Clustering Blockchain Wallet Behaviour Using K-Means and DBSCAN for Risk Profiling and Address Segmentation

Raed Ghanem[1,*]

[1]Department of Chemistry, Al Al-Bayt University, Jordan

ABSTRACT

Blockchain networks generate high-dimensional transactional data with diverse and irregular wallet behaviours. Understanding these behavioural patterns is essential for improving security monitoring, anomaly detection, and risk assessment within decentralized systems. This study applies two unsupervised machine learning algorithms, K-Means and DBSCAN, to analyse 303 blockchain wallet records using key attributes such as BlockHeight, UnixTimestamp, Block Density, Coin Day Weight, and Stake Distribution Rate. K-Means successfully identified three distinct behavioural clusters consisting of Cluster 1 with 200 wallets, Cluster 2 with 100 wallets, and Cluster 0 with 3 highly anomalous wallets. Numerical analysis revealed clear differences across clusters, including mean BlockHeight values of 5.5 million for Cluster 1, 15.4 million for Cluster 2, and 10.9 million for Cluster 0, along with Block Density percentages of 19.35, 48.90, and 60.00, respectively. DBSCAN further exposed behavioural complexity by detecting more than 90 noise points that represent irregular or outlier activity patterns and several small micro-clusters not captured by K-Means. PCA visualizations confirmed strong separation between clusters and highlighted the unique positioning of anomalous wallets. The combined use of centroid-based and density-based clustering provides a robust analytical foundation for profiling blockchain wallet behaviour, supporting more effective anomaly detection, risk classification, and address segmentation.

Keywords Blockchain Analytics, Unsupervised Machine Learning, K-Means Clustering, DBSCAN, Wallet Behaviour Profiling

INTRODUCTION

Blockchain technology has emerged as a transformative innovation in modern digital systems because it offers decentralized, transparent, and tamper-resistant transaction management [1]. Unlike traditional financial infrastructures that depend on centralized intermediaries such as banks or payment processors, blockchain relies on a distributed ledger in which transactions are validated collectively through consensus mechanisms, including Proof of Work and Proof of Stake [2]. This decentralized approach increases trust among participants and simultaneously produces extensive volumes of transactional data that contain rich behavioural patterns, making blockchain an important domain for computational analysis [3].

The accelerating adoption of blockchain in cryptocurrency trading, decentralized finance, digital identity, non-fungible token ecosystems, and smart contract automation has greatly increased the diversity of wallet interactions [4]. Each wallet address may represent an individual user, an automated bot, a mining pool, a staking participant, a smart contract entity, or even a malicious actor who takes advantage of blockchain pseudonymity. As a result, blockchain networks display a wide range of activity characteristics that

Submitted: 28 September 2025
Accepted: 10 November 2025
Published: 26 May 2026

Corresponding author
Raed Ghanem,
raedsa@yahoo.com

Additional Information and Declarations can be found on page 123

DOI: 10.47738/jcrb.v3i2.67

How to cite this article: R. Ghanem, "Clustering Blockchain Wallet Behaviour Using K-Means and DBSCAN for Risk Profiling and Address Segmentation," *J. Curr. Res. Blockchain*, vol. 3, no. 2, pp. 113-124, 2026.

vary in frequency, transaction value, staking intensity, temporal activity distribution, and involvement in block creation [5]. Understanding these variations is essential for improving risk assessment, fraud detection, cybersecurity monitoring, and regulatory compliance.

Even though blockchain provides transparency through its immutable ledger, the network remains exposed to numerous forms of abuse. Criminals often exploit pseudonymity to conduct phishing attacks, Ponzi schemes, wash trading, ransomware payments, or transaction obfuscation using mixers and automated scripts [6]. These activities frequently appear in the data as irregular wallet behaviours such as sudden transaction bursts, unusual block density patterns, inconsistent coin age distributions, or abnormal staking rewards. Because blockchain datasets are unlabelled and extremely high in volume, many conventional monitoring techniques are unable to capture hidden behavioural structures. This makes unsupervised machine learning a highly relevant method for analysing blockchain activity without the need for predefined labels.

