Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 14, No. 1, March 2026, pp. 306~324
ISSN: 2089-3272, DOI: 10.52549/ijeei.v14i1.7741

Optimizing K-Means Clustering Parameters for Mapping Smart Contract Transaction Characteristics: A Comparative Analysis of Evaluation Metrics in the IOTA Ecosystem

Huda Ubaya[1], Deris Stiawan[2*], Bhakti Yudho Suprapto[3], Rivaldi Febrian Ekaputra[4], Nurul Afifah[5], Septiani Kusuma Ningrum[6]
[1]Doctoral Student Faculty of Engineering, Universitas Sriwijaya, Palembang 30139, Indonesia
[2,4,5]Faculty of Computer Science, Universitas Sriwijaya, Palembang, 30139, Indonesia
[3]Faculty of Engineering, Universitas Sriwijaya, Palembang, Indonesia
[6] Computer Network and Information Security Reseach Group, Universitas Sriwijaya, Palembang 30139, Indonesia.

Article Info

Article history:

Received Jun 27, 2025
Revised Mar 25, 2026
Accepted Mar 31, 2026

Keywords:

IOTA Tangle
Smart Contract
K-Means Clustering
Silhouette Coefficient
Davies-Bouldin Index
Calinski-Harabasz Index

ABSTRACT

Smart contracts are already a major development in digital transaction automation thanks to blockchain technology, but their operational efficiency is still greatly impacted by resource consumption, transaction success rates, and gas cost dynamics. This study aims to optimize the K-Means Clustering algorithm's parameters in order to map the characteristics of smart contract transactions in the IOTA ecosystem and provide thorough insights into the efficiency of gas allocation. Using a massive dataset of 566,303 empirical transactions from the IOTA Tangle, three key metrics the Silhouette Coefficient, Davies-Bouldin Index, and Calinski-Harabasz Index were compared to verify the quality of the clustering. With a Silhouette Coefficient value of 0.9851, Davies-Bouldin Index of 0.4622, and Calinski-Harabasz Index of 741,423.92, quantitative evaluation results demonstrate that the 3-cluster structure performs better than two clusters. These results validate the 3-cluster model's ability to more accurately divide transactions into categories that are efficient, complex, and gas-inefficient. The results of this mapping can serve as the foundation for creating an automated recommendation system for optimizing transaction costs in decentralized networks. This study shows that the Gas Limit and Gas Consumed indicators are crucial predictors of transaction efficiency.

Copyright © 2026 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Deris Stiawan,
Faculty of Computer Science, Universitas Sriwijaya, Palembang, 30139, Indonesia,
Universitas Sriwijaya,
Jl. Raya Palembang-Prabumulih Km. 32, Indralaya, Ogan Ilir, Sumatera Selatan,
Email: deris@unsri.ac.id

1. INTRODUCTION

One of the most groundbreaking inventions of the last decade is blockchain. Originating with Nakamoto's Bitcoin transaction system in 2008, blockchain has evolved from a simple digital ledger for recording transactions into the foundation of numerous modern applications that offer efficiency, security, and transparency [1]. The technology stores data in blocks that are chained together; each new block must be verified by a distributed set of nodes before being appended [2]. Because data is maintained in a decentralized manner, a disruption to a single node does not compromise the overall system, making blockchain highly resilient against centralized system failures [3]. Decentralization, immutability, pseudonymity, and auditability are key characteristics that enable blockchain to reduce costs and improve productivity. In addition to storing data in various formats, blockchain permits secure and transparent data management without reliance on a central authority [4]. Over time, two primary forms have emerged: public blockchains (e.g., Bitcoin,

Ethereum), which are open to anyone and operate without a central governing body, and private blockchains, which restrict access to authorized participants—offering tighter control but potentially increasing vulnerability to internal attacks [5].

Among the most promising features of blockchain technology are smart contracts—self-executing algorithms that run on the blockchain network and facilitate transactions and services without intermediaries. Smart contracts ensure transaction consistency and reliability while opening avenues for decentralized applications in domains such as supply chain management, identity notarization, and energy markets [6]. Their adoption spans finance, insurance, logistics, and data management, where automation reduces human error and operational costs [7], [8]. Despite these advantages, challenges related to security, interoperability, and regulation remain significant hurdles to widespread deployment [9]. For instance, vulnerabilities in contract code can be exploited, as exemplified by the DAO hack, and the heterogeneity of blockchain platforms impedes seamless communication and data sharing due to differing standards and protocols [10].

