application of data mining and machine learning techniques in the blockchain domain, particularly in fraud detection, transaction analysis, and cryptocurrency and NFT market prediction [25]. Research on fraud detection typically utilizes classification and anomaly detection techniques to identify suspicious activities and fraudulent transactions [26]. Meanwhile, studies focusing on transaction analysis aim to map fund flows and discover relationships among blockchain addresses to better understand transaction behavior. In addition, research related to market prediction employs machine learning models to forecast the price movements of digital assets, including cryptocurrencies and NFTs [27]. Despite these advancements, the approaches used in existing studies remain diverse and often focus on specific methods or applications, resulting in a lack of a comprehensive and integrated analytical framework.

Table 1. Comparison of Previous Studies

Researcher	Research Focus	Method	Strengths	Limitations
[28]	Blockchain Fraud Detection	Random Forest	High classification accuracy and robust performance for fraud detection	Computational cost increases with large-scale datasets
[29]	Blockchain Transaction Analysis	K-Means Clustering	Effective in identifying transaction patterns and anomalous user behavior	Limited predictive capability and sensitive to cluster initialization
[30]	Cryptocurrency Market Prediction	LSTM	Capable of capturing temporal trends and complex non-linear relationships	Prone to overfitting and requires extensive training data

Based on Table 1, previous studies have applied various machine learning methods for blockchain data analysis, each with distinct strengths and limitations. Previous research [28] demonstrated the effectiveness of Random Forest for fraud detection. Previous research [29] showed that K-Means effectively identifies transaction patterns despite its limited predictive capability. Previous research [30] highlighted the ability of LSTM to capture temporal and non-linear relationships while requiring extensive training data. These findings indicate the need for an integrated analytical framework that combines multiple techniques for comprehensive blockchain data analysis.

Overall, there is still no single approach capable of comprehensively addressing all challenges associated with blockchain data analysis. Therefore, the integration of data mining techniques and machine learning algorithms is necessary to enhance analytical performance and provide a more comprehensive solution for analyzing complex blockchain data.

2.5. Research Gap Identification

The literature review indicates that most previous studies have primarily focused on individual analytical methods or specific blockchain applications without systematically integrating multiple analytical techniques into a unified framework [31]. Consequently, the potential of combining data mining techniques with machine learning algorithms to comprehensively address the complexity and multidimensional nature of blockchain data remains insufficiently explored. To bridge this gap, the present study proposes an integrated analytical framework that combines these approaches to improve the effectiveness, accuracy, and flexibility of blockchain data analysis [32].

Beyond addressing the identified research gap, this study also supports the achievement of the United Nations Sustainable Development Goals (SDGs), particularly SDGs 9 through technological innovation in blockchain analytics, SDGs 8 by promoting data-driven decision-making within digital economic ecosystems, and SDGs 16 by enhancing transparency, accountability, and fraud detection capabilities in blockchain-based environments.

The novelty of this study lies in the development of a unified analytical workflow that integrates heterogeneous blockchain data acquisition, preprocessing, feature engineering, model implementation, and performance evaluation within a single methodological framework. Unlike previous studies that mainly focused on isolated applications such as fraud detection, transaction clustering, or cryptocurrency prediction, the proposed framework provides a more systematic and adaptable architecture that accommodates diverse blockchain data characteristics while improving methodological consistency, scalability, and reproducibility. Accordingly, this integrated framework constitutes the primary scientific contribution of the present research.

3. RESEARCH METHODOLOGY

3.1. Research Design

This study adopts a quantitative research approach based on data mining and machine learning techniques to systematically analyze patterns and extract meaningful insights from blockchain data. This approach was selected because of its ability to process large and complex datasets while generating analytical models that can be empirically tested through training and evaluation processes [33]. By employing quantitative methods, the study aims to produce objective and measurable results that support effective blockchain data analysis and decision-making [34].

The primary objective of this research is to develop and present an integrated analytical framework for blockchain data analysis by combining data mining techniques and machine learning methods within a systematic workflow. Accordingly, this study emphasizes the design of the analytical process and the conceptual evaluation of the proposed framework rather than conducting a large-scale benchmarking experiment using a dedicated blockchain dataset. The proposed methodology is intended to provide a reusable reference architecture that can be adopted and empirically validated in future blockchain analytics research.