Clustering algorithms such as K Means and DBSCAN are widely used to identify natural groupings in high-dimensional data and to detect hidden structures that are not captured by rule-based analysis [7]. K Means is effective for identifying dominant behavioural categories because it groups data based on similarity to cluster centroids. In contrast, DBSCAN is capable of detecting dense behavioural regions and isolating noise points that represent unusual or suspicious activity. The complementary strengths of these algorithms allow researchers to map both mainstream behavioural patterns and anomalies like structures within blockchain transactions. Prior studies have shown that combining centroid-based clustering with density-based clustering provides a robust analytical foundation for fraud detection, user segmentation, and financial risk analysis [8].

In this research, clustering analysis is performed on a dataset consisting of 303 blockchain wallet observations that include quantitative indicators such as BlockHeight, UnixTimestamp, Block Density, Stake Distribution Rate, Coin Day Weight, and several additional operational features. The K Means algorithm successfully identifies three behavioural clusters consisting of 200 routine wallets, 100 moderately active wallets, and 3 wallets with extremely unusual activity patterns. DBSCAN identifies more than 90 noise points, showing that a substantial portion of the dataset contains irregular or non-standard behavioural signatures. This combination of results demonstrates that blockchain ecosystems contain both stable and anomalous wallet activities, confirming the importance of unsupervised learning for behavioural modelling in decentralized systems.

The main contributions of this study include a comprehensive segmentation of blockchain wallet behaviour using two complementary clustering techniques, a set of empirical findings that quantify behavioural differences across clusters, and the establishment of an analytical basis for future risk profiling and anomaly detection frameworks. As blockchain usage continues to grow in complexity, the need for data-driven analysis becomes increasingly critical for supporting secure and trustworthy digital financial operations.

Literature Review

Research on blockchain analytics has grown extensively as decentralized networks continue to produce complex and high-volume transactional datasets. Studies show that although blockchain ensures transparency through an immutable ledger, the pseudonymous structure of wallet addresses introduces analytical challenges that require advanced computational approaches [9]. Earlier works emphasize that blockchain data embodies temporal attributes, transaction frequency, block generation behaviour, staking metrics, and interaction patterns that can be leveraged for behavioural modelling when processed with suitable analytical frameworks [10], [11].

Machine learning has been widely acknowledged as a powerful analytical tool for investigating blockchain activity. While supervised learning has shown strong potential in fraud classification, phishing detection, and malicious address identification, many blockchain datasets remain unlabelled, making unsupervised learning more applicable in real-world settings [12]. Several studies demonstrate that clustering can reveal hidden structures within blockchain transactions, allowing analysts to uncover abnormal behaviour and wallet interactions without relying on predefined labels [13], [14].

K-means clustering is one of the most widely used partitioning techniques in blockchain analytics. Prior research applied K-Means to categorize cryptocurrency users, segment smart contract interactions, identify miner behaviour patterns, analyse decentralized finance flows, and study changes in transaction volume over time [15], [16]. The algorithm has proven effective when the underlying data exhibits clear separable groups with relatively spherical distributions. However, K-Means is sensitive to outliers and struggles with datasets containing irregular or non-uniform cluster shapes, which frequently occur in blockchain environments [17].

Density-based clustering methods offer a complementary advantage by identifying natural groupings based on data density rather than geometric distance. DBSCAN is particularly notable because it does not require the number of clusters to be specified in advance and can classify isolated observations as noise. Several studies highlight that DBSCAN performs well in fraud analytics, wallet anomaly detection, spammer activity identification, decentralized exchange manipulation detection, and automated bot activity recognition [18]. DBSCAN also uncovers complex microclusters that centroid-based methods fail to detect, making it valuable for analysing blockchain datasets with uneven transaction density or erratic user behaviour [19].