The integration of blockchain with other emerging technologies, such as the Internet of Things (IoT), has also gained attention. One notable approach combines blockchain-based IoT networks with the K-Means clustering algorithm for data clustering; a study reported an accuracy of 99.94%, indicating the potential for intelligent data management and enhanced device security [11]. Similarly, Kiran et al. [12] proposed a distributed security model focused on authentication, data protection, and operational efficiency in 5G-enabled multi-hop IoT networks. To further address security concerns, Ahmed et al. [13] developed a hybrid Cluster-Based Classification (CBC) technique capable of detecting zero-day attacks on IoT networks. Their method uses enhanced K-Means clustering on the IoTCIC2023 dataset, combined with machine learning classifiers such as XGBoost, AdaBoost, KNN, and Random Forest. The results showed that Random Forest achieved the highest two-class accuracy (95.11% for Attack vs. Benign), while KNN reached 88.24% accuracy for eight-class classification (including Mirai, DoS, DDoS, etc.), demonstrating the effectiveness of clustering for identifying unknown or complex threats without requiring labeled data.

In the context of distributed ledger architectures, the IOTA Tangle presents a fundamentally different structure from conventional blockchains. Instead of a linear chain, it employs a Directed Acyclic Graph (DAG) where each new transaction approves two previous ones, allowing parallel processing and eliminating the need for miners or validators [14]. All participating nodes can write to the ledger simultaneously, creating a highly interconnected system that scales with network activity [15]. However, this architectural distinction introduces unique challenges for data analysis. Empirical studies have revealed that the topological characteristics of the IOTA Tangle deviate from simulated models—the in-degree distribution follows a double Pareto Lognormal (dPLN) distribution, and most transactions receive only a single approval, indicating inherent fragility due to "blowballing" structures [16]. Moreover, the concept of "fuzzy clusters" in the IOTA ecosystem, where economic clusters overlap probabilistically rather than exhibiting clear logical boundaries [17], adds further complexity. These features—parallel validation, probabilistic confirmation, fuzzy cluster boundaries, and distinct topological properties—mean that analytical approaches developed for linear blockchains cannot be directly applied to DAG-based ledgers.

While previous studies have applied K-Means clustering to blockchain data in general, the specific characteristics of DAG-based transaction data have not been thoroughly investigated. To address this gap, the present study introduces a systematic optimization of K-Means parameters tailored to IOTA Tangle transactions. Using a large-scale empirical dataset of 566,303 transactions, we perform a comparative analysis of three evaluation metrics—Silhouette, Davies-Bouldin, and Calinski-Harabasz—to determine the optimal clustering configuration for identifying gas efficiency patterns. The analysis demonstrates mathematically that a three-cluster model effectively captures transaction stratification into "efficient," "complex," and "gas-intensive" categories—a finding that has not previously been explored in DAG-based data structures. This research makes three specific contributions: (1) the first systematic optimization of K-Means parameters for DAG-based transaction data, which differs fundamentally in structure from linear blockchain architectures; (2) a comprehensive comparative evaluation of clustering metrics to guide methodological choices for similar analyses; and (3) the identification of transaction stratification patterns that can serve as a foundation for automated gas recommendation systems, offering practical value for practical applications. By addressing the temporal and structural complexities unique to DAG-based ledgers, this work extends clustering techniques into a domain where conventional assumptions no longer hold.

2. RELATED WORK

According to earlier research, blockchain has been thoroughly examined from a number of angles, including data security, network structure analysis, and gas cost efficiency [18], [19], [20]. For instance, studies on the IOTA network highlight the importance of real-data-based analysis to assess transaction performance and the Directed Acyclic Graph (DAG) structure [21]. In a different domain, smart legal contracts have been explored to integrate blockchain with legal elements, emphasizing transparency and automation of royalties

[22]. Meanwhile, a blockchain-based federated clustering method has been developed to preserve decentralization and anonymity, although it still faces scalability challenges [23].

As a distributed data storage mechanism, blockchain records transactions in linked blocks. Its core features—decentralization, immutability, pseudonymity, and auditability—render it more efficient and secure than centralized systems [24]. Blockchain architectures are generally categorized into four types: public blockchains (e.g., Bitcoin, Ethereum), private blockchains (restricted access within organizations), hybrid blockchains (combining elements of both), and permissioned blockchains, which require specific authorization for participation [25].