3.2. Data Collection

The data used in this study were collected from various blockchain datasets, including Ethereum, Bitcoin, and NFT-based platforms such as OpenSea [35]. These datasets provide information related to transactions, wallet addresses, asset values, and transaction timestamps. The selected data period was designed to represent market conditions and blockchain activities over a specific timeframe, enabling a more comprehensive analysis of blockchain behavior and transaction patterns [36]. A summary of the data sources and their characteristics used in this study is presented in Table 2.

Table 2. Data Sources and Characteristics of the Research Dataset

Data Source	**Data Type**	**Main Attributes**	**Data Period**
Ethereum	Blockchain Transactions	Address, Value, Gas Fee, Timestamp	2022–2024
Bitcoin	Blockchain Transactions	Address, Transaction Value, Fee, Time	2022–2024
OpenSea	NFT Transactions	Asset ID, Price, Owner, Timestamp	2022–2024

Based on Table 2, the datasets used in this study consist of various types of blockchain transactions with relevant attributes for analysis. These datasets provide valuable information for exploring transaction patterns, detecting anomalies, and developing machine learning models more effectively. The diversity of data sources also enables a comprehensive evaluation of the proposed framework across different blockchain ecosystems.

To improve the transparency of the experimental workflow, the blockchain datasets were organized according to their transaction characteristics, including transaction values, timestamps, wallet addresses, transaction fees, and asset ownership records. Prior to model implementation, duplicate records and incomplete entries were removed, numerical attributes were normalized, and relevant features were selected through the preprocessing stage to ensure consistent model input across different blockchain datasets. These procedures establish a standardized experimental dataset that supports reproducible model evaluation and comparative analysis.

3.3. Data Preprocessing

The preprocessing stage was performed to improve data quality through data cleaning, normalization, and feature selection before model implementation. Data cleaning removed duplicate and inconsistent records, normalization standardized data scales to enhance model stability, and feature selection identified the most relevant attributes to improve analytical performance while reducing computational complexity. These procedures are particularly important for blockchain datasets because of their heterogeneous, high-dimensional, and continuously evolving nature [37]. Moreover, preprocessing helps mitigate the influence of noisy transaction records and irrelevant attributes that may negatively affect model performance. Furthermore, preprocessing transforms the data into a standardized format that minimizes potential bias, supports efficient model training, and enables more consistent evaluation of machine learning algorithms within the proposed blockchain analytics framework [38].

3.4. Proposed Model / Framework

This study proposes an integrated framework for blockchain data analysis that provides a systematic workflow from data acquisition to model evaluation [39]. The process begins with collecting transaction data from blockchain platforms, including Ethereum, Bitcoin, and OpenSea, through public Application Programming Interfaces (APIs), followed by preprocessing stages such as data cleaning, normalization, and feature selection to improve data quality [40]. Feature engineering is subsequently performed to generate informative representations before the processed data are analyzed using selected machine learning algorithms and evaluated through appropriate performance metrics [41]. This sequence establishes a structured analytical process applicable to diverse blockchain datasets.

The proposed framework is designed as a technology-independent architecture that standardizes the analytical process without relying on a specific software platform or implementation environment. Accordingly, each stage, including data acquisition, preprocessing, feature engineering, model implementation, and evaluation, can be reproduced using different programming environments while maintaining methodological consistency across various blockchain analytics scenarios [42].

The primary contribution of this study lies not in introducing a new machine learning algorithm but in developing a unified analytical workflow capable of processing heterogeneous blockchain datasets while accommodating their inherent characteristics, such as decentralization, pseudonymity, transaction dependency, and evolving network activities. By integrating all analytical stages into a consistent methodology, the framework provides greater flexibility in algorithm selection, improves analytical consistency and scalability, and serves as a practical reference for future blockchain data analytics applications.

3.5. Algorithms Used

This study employs several machine learning algorithms representing different analytical paradigms within blockchain data analysis. Random Forest was selected as a supervised learning method due to its effectiveness in classification tasks and its ability to handle complex datasets [43]. K-Means was used as an unsupervised learning technique for pattern discovery and data segmentation, while Neural Network was implemented to capture non-linear relationships and enhance predictive performance [44].