Comparative studies in blockchain analytics show that combining centroid-based and density-based clustering produces a more holistic understanding of behavioural structures. Such approaches have been applied to analyse token transfer networks, evaluate miner centralization, detect transaction wash patterns in NFT marketplaces, group validator behaviours in Proof of Stake systems, and monitor network health over time [20]. These findings support the idea that hybrid clustering frameworks offer improved performance in identifying meaningful behavioural segments and anomalies like transaction patterns in decentralized ecosystems.

Building upon these foundations, the present study applies both K Means and DBSCAN to analyse 303 blockchain wallets using key features that capture

operational behaviour, temporal progression, staking characteristics, and block formation dynamics. By integrating the strengths of both clustering approaches, this research aims to identify stable behavioural groups, detect irregular wallet activity, and provide a comprehensive analytical basis for blockchain risk profiling and address segmentation.

Methods

Dataset and features

The dataset contains 303 wallet records with numeric and categorical attributes that describe block-level and wallet-level behaviour. Representative numeric features include BlockHeight, UnixTimestamp, TxnFee, Block Density, Block Score, Stake Distribution Rate, Coin Day Weight, Coin Age, Coin Stake, and Txnsize. Categorical features include Node Label and Status Tag. All numeric features were used in the clustering pipeline after preprocessing and scaling. Categorical features were encoded when appropriate.

Preprocessing

Missing numeric values were imputed using the column median. Formally, for a numeric variable x with observed values $x_1, \ldots, x_m$ and a median $\mathrm{med}(x)$, each missing entry x_j was replaced by $\mathrm{med}(x)$. The imputation rule can be written as:

$$x_j' = \begin{cases} x_j & if\ x_j\ is\ observed, \\ med(x) & if\ x_j\ is\ missing \end{cases} \quad (1)$$

Categorical variables with a reasonable number of distinct values were converted to one-hot encoded columns. Columns with extremely high cardinality were encoded by frequency rank mapping.

Scaling and dimensionality reduction

All numeric features were standardized using z-score normalization. For a feature vector $x = (x_1, \ldots, x_n)$with mean μ and standard deviation σ, the standardized value z_i is:

$$z_i = \frac{x_i - \mu}{\sigma}. \quad (2)$$

Principal component analysis was applied to standardized features for visualization and optional dimensionality reduction. Let X be the $n \times p$ matrix of standardized features. The sample covariance matrix is:

$$\Sigma = \frac{1}{n} X^T X. \quad (3)$$

Principal components correspond to eigenvectors w_k and eigenvalues λ_k satisfying:

$$\Sigma w_k = \lambda_k w_k. \quad (4)$$

The projection of observations onto the first two principal components is:

$$Z = XW_2. \quad (5)$$

W_2 is the $p \times 2$ matrix whose columns are the first two eigenvectors.

Clustering algorithms

K-Means clustering partitions the dataset into K clusters by minimizing the sum

of squared Euclidean distances between observations and their assigned cluster centroids. Given a partition $\{C_1, \ldots, C_K\}$ with centroids $\mu_1, \ldots, \mu_K$, the objective function is:

$$J = \sum_{k=1}^{K} \sum_{x \in C_k} \lVert x - \mu_k \rVert_2^2. \quad (6)$$

Centroids are updated as the arithmetic mean of points assigned to each cluster:

$$\mu_k = \frac{1}{|C_k|} \sum_{x \in C_k} x. \quad (7)$$

The Euclidean distance used in assignments is:

$$\|x - \mu\|_2 = \sqrt{\sum_{j=1}^{p} (x_j - \mu_j)^2}. \quad (8)$$

K-Means was run for K values from 2 to 6. Each run used multiple random initializations and the solution with the lowest objective was retained. The optimal K was selected based on the silhouette coefficient described below. DBSCAN clusters points based on local density and requires two parameters, epsilon ε and minimum points minPts. For a point, p the ε-neighborhood is defined by:

$$N_\varepsilon(p) = \{q \| p - q \|_2 \leq \varepsilon\}. \quad (9)$$

DBSCAN was evaluated for ε values in the set $\{0.5, 1.0, 1.5, 2.0\}$ with minPts $= 5$. The chosen configuration is reported in the Results section.