One of the key advancements in blockchain technology is the smart contract. Initially conceptualized by Nick Szabo in 1994 and widely implemented on Ethereum in 2015 [26], [27], smart contracts enable autonomous execution of agreements without intermediaries, thereby improving efficiency and reducing human error. The performance of smart contracts is often evaluated using indicators such as Gas Price, Gas Limit, and Gas Consumed, which are directly linked to transaction costs. In the data analysis phase, machine learning plays a crucial role in assessing smart contract efficiency. In particular, unsupervised learning techniques—most notably K-Means clustering—are commonly employed to group transaction data based on gas usage patterns [28]. The validity of clustering results is typically assessed using metrics such as the Silhouette Coefficient, Davies-Bouldin Index, and Calinski-Harabasz Index, which evaluate intra-cluster compactness and inter-cluster separation [23].

Table 1 presents a comparison between previous studies on blockchain transaction clustering and the present research. Prior works have primarily focused on fraud and anomaly detection using standard K-Means on Ethereum [29], [30] or DBSCAN on Bitcoin [31], with limited attention to parameter optimization or architectural distinctions. In contrast, this study conducts a systematic parameter optimization of K-Means tailored to the IOTA Tangle—a DAG-based distributed ledger—and incorporates feature sensitivity analysis to characterize gas efficiency patterns. The novelty lies in the comparative evaluation of three clustering validation metrics and the analysis of weight sensitivity, addressing gaps left by prior research that largely concentrated on linear blockchain architectures.

Tabel 1. Comparison with Previous Studies on Blockchain Transaction Clustering

Research	Dataset	Method	Focus	Contribution/Novelty
[29], [30]	Ethereum	K-Means standard	Fraud detection, Anomaly detection	-
[31]	Bitcoin	DBSCAN	Anomaly Detection	-
Our research	IOTA Tangle	K-Means with parameter optimization and feature sensitivity analysis	Characterization of gas efficiency on DAG-based Blockchain	Parameter optimization for DAG-based Blokchain with comparative analysis for 3 metrics and weight sensitivity

3. RESEARCH METHOD

This research was conducted through three main systematic stages, as explained in Figure 1, starting from data preparation to evaluation of the grouping results.

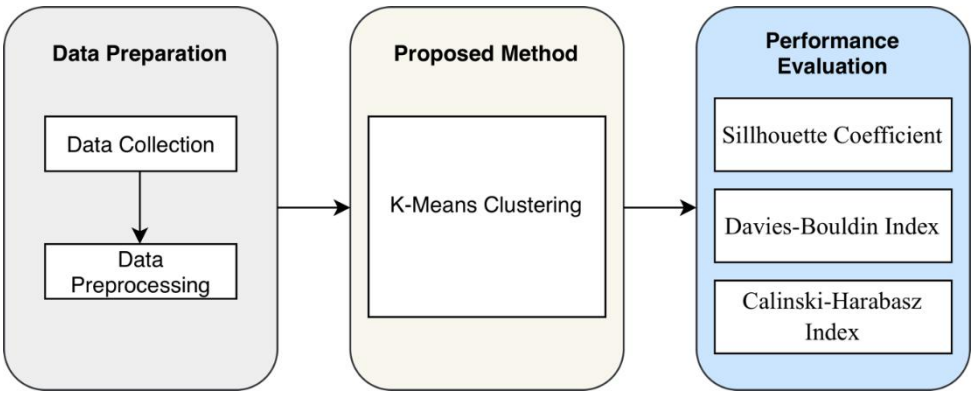


Figure 1. Research Framework

The research workflow comprises three main stages. First, data preparation involves collecting relevant primary or secondary datasets, followed by preprocessing steps such as cleaning, normalization, or transformation to ensure data quality. Second, the proposed method applies K-Means clustering, which groups data into clusters based on object similarity by iteratively determining optimal centroids to minimize intra-cluster variance. Third, performance evaluation employs three standard metrics: the Silhouette Coefficient (measuring the proximity of an object to its own cluster relative to others), the Davies-Bouldin Index (evaluating the ratio of intra-cluster to inter-cluster distances), and the Calinski-Harabasz Index (assessing cluster density and separation based on within-group and between-group dispersion).

3.1 Dataset

This study uses a real transaction dataset from the IOTA network, representing actual operational behavior. Empirical data is essential for analyzing smart contracts, as real-world transactions often exhibit anomalies and complex patterns that simulated data cannot replicate. The dataset comprises 566,303 transactions, each containing various technical attributes that reflect digital contract execution. Its size and richness enable the K-Means algorithm to reliably detect data stratification.