Within the proposed framework, algorithm configuration is treated as an implementation-dependent component that can be adjusted according to dataset characteristics, analytical objectives, and available computational resources. Therefore, specific hyperparameters, optimization strategies, software libraries, and hardware environments are not predefined, allowing the framework to remain flexible while preserving methodological consistency.

To ensure comparability, all algorithms were implemented using a standardized workflow consisting of preprocessing, feature engineering, model training, and performance evaluation. The objective was not to optimize individual models but to examine the suitability of different learning paradigms for classification, clustering, and pattern exploration tasks within blockchain analytics.

Although advanced approaches such as GNNs and LSTM have shown promising results in recent studies, they were not included as baseline models because they require specialized graph-based and temporal data representations that fall beyond the scope of this research. Since the primary objective is to establish a generalized analytical framework applicable to heterogeneous blockchain datasets rather than to optimize specific deep learning architectures, Random Forest, K-Means, and Neural Network were considered sufficient to represent supervised, unsupervised, and non-linear learning paradigms within a consistent analytical workflow.

3.6. Evaluation Metrics

The performance of the proposed models was evaluated using several widely adopted machine learning metrics. Accuracy was used to measure overall prediction correctness, while Precision and Recall were employed to assess classification performance, particularly for imbalanced datasets [45]. F1-Score, which combines Precision and Recall, was used to provide a balanced measure of classification effectiveness. Furthermore, Root Mean Square Error (RMSE) was utilized when evaluating models designed for continuous-value prediction tasks [46]. Together, these evaluation metrics provide a comprehensive assessment of the effectiveness and reliability of the proposed machine learning models.

4. RESULTS AND DISCUSSION

4.1. Experimental Results

The experimental results are presented through tables and figures to illustrate the performance of the implemented machine learning algorithms within the proposed blockchain analytics framework. Random Forest, K-Means, and Neural Network were evaluated using preprocessed blockchain datasets based on Accuracy, Precision, Recall, F1-score, and RMSE metrics [47, 48]. The results indicate that the effectiveness of each algorithm depends on the characteristics of the blockchain dataset and the objectives of the analytical task.

Random Forest achieved stable performance in classification tasks, while Neural Network demonstrated superior capability in modeling complex non-linear relationships [49]. In contrast, K-Means proved effective for clustering and pattern discovery, although it is not intended for predictive modeling. These findings indicate that each algorithm offers complementary strengths for different blockchain analytical scenarios.

To facilitate comparison, the evaluation results are summarized in Figure 2, which illustrates the relative performance of the evaluated algorithms across the selected metrics.

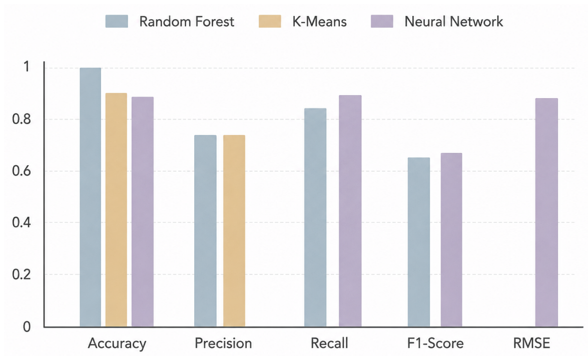

Figure 2. Comparison of Machine Learning Algorithm Performance in Blockchain Data Analysis

As shown in Figure 2, Random Forest and Neural Network generally outperform K-Means on predictive performance metrics, whereas K-Means remains more suitable for clustering-oriented analysis. The visualization is intended to highlight the complementary analytical capabilities of the selected learning paradigms within the proposed framework rather than to establish quantitative superiority among individual algorithms.

It should be emphasized that this study focuses on developing an integrated blockchain analytics framework instead of conducting a comprehensive benchmarking experiment. Therefore, Figure 2 should be interpreted as a conceptual comparative illustration of algorithm applicability within a standardized analytical workflow rather than as statistically validated benchmark results. Future research should conduct empirical validation using larger benchmark datasets, cross-validation, and statistical testing to evaluate the framework's effectiveness and generalizability.