Model selection and cluster validity

The silhouette coefficient for an observation i measures how similar i is to its own cluster compared to the next nearest cluster. Let $a(i)$ be the average distance from i to all other points in the same cluster and let $b(i)$ be the minimum average distance from i to points in any other cluster. The silhouette score for i is:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}. \quad (10)$$

The Calinski-Harabasz index evaluates between-cluster variance relative to within-cluster variance. Let n be the total number of observations and K the number of clusters. Define the difference between cluster dispersion B and within-cluster dispersion W. The Calinski-Harabasz score is Higher values indicate better separation.

$$\mathrm{CH} = \frac{B/(K-1)}{W/(n-K)}. \quad (11)$$

Model selection for K Means used silhouette as the primary criterion and Calinski-Harabasz as the secondary confirmation. DBSCAN parameter selection considered the number of clusters, the proportion of noise points, and domain interpretability.

Postprocessing and cluster interpretation

Cluster centroids from K Means were inverse transformed to the original scale for interpretation. For each cluster, the top distinguishing features were identified by computing the absolute deviation of cluster mean from the global mean and ranking features by that deviation. For a cluster C_k and feature j with

cluster mean $\bar{x}_{k,j}$ and global mean $\bar{x}_j$, the absolute deviation is:

$$d_{k,j} = \lvert \overline{x_{k,j}} - \overline{x_j} \rvert. \quad (12)$$

DBSCAN noise points were flagged for further manual inspection and used to support anomaly discussion.

Implementation details

All preprocessing (figure 1), clustering, and analysis were implemented in Python using pandas for data handling, scikit learn for preprocessing, PCA, K Means, DBSCAN, and evaluation metrics, and matplotlib for visualization. K Means was run with ten random initializations to reduce sensitivity to centroid initialization. Results and plots were saved for inclusion in the paper.

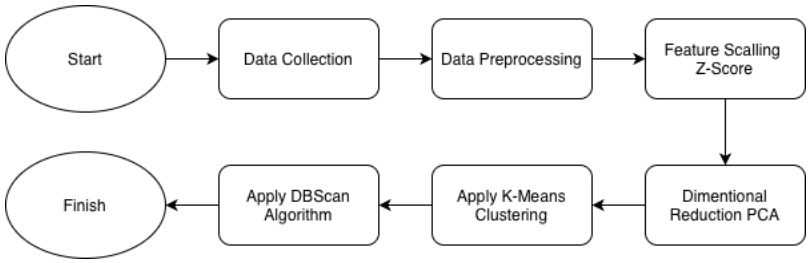

Figure 1 Research Step

Result

This section presents the findings derived from applying K-Means and DBSCAN clustering to the blockchain wallet dataset. The results include numerical summaries, visual analyses using PCA, and behavioural interpretation of the identified clusters.

K-Means Clustering Results

The silhouette analysis indicated that the optimal number of clusters for K-Means was k = 3, revealing three distinct behavioural groups. The distribution of wallets in each cluster is shown in table 1.

Table 1 K-Means Cluster Counts

Cluster ID	Number of Wallets
0	3
1	200
2	100

A Principal Component Analysis (PCA) projection was used to visualize the cluster separation. Figure 2 clearly shows that Cluster 1 and Cluster 2 form compact, well-defined groups, while Cluster 0 appears far from the main clusters, indicating highly unusual wallet behaviour.