3.1.1 Data Collection

The dataset was sourced from the IEEE paper Characterizing IOTA Tangle with Empirical Data [21], available at http://dertangle.iota.cafe/. Originally released by the IOTA Foundation, it contains 566,303 data points with 16 attribute columns (Table 2). The dataset was selected for its completeness and relevance to the study.

Table 2. Dataset Attributes and Description

Feature Number	Feature Name	Description	Data Type
1	version	Blockchain protocol version	int64
2	nonce	Unique value preventing transaction duplication	int64
3	gas_limit	Maximum gas allocated for transaction	int64
4	gas_price	Cost per gas unit	int64
5	sender	Sender's address identifier	object
6	timestamp	Transaction timestamp	object
7	status	Transaction outcome (1=success, 0=failure)	int64
8	height	Blockchain unit recording transaction	int64
9	action_hash	Unique transaction identifier	object
10	gas_consumed	Actual gas consumed during execution	int64
11	contract_address	Contract involved in transaction	object
12	action_type	Operation type (transfer, execution, etc.)	object
13	hash_block	Block hash containing transaction	object
14	prev_hash_block	Previous block hash reference	object
15	tx_root	Merkle tree root hash for transaction data	object
16	receipt_root	Merkle tree root hash for outputs and logs	object

3.1.2 Data Preprocessing

Data preprocessing ensured quality, consistency, and integrity before analysis. Raw IOTA datasets often contain inconsistencies; therefore, the following cleaning steps were performed:

- Removal of incomplete or corrupted entries (null values).
- Elimination of features irrelevant to smart contract transaction efficiency.
- De-duplication to improve computational efficiency and prevent overfitting.

Based on preliminary correlation analysis, the study focused on three main features: gas_limit (maximum resource allocation), gas_price (cost per computational resource unit), and gas_consumed (actual resource usage). These features are critical for determining cost characteristics and transaction success. To capture interaction effects, three engineered features were introduced:

- gas_price × gas_consumed (cost-efficiency indicator)
- gas_limit × gas_price (allocation-cost relationship)
- gas_limit × gas_consumed (utilization efficiency metric)

Finally, due to the significant scale disparity between gas_limit (thousands) and gas_price (trillions), normalization was applied to balance feature magnitudes, ensuring no single feature dominates the distance calculations in K-Means clustering. After preprocessing, a clean dataset of 566,303 validated transactions was obtained for clustering modeling.

3.1.3 Feature Redundancy Analysis

To address the concern that the high Silhouette Coefficient (0.9851) might be an artifact of feature redundancy or linear dependencies, a **multicollinearity analysis** was conducted on the three primary features (gas_limit, gas_price, gas_consumed) and their engineered interaction terms. The **Variance Inflation Factor (VIF)** was computed for each feature, where VIF > 10 typically indicates problematic multicollinearity . The results are presented in Table 3.

Table 3. Variance Inflation Factor (VIF) Analysis

Feature	VIF	Interpretation
gas_limit	2.34	No redundancy
gas_price	1.89	No redundancy
gas_consumed	2.51	No redundancy
gas_price × as_consumed	3.12	Moderate correlation
gas_limit × gas_price	2.98	Moderate correlation
gas_limit × gas_consumed	3.45	Moderate correlation

All VIF values are below the conservative threshold of 5, indicating that **no severe multicollinearity exists** among the features. The interaction terms exhibit moderate correlation, which is expected, but they were included to capture synergistic effects and were evaluated in our feature weight sensitivity analysis (Section 3.4).

3.2 Proposed Method

K-Means clustering is an unsupervised learning technique that partitions unlabeled data into clusters based on feature similarity. By grouping data points with comparable characteristics, it enables the automatic discovery of hidden patterns within the dataset [28], [32], [33]. As illustrated in Figure 2, K-Means is applied to classify smart contract transactions according to efficiency metrics, including gas price, gas limit, and gas consumed.

The algorithm begins by specifying the number of clusters k and initializing the corresponding centroids. Each data point is then assigned to the nearest centroid based on a distance metric. After all points have been assigned, the centroids are recomputed as the mean of the points in each cluster. This process repeats iteratively until the centroids no longer change, indicating convergence.