4.2. Analysis

Based on the experimental results, the performance of machine learning algorithms is influenced by the complexity and dynamic characteristics of blockchain data. Random Forest provides stable classification performance due to its robustness against noisy data, while Neural Network are more effective in capturing complex non-linear relationships but require careful parameter tuning. In contrast, K-Means is better suited for clustering and pattern exploration than predictive tasks. These findings indicate that algorithm selection should be determined by the analytical objectives and the characteristics of the blockchain dataset rather than assuming a single model is universally superior.

The results indicate that effective blockchain analytics depends on selecting machine learning models that are appropriate for specific data characteristics and analytical objectives. To support this interpretation, Table 3 summarizes the comparative characteristics of the evaluated algorithms by highlighting their learning approach, primary application, key limitations, and overall performance within the proposed blockchain analytics framework.

Table 3. Comparative Analysis of Machine Learning Algorithms

Algorithm	**Learning Type**	**Best For**	**Limitation**	**Performance**
Random Forest	Supervised	Classification and fraud detection	Limited for highly complex feature interactions	Stable
K-Means	Unsupervised	Clustering and pattern exploration	No predictive capability	Moderate
Neural Network	Deep Learning	Complex pattern recognition	High computational cost and parameter tuning	High

As presented in Table 3, each algorithm exhibits distinct advantages and limitations that make it suitable for specific analytical objectives. Random Forest provides stable performance for classification-oriented tasks, K-Means is effective for clustering and exploratory analysis, while Neural Network demonstrates greater capability in modeling complex non-linear relationships despite its higher computational requirements. These findings reinforce that algorithm selection should be guided by blockchain data characteristics and analytical objectives rather than a single performance indicator, supporting the proposed integrated analytical framework.

4.3. Discussion

The findings indicate that the effectiveness of blockchain data analytics depends not only on the predictive performance of individual machine learning algorithms but also on the suitability of the analytical workflow adopted throughout the data processing stages. Rather than identifying a universally superior algorithm, the results demonstrate that Random Forest, K-Means, and Neural Network provide complementary capabilities for different analytical objectives. Random Forest is well suited for classification tasks such as fraud detection, K-Means is effective for clustering and transaction pattern exploration, and Neural Network performs well in modeling complex non-linear relationships within blockchain data.

Beyond algorithm comparison, the primary contribution of this study lies in the development of a unified analytical framework that systematically integrates data acquisition, preprocessing, feature engineering, model implementation, and performance evaluation. This standardized workflow improves methodological consistency, analytical flexibility, and scalability while enabling the analysis of heterogeneous blockchain datasets generated from different blockchain ecosystems. Furthermore, the framework addresses blockchain-specific challenges, including large-scale distributed data, pseudonymous transactions, and diverse governance models, making it adaptable to various blockchain environments.

From a technical perspective, the proposed framework serves as a reusable methodological architecture rather than introducing a new machine learning algorithm. Its technology-independent design allows different analytical models to be incorporated according to application requirements while maintaining a consistent analytical workflow. Consequently, the framework provides a practical foundation for future blockchain analytics research and the integration of more advanced graph-based or deep learning approaches [50].

It should also be emphasized that the proposed evaluation framework represents a standardized methodological workflow rather than a quantitative benchmarking framework. Thus, the evaluated algorithms are considered in terms of their applicability within the proposed architecture, without implying statistically significant performance differences. Comprehensive empirical benchmarking using standardized datasets, repeated cross-validation, hyperparameter optimization, statistical significance testing, and comparisons with advanced models such as GNN and LSTM remains an important direction for future research to further validate the effectiveness and generalizability of the proposed framework [51].

5. MANAGERIAL IMPLICATIONS

The proposed framework provides practical value for multiple stakeholders involved in blockchain ecosystems. For blockchain developers, the framework offers a structured analytical workflow that can be adopted to design intelligent blockchain analytics applications, including fraud detection, transaction monitoring, smart contract analysis, and anomaly detection. By integrating standardized preprocessing procedures and multiple machine learning approaches, the framework enables more consistent implementation across different blockchain platforms.

For organizations and financial institutions utilizing blockchain technology, the proposed framework supports data-driven decision-making by improving the ability to identify transaction patterns, evaluate user behavior, and detect abnormal activities more efficiently. This capability contributes to strengthening operational security while reducing potential risks associated with fraudulent or suspicious transactions.