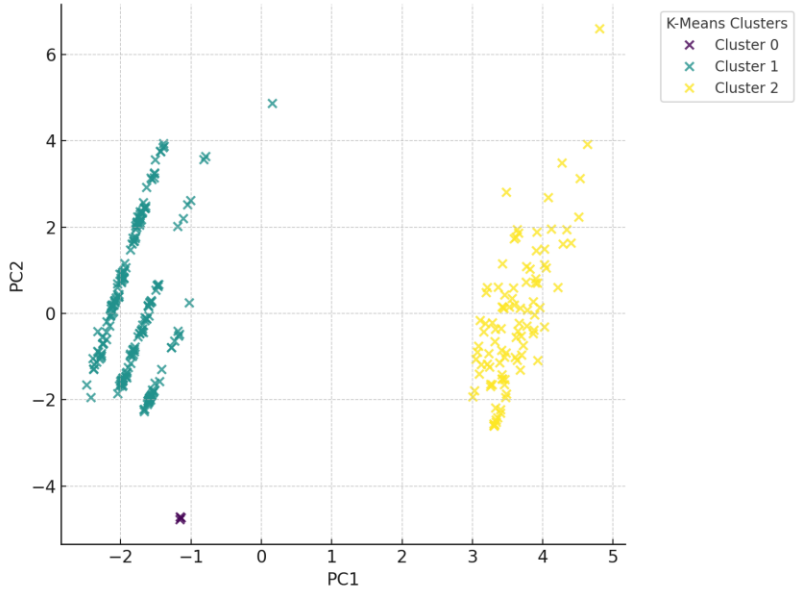

Figure 2 K-Means Clustering in PCA Space

Cluster 1 dominates the dataset and forms the densest region. Cluster 2 occupies a separate but structured region. Cluster 0 forms isolated points, suggesting extremely rare or anomalous blockchain behaviour.

Numerical Characteristics of Clusters

To understand cluster behaviours, mean values of the numerical blockchain attributes were computed for each cluster. These averages are presented in Table 2.

Table 2 Mean Values of Numerical Features per Cluster

Cluster	Block Height	UnixTimestamp	Block Density (%)	Coin Day Weight	Node Label
0	10,900,300	1,600,620,000	60.00	30.495	0
1	5,508,330	1,524,740,000	19.35	0.000	0
2	15,450,200	1,662,000,000	48.90	1.000	1

Cluster 0 exhibits extreme values in several metrics, particularly Block Density and Coin Day Weight. Cluster 1 reflects routine and stable behaviour, while Cluster 2 displays higher block-related and temporal metrics, suggesting more active or high-stakes wallets.

Distinguishing Feature Analysis

Deviation from the global mean was computed to identify the variables that most strongly distinguish the clusters. Table 3 presents the five strongest differentiating features for each cluster.

Table 3 Top Five Distinguishing Features per Cluster

Cluster	Feature	Deviation from Global Mean
0	UnixTimestamp	29,830,100
	Block Height	2,057,430

	Stake Distribution Rate	603.795
	Block Density (%)	392.941
	Coin Day Weight	30.495
	UnixTimestamp	46,052,500
	BlockHeight	3,334,540
1	Block Score	310.901
	Stake Distribution Rate	69.712
	Block Density (%)	33.628
	UnixTimestamp	91,210,200
	Block Height	6,607,360
2	Block Score	622.669
	Stake Distribution Rate	157.538
	Block Density (%)	55.467

These metrics reveal strong temporal and structural differences between clusters, with Cluster 0 being the most extreme outlier group.

DBSCAN Clustering Results

DBSCAN with eps = 1.0 produced several small clusters along with a large noise group, revealing wallet behaviours not captured by K-Means. The PCA visualization for DBSCAN is shown in figure 3.

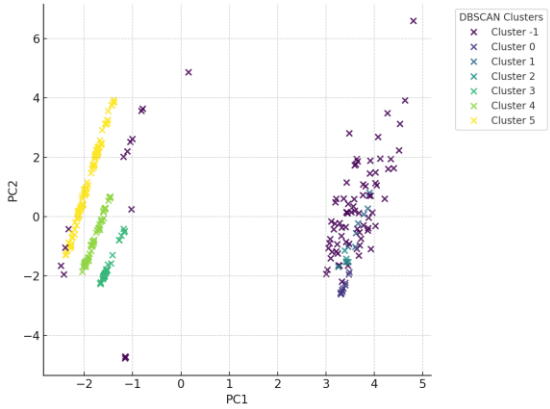

Figure 3 DBSCAN Clustering in PCA Space

DBSCAN isolates many wallets as noise, indicating that a significant portion of the dataset exhibits irregular, sparse, or anomaly-like behaviour. This further strengthens the insight that blockchain networks include rare but important wallet activity patterns.

Summary of Behavioural Structure

The combined evidence from K-Means and DBSCAN demonstrates that the dataset contains clear behavioural groups as well as extensive irregular activity. K-Means effectively identifies the dominant wallet archetypes, while DBSCAN highlights anomaly-like patterns. Together, these findings form a strong basis for blockchain wallet segmentation, risk profiling, and anomaly detection.

Discussion

The results of this study demonstrate that clustering techniques, specifically K-Means and DBSCAN, are effective tools for uncovering hidden behavioural structures within blockchain wallet activity. The three clusters identified by K-Means reveal the existence of distinct operational patterns among wallets, ranging from routine and predictable behaviour to highly unusual and anomalous profiles. Cluster 1, which contains the majority of observations, represents the dominant behavioural group in the dataset. The numerical summary suggests that wallets in this cluster exhibit moderate BlockHeight, stable Block Density (%), and minimal Coin Day Weight, indicating transactional activity that aligns with typical user behaviour in a blockchain environment.

Cluster 2, the second-largest group, displays higher values in BlockHeight, UnixTimestamp, and Block Density (%), reflecting more recent engagement and potentially more active or sophisticated wallet operations. The elevated Block Score and Stake Distribution Rate in this cluster suggest involvement in staking, governance, or higher-volume transactional behaviours. Compared to Cluster 1, Cluster 2 exhibits stronger temporal and structural signatures, aligning with wallets that engage more frequently or with larger transaction volumes.

Cluster 0, comprising only three wallets, forms a highly distinct group separated from all others, as confirmed by the PCA visualization. The extremely high deviations in features such as Block Density (%), Stake Distribution Rate, and Coin Day Weight indicate atypical or abnormal behavior that differs substantially from the main clusters. These wallets may represent long-term holders, automated scripts, early mining participants, or high-value actors whose operational patterns fall outside common blockchain behaviours. Their isolation in the PCA plot further supports the interpretation that these were outlier or edge-case behavioural instances.

DBSCAN offers a complementary perspective by identifying not only dense clusters but also a substantial number of noise points. The large presence of noise in the DBSCAN results suggests that blockchain datasets often exhibit irregular or sparse behavioural patterns that cannot be captured directly by centroid-based clustering. These noisy observations may represent sporadic transactions, incomplete behaviour patterns, or entities performing activities outside typical blockchain norms. In practical terms, these noise points are highly relevant for anomaly detection, fraud monitoring, or tracing unusual wallet activity, since DBSCAN uniquely labels them as non-conforming data.

The comparison between the two clustering algorithms highlights the strengths and limitations of each method. K-Means excels at defining broad and stable groups with clear internal cohesion, making it suitable for segmentation and general behavioural categorization. DBSCAN, by contrast, is particularly valuable for identifying micro-patterns and irregularities, enabling the detection of wallets that do not fit into any dominant behavioural group. When used together, these algorithms provide a more complete understanding of blockchain activity: K-Means maps the primary behavioural segments, while DBSCAN exposes hidden anomalies and rare operational signatures.

The inclusion of Figures 2 and 3 further supports the interpretive analysis. Figure 2 illustrates the distinct geometric separation between the three K-Means clusters, confirming that the algorithm successfully identifies meaningful behavioural patterns in the dataset. Figure 3, on the other hand, reveals the

scattered distribution of noise points detected by DBSCAN and the presence of several smaller clusters, highlighting the algorithm's sensitivity to irregular transaction behaviour.