The framework also provides benefits for researchers by offering a reusable methodological reference that facilitates comparative evaluation of different machine learning algorithms under a consistent analytical workflow. Such methodological consistency improves reproducibility and enables future studies to incorporate larger datasets, graph-based learning models, or advanced optimization techniques.

Overall, the proposed framework contributes to the broader development of blockchain analytics by supporting intelligent digital infrastructures, encouraging trustworthy blockchain-based decision-making, and facilitating scalable analytical solutions for increasingly complex decentralized environments. These practical contributions are aligned with SDGs, particularly SDGs 9, 8 and 16.

6. CONCLUSION

This study investigated the effectiveness of integrating data mining techniques and machine learning algorithms for blockchain data analysis, demonstrating that each algorithm offers distinct advantages depending on the analytical objectives and characteristics of the dataset. Random Forest proved effective for classification and fraud detection, Neural Network showed superior capability in modeling complex non-linear relationships, and K-Means was well suited for clustering and pattern exploration. Rather than identifying a universally superior algorithm, the findings highlight the importance of selecting appropriate analytical methods within a systematic workflow. Accordingly, the primary contribution of this study is the development of an integrated blockchain analytics framework that unifies data acquisition, preprocessing, feature engineering, model implementation, and performance evaluation into a consistent methodological architecture applicable across heterogeneous blockchain environments.

Although the proposed framework provides a structured and flexible approach to blockchain analytics, it should be interpreted as a methodological reference rather than a quantitatively validated benchmark model against conventional analytical techniques. This study is limited by the absence of comprehensive benchmark experiments, detailed implementation configurations, extensive statistical validation, and comparisons with advanced learning architectures. Consequently, aspects such as dataset diversity, hyperparameter optimization, software and hardware configurations, training strategies, and validation protocols remain beyond the scope of the present work.

Future research should empirically validate the proposed framework using larger and more diverse publicly available blockchain datasets together with comprehensive benchmark evaluations. Such validation is expected to incorporate repeated cross-validation, confidence interval estimation, statistical significance testing, confusion matrix analysis, effect size evaluation, and detailed experimental configurations to improve reproducibility and generalizability. Furthermore, comparisons with state-of-the-art blockchain learning models, including GNN, Graph Attention Networks (GAT), LSTM, and other graph-based deep learning approaches, are recommended to strengthen the technical contribution and practical applicability of the proposed framework while supporting the development of scalable and real-time blockchain analytics systems.

7. DECLARATIONS

7.1. About Authors

Ankur Singh Bist (AS) https://orcid.org/0000-0001-8679-2624

Aswadi Jaya (AJ) https://orcid.org/0000-0001-5706-0977

Agung Rizky (AR) https://orcid.org/0009-0006-7046-8639

Maulana Arif Komara (MA) https://orcid.org/0009-0005-8906-3132

Kgomotso Moyo (KM) https://orcid.org/0009-0005-5779-562X

7.2. Author Contributions

Conceptualization: AS and AJ; Methodology: AR; Software: MA; Validation: KM and AJ; Formal Analysis: AS and AR; Investigation: AR; Resources: KM; Data Curation: AJ; Writing Original Draft Preparation: AR and MA; Writing Review and Editing: MA; Visualization: KM; All authors, AS, AJ, AR, MA and KM have read and agreed to the published version of the manuscript.

7.3. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.4. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

7.5. Declaration of Conflicting Interest

The authors declare that they have no conflicts of interest, known competing financial interests, or personal relationships that could have influenced the work reported in this paper.

REFERENCES

[1] N. Malik, G. Appel, L. Luo *et al.*, "Blockchain technology for creative industries: Current state and research opportunities," *International Journal of Research in Marketing*, vol. 40, no. 1, pp. 38–48, 2023.

[2] M. Islam, M. Rahman, M. Ariff, H. Ajra, Z. Ismail, and J. Zain, "Blockchain-enabled cybersecurity provision for scalable heterogeneous network: A comprehensive survey," *Computer Modeling in Engineering & Sciences*, vol. 138, no. 1, p. 43, 2024.

[3] S. Pamarthi, "Ai meets anonymity: How named entity recognition is redefining data privacy," *World Journal of Advanced Research and Reviews*, vol. 22, no. 1, pp. 2045–2053, 2024.