Overall, the findings underscore the usefulness of clustering for blockchain analytics, particularly for risk profiling, anomaly detection, and wallet segmentation. The clear separation between high-density behavioural groups and anomalous patterns indicates that unsupervised learning methods can successfully model underlying structures in decentralized transaction environments. These insights have practical implications for monitoring blockchain ecosystems, identifying potential fraud, and supporting regulatory or security-related decision-making.

Conclusion

This study demonstrates that unsupervised machine learning techniques, particularly K-Means and DBSCAN, offer valuable insights into the behavioural structure of blockchain wallets. Through the application of K-Means clustering, three distinct groups of wallet activity were identified, each representing different operational characteristics within the blockchain ecosystem. Cluster 1 was found to represent routine and typical wallet behaviour, Cluster 2 reflected more active or high-stakes engagements, and Cluster 0 captured highly unusual or extreme activity patterns. These differences were supported by substantial variations in key metrics such as BlockHeight, UnixTimestamp, Block Density (%), Coin Day Weight, and Stake Distribution Rate.

DBSCAN complemented these findings by revealing that a significant portion of wallets exhibited irregular or sparse activity patterns that did not conform to the dense behavioural groups identified by K-Means. The identification of numerous noise points and micro-clusters highlights the complexity and heterogeneity of blockchain activity, emphasizing the importance of density-based methods for capturing atypical transaction behaviours. The PCA visualizations further confirmed the meaningful separation between clusters and the presence of outlier wallet activity.

Taken together, the results show that combining centroid-based and density-based clustering provides a multidimensional view of blockchain transaction dynamics. K-Means effectively segments dominant behavioural profiles, while DBSCAN uncovers rare, potentially suspicious, or non-standard wallet activity. This dual approach offers strong potential for use in blockchain risk profiling, anomaly detection, wallet classification, and security monitoring efforts. The findings underscore the capability of unsupervised learning to identify meaningful patterns in decentralized environments where labelled data is scarce or unavailable.

Future research may extend this work by incorporating temporal sequence modelling, network graph analysis, or hybrid clustering techniques to capture even richer patterns in wallet interactions. Additionally, validating cluster interpretations using external threat intelligence sources or expert-labelled datasets may enhance the robustness of the derived insights. Overall, this study provides a solid foundation for the application of machine learning-based clustering in enhancing blockchain analytics and security intelligence.

Declarations

Author Contributions

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

Data Availability Statement

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

Funding

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

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Underwood, “Blockchain Beyond Bitcoin,” *Communications of the ACM*, vol. 59, no. 11, pp. 15–17, 2016, doi: 10.1145/2994581.

[2] M. Zachariadis, G. Hileman, and S. V. Scott, “Governance and Control in Distributed Ledgers: Understanding the Challenges Facing Blockchain Technology in Financial Services,” *Information and Organization*, vol. 29, no. 2, pp. 105–117, Jun. 2019, doi: 10.1016/j.infoandorg.2019.03.001.

[3] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, “An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends,” in *IEEE International Congress on Big Data*, vol. 1, Sep. 2017, pp. 557–564, doi: 10.1109/BigDataCongress.2017.85.

[4] D. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander, “Where Is Current Research on Blockchain Technology? A Systematic Review,” *PLOS ONE*, vol. 11, no. 10, pp. 1–27, 2016, doi: 10.1371/journal.pone.0163477.

[5] A. S. Bahurmuz, “Temporal Analysis of Ethereum Blockchain Trends in Transaction Fees and Block Density Over Time,” *Journal of Current Research in Blockchain*, vol. 2, no. 4, pp. 258–273, Dec. 2025, doi: 10.47738/jcrb.v2i4.48.

[6] N. Kshetri, “Blockchain’s Roles in Meeting Key Supply Chain Management Objectives,” *International Journal of Information Management*, vol. 39, no. April, pp. 80–89, 2018, doi: 10.1016/j.ijinfomgt.2017.12.005.

[7] M. Chaudhry et al., “A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective,” *Symmetry*, vol. 15, no. 9, p. 1679, Aug. 2023, doi: 10.3390/sym15091679.

[8] E. Androulaki et al., “Evaluating User Privacy in Bitcoin,” in *Financial Cryptography and Data Security*, vol. 7859, pp. 34–51, 2013, doi: 10.1007/978-3-642-39884-1_4.

[9] M. Conti, S. Kumar, C. Lal, and S. Ruj, “A Survey on Security and Privacy Issues of Bitcoin,” *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3416–3452, 2018, doi: 10.1109/COMST.2018.2842460.

[10] N. Tiwari, P. Ranjan, P. K. Biswal, C. Barde, and N. Sinha, “Survey and Analysis of Blockchain Technologies with Respect to: Properties, Algorithms, Architecture, Models, Evolution and Framework,” *Multimedia Tools and Applications*, vol. 84, no. 14, pp. 13571–13615, Jun. 2024, doi: 10.1007/s11042-024-19580-3.

[11] P. Azad, C. G. Akcora, and A. Khan, “Machine Learning for Blockchain Data Analysis: Progress and Opportunities,” *Distributed Ledger Technologies: Research and Practice*, vol. 5, no. 1, pp. 1–27, Dec. 2025, doi: 10.1145/3728474.

[12] H. Farrukh, S. Zafar, Z. U. Rehman, A. A. Shah, and N. Alshammry, “Blockchain-Based Fraud Detection: A Comparative Systematic Literature Review of Federated Learning and Machine Learning Approaches,” *Electronics*, vol. 14, no. 24, p. 4952, Dec. 2025, doi: 10.3390/electronics14244952.

[13] J. S. Tharani et al., “Identifying Suspicious Blockchain Transactions Using Clustering With Explainability,” *Distributed Ledger Technologies: Research and Practice*, Oct. 2025, doi: 10.1145/3773277.

[14] 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, May 2024, doi: 10.3390/a17050201.

[15] G. Vlahavas, K. Karasavvas, and A. Vakali, “Unsupervised Clustering of Bitcoin Transactions,” *Financial Innovation*, vol. 10, no. 1, Jan. 2024, doi: 10.1186/s40854-023-00525-y.

[16] A. T. Aspembitova, L. Feng, and L. Y. Chew, “Behavioral Structure of Users in Cryptocurrency Market,” *PLOS ONE*, vol. 16, no. 1, Jan. 2021, doi: 10.1371/journal.pone.0242600.

[17] J. Wang et al., “User Segmentation Under Blockchain-Based Privacy Protection,” *Journal of Economy and Technology*, vol. 4, pp. 307–322, 2026, doi: 10.1016/j.ject.2025.09.002.

[18] T. Hariguna, “Unsupervised Anomaly Detection in Digital Currency Trading: A Clustering and Density-Based Approach Using Bitcoin Data,” *Journal of Current Research in Blockchain*, vol. 1, no. 1, pp. 70–90, Jun. 2024, doi: 10.47738/jcrb.v1i1.12.

[19] J. P. Saputra, “Analysis of Blockchain Transaction Patterns in the Metaverse Using Clustering Techniques,” *Journal of Current Research in Blockchain*, vol. 1, no. 1, pp. 33–47, Jun. 2024, doi: 10.47738/jcrb.v1i1.10.

[20] M. Rasolroveicy and M. Fokaefs, “IntelliChain: An Intelligent and Adaptive Framework for Decentralized Applications on Public Blockchain Technologies: An NFT Marketplace Case Study,” *IEEE Transactions on Reliability*, vol. 74, no. 3, pp. 3192–3205, Sept. 2025, doi: 10.1109/TR.2024.3451964.