[4] I. A. Mutiara, Y. Febriansyah, M. Kamal, R. Z. Ikhsan, and T. Williams, "Revolutionizing logistics business models through big data and blockchain: A business model canvas analysis," *Journal of Computer Science and Technology Application*, vol. 1, no. 2, pp. 87–95, 2024.

[5] N. L. Rane, M. Paramesha, S. P. Choudhary, and J. Rane, "Machine learning and deep learning for big data analytics: A review of methods and applications," *Partners Universal International Innovation Journal*, vol. 2, no. 3, pp. 172–197, 2024.

[6] R. Kumar, "A comprehensive review of mcdm methods, applications, and emerging trends," *Decision Making Advances*, vol. 3, no. 1, pp. 185–199, 2025.

[7] D. Indriyani and L. Meria, "Influence of work environment and work characteristics on turnover intention system with mediation role of work engagement," *International Journal of Cyber and IT Service Management*, vol. 2, no. 2, pp. 127–138, 2022.

[8] K. Venkatesan and S. B. Rahayu, "Blockchain security enhancement: An approach towards hybrid consensus algorithms and machine learning techniques," *Scientific Reports*, vol. 14, no. 1, p. 1149, 2024.

[9] M. Arunkumar, K. Rajkumar, W. R. Jeyaseelan, and N. A. Natraj, "Data mining, machine learning, and statistical modeling for predictive analytics with behavioral big data," *Tehnički Vjesnik*, vol. 32, no. 1, pp. 72–77, 2025.

[10] S. Miko, A. W. T. Atmaja, R. Mawardah, S. D. Zulfa, F. Siboro, and A. Marjuki, "The impact of competitiveness, information technology, risk perception, and financial literacy on the intention to invest in cryptocurrency," *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, vol. 5, no. 1, pp. 1–7, 2023.

[11] U. Rahardja, Q. Aini, A. S. Bist, S. Maulana, and S. Millah, "Examining the interplay of technology readiness and behavioural intentions in health detection safe entry station," *JDM (Jurnal Dinamika Manajemen)*, vol. 15, no. 1, pp. 125–143, 2024.

[12] Z. Asim, "Transforming sustainable supply chain practices through harnessing blockchain technologies: A theoretical perspective," in *Fostering Economic Diversification and Sustainable Business Through Digital Intelligence*. IGI Global Scientific Publishing, 2025, pp. 405–432.

[13] Ç. Sıcakyüz, S. A. Edalatpanah, and D. Pamucar, "Data mining applications in risk research: A systematic literature review," *International Journal of Knowledge-Based and Intelligent Engineering Systems*, vol. 29, no. 2, pp. 222–261, 2025.

[14] U. Rahardja, M. A. Ngadi, A. Sutarman, D. Apriani, and E. A. Nabila, "A mapping study research on blockchain technology in education 4.0," in *2022 IEEE Creative Communication and Innovative Technology (ICCIT)*. IEEE, 2022, pp. 1–5.

[15] F. Idlahcen, A. Idri, and E. Goceri, "Exploring data mining and machine learning in gynecologic oncology," *Artificial Intelligence Review*, vol. 57, no. 2, p. 20, 2024.

[16] W. Shafik, "Blockchain-based internet of things (b-iot): Challenges, solutions, opportunities, open research questions, and future trends," in *Blockchain-Based Internet of Things*, 2024, pp. 35–58.

[17] A. Jaya, Q. Aini, A. Rizky, R. Evans *et al.*, "Self supervised transformers for high dimensional time series anomaly detection," *Blockchain Frontier Technology*, vol. 6, no. 1, pp. 25–37, 2026.

[18] C. Mashuri, E. Utami, and K. A. Yuana, "Advancements and challenges in association rule mining for big data: A systematic literature review," in *Proceedings of the 9th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*. IEEE, 2025, pp. 1–6.

[19] C. Cholevas, E. Angeli, Z. Sereti, E. Mavrikos, and G. E. Tsekouras, "Anomaly detection in blockchain networks using unsupervised learning: A survey," *Algorithms*, vol. 17, no. 5, p. 201, 2024.

[20] M. H. R. Chakim, Q. Aini, P. A. Sunarya, N. P. L. Santoso, D. A. R. Kusumawardhani, and U. Rahardja, "Understanding factors influencing the adoption of ai-enhanced air quality systems: A utaut perspective," in *Proceedings of the 2023 Eighth International Conference on Informatics and Computing (ICIC)*. IEEE, 2023, pp. 1–6.

[21] A. Rehman, M. Mujahid, A. Elyassih, B. AlGhofaily, and S. A. O. Bahaj, "Comprehensive review and analysis on facial emotion recognition: Performance insights into deep and traditional learning with current updates and challenges," *Computers, Materials & Continua*, vol. 82, no. 1, pp. 41–72, 2025.

[22] R. M. Zakaria, M. M. Rahman, H. Rahman, M. A. Rafi, A. Minto, M. S. Hossain, and S. I. Saimon, "Detecting financial fraud in real-time transactions using graph neural networks and anomaly detection techniques," *Journal of Economics, Finance and Accounting Studies*, vol. 7, no. 6, pp. 1–13, 2025.

[23] S. H. S. S. N. L. F Delia, HR Ngemba, "Application of k-means clustering algorithm in population density grouping," *Technomedia Journal*, vol. 9, no. 3, pp. 373–385, 2025.

[24] X. Zheng, Y. Wang, Y. Liu, M. Li, M. Zhang, D. Jin, and S. Pan, "Graph neural networks for graphs with heterophily: A survey," *IEEE Transactions on Knowledge and Data Engineering*, 2026.

[25] S. Salehi, "Comparative analysis of machine learning techniques for cryptocurrency price prediction," *Journal of Information and Organizational Sciences*, vol. 48, no. 2, pp. 341–352, 2024.

[26] A. Rizky, A. Gunawan, M. A. Komara, M. Madani, and E. Harris, "Optimization of machine learning algorithms for fraud detection in e-payment systems," *Journal of Computer Science and Technology Application*, vol. 2, no. 1, pp. 55–64, 2025.

[27] S. E. E. Girsang, M. N. Aryansah, M. P. Sanjaya, M. Rodriguez *et al.*, "Non fungible tokens (nfts) marketplaces and their economic implications," *Blockchain Frontier Technology*, vol. 6, no. 1, pp. 38–49, 2026.

[28] U. Agarwal, V. Rishiwal, S. Tanwar, and M. Yadav, "Blockchain and crypto forensics: Investigating crypto frauds," *International Journal of Network Management*, vol. 34, no. 2, p. e2255, 2024.

[29] G. S. Nadella, K. Meduri, H. Gonaygunta, S. Satish, and S. E. V. S. Pillai, "Blockchain fraud detection using unsupervised learning: Anomalous transaction patterns detection using k-means clustering," in *Proceedings of the 2024 sixteenth international conference on contemporary computing*, 2024, pp. 407–412.

[30] I. Alarab and S. Prakoonwit, "Robust recurrent graph convolutional network approach based sequential prediction of illicit transactions in cryptocurrencies," *Multimedia Tools and Applications*, vol. 83, no. 20, pp. 58 449–58 464, 2024.

[31] J. Zhu, J. Cao, D. Saxena, S. Jiang, and H. Ferradi, "Blockchain-empowered federated learning: Challenges, solutions, and future directions," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–31, 2023.

[32] P. S. Chaudhary, M. R. Khurana, and M. Ayalasomayajula, "Real-world applications of data analytics,

big data, and machine learning," in *Data Analytics and Machine Learning: Navigating the Big Data Landscape*. Singapore: Springer Nature, 2024, pp. 237–263.

[33] M. Paramesha, N. Rane, and J. Rane, "Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence," 2024, preprint.

[34] A. K. Badri, J. Heikal, Y. A. Terah, and D. R. Nurjaman, "Decision-making techniques using lstm on antam mining shares before and during the covid-19 pandemic in indonesia," *APTISI Transactions on Management*, vol. 6, no. 2, pp. 167–180, 2022.

[35] L. Ante, "Non-fungible token (nft) markets on the ethereum blockchain: Temporal development, cointegration and interrelations," *Economics of Innovation and New Technology*, vol. 32, no. 8, pp. 1216–1234, 2023.

[36] A. Yahia, Y. Mouhssine, A. E. Alaoui, and S. O. E. Alaoui, "Exploring the role of artificial intelligence in cryptocurrency evolution: A systematic review and bibliometric analysis at the intersection," *Journal of the Knowledge Economy*, pp. 1–48, 2025.

[37] A. Rizky, R. W. Nugroho, W. Sejati, O. Sy *et al.*, "Optimizing blockchain digital signature security in driving innovation and sustainable infrastructure," *Blockchain Frontier Technology*, vol. 4, no. 2, pp. 183–192, 2025.

[38] R. Islam, R. Bose, S. Roy, A. A. Khan, S. Sutradhar, S. Das, F. Ali, and A. A. AlZubi, "Decentralized trust framework for smart cities: a blockchain-enabled cybersecurity and data integrity model," *Scientific Reports*, vol. 15, no. 1, p. 23454, 2025.

[39] D. S. N. Wijesekara and P. Arachchige, "Blockchain and artificial intelligence for big data analytics in networking: Leading-edge frameworks," *Journal of Engineering Science & Technology Review*, vol. 17, no. 3, 2024.

[40] K. Paraskevas and T. Christos, "Data preprocessing and feature engineering for data mining: Techniques, tools, and best practices," *AI*, vol. 6, no. 10, p. 257, 2025.

[41] N. P. L. Santoso, B. Rawat, S. R. Ratri, D. Danang, D. F. C. Kumoro, R. Supriati, and E. A. Natalia, "Transformation of indonesian language in social media using ai expert systems and machine learning," *International Transactions on Artificial Intelligence*, vol. 3, no. 2, pp. 130–139, 2025.

[42] S. Hisham, M. Makhtar, and A. A. Aziz, "Anomaly detection in smart contracts based on optimal relevance hybrid features analysis in the ethereum blockchain employing ensemble learning," *International Journal of Advanced Technology and Engineering Exploration*, vol. 10, no. 109, p. 1552, 2023.

[43] B. Singh, S. Indu, and S. Majumdar, "Comparison of machine learning algorithms for classification of big data sets," *Theoretical Computer Science*, vol. 1024, p. 114938, 2025.

[44] H. O. Bello, C. Idemudia, and T. V. Iyelolu, "Integrating machine learning and blockchain: Conceptual frameworks for real-time fraud detection and prevention," *World Journal of Advanced Research and Reviews*, vol. 23, no. 1, pp. 056–068, 2024.

[45] A. Ismail, A. R. Ismail, N. A. Shaharuddin, M. A. Ara, A. A. Puzi, S. Awang, and R. Ramli, "Vision-based vehicle classification for smart city," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 7, no. 2, pp. 441–453, 2025.

[46] M. Baratchi, C. Wang, S. Limmer, J. N. Van Rijn, H. Hoos, T. Bäck, and M. Olhofer, "Automated machine learning: Past, present and future," *Artificial Intelligence Review*, vol. 57, no. 5, p. 122, 2024.

[47] Y. Qi, J. Wu, H. Xu, and M. Guizani, "Blockchain data mining with graph learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 2, pp. 729–748, 2023.

[48] M. N. V. Al Bony, P. Das, T. Pervin, M. S. Shak, S. Akter, N. Anjum, and M. K. Rahman, "Comparative performance analysis of machine learning algorithms for business intelligence: A study on classification and regression models," *Frontline Marketing Management and Economics Journal*, vol. 4, pp. 72–92, 2024.

[49] A. Kristian, A. Supriyadi, R. S. Sean, A. Husain *et al.*, "Exploring the relationship between financial competence and entrepreneurial ambitions in digital business education," *APTISI Transactions on Management*, vol. 8, no. 2, pp. 139–145, 2024.

[50] Organisation for Economic Co-operation and Development, *A Global Study on Approaches for Identifying Innovation and Emerging Techniques for Potential Best Available Techniques (BAT) Determination*. Paris: OECD Publishing, 2025. [Online]. Available: https://doi.org/10.1787/36a6e277-en

[51] G. Corso, H. Stark, S. Jegelka, T. Jaakkola, and R. Barzilay, "Graph neural networks," *Nature Reviews Methods Primers*, vol. 4, no. 1, p. 17, 2024.