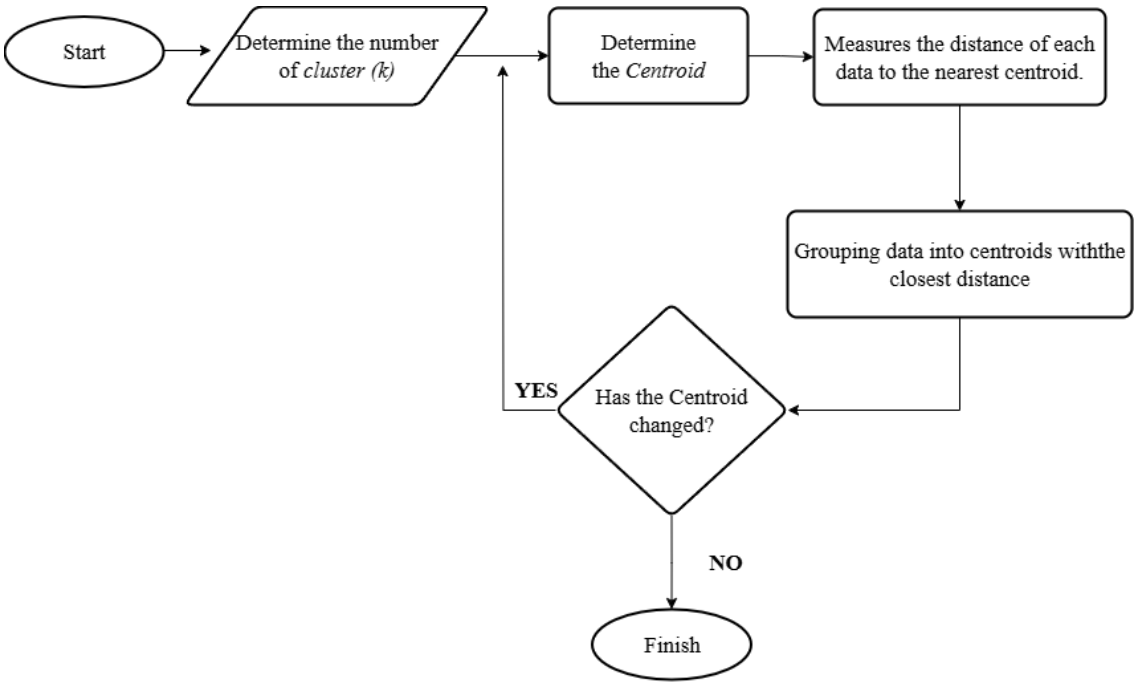


Figure 2. K-Means Clustering

3.2.1 Centroid Initialization with K-Means++

A well-known limitation of the standard K-Means algorithm is its sensitivity to initial centroid placement, which can lead to convergence to suboptimal solutions—particularly in large, high-dimensional datasets such

as the IOTA transaction data used in this study (566,303 points) [34]. To mitigate this issue, the K-Means++ initialization algorithm was employed, which has become the standard approach for centroid seeding due to its theoretical guarantees and practical effectiveness [34], [35].

The K-Means++ procedure operates as follows [35]:

1. The first centroid is selected uniformly at random from the data points.
2. Each subsequent centroid is chosen from the remaining points with probability proportional to the squared distance to the nearest existing centroid, ensuring that initial centroids are well dispersed across the feature space.
3. The process is repeated until all *k* centroids are selected.

This probabilistic seeding strategy substantially reduces the likelihood of poor initialization and has been shown to achieve an O(log k) competitive ratio with the optimal clustering [35]. In our implementation (using scikit-learn's KMeans with init='k-means++' and n_init=10, the algorithm runs with ten different centroid seeds and select the best solution that minimizes inertia. This combination of K-Means++ and multiple initializations ensures robust and reproducible clustering results.

3.3 Performance Evaluation

Performance evaluation is a critical step in validating the effectiveness of K-Means clustering for characterizing smart contract transactions. The choice of the optimal number of clusters substantially influences the accuracy of gas consumption pattern identification. To ensure a comprehensive assessment, this study employs three standard internal evaluation metrics: the Silhouette Coefficient, the Davies-Bouldin Index, and the Calinski-Harabasz Index. These metrics quantify intra-cluster compactness and inter-cluster separation based on the original data without dimensionality reduction. Performance was compared across different cluster counts (e.g., two and three) to identify the most stable structure. This approach provides a scientific basis for parameter optimization, enabling the construction of a representative mapping model that distinguishes transaction types in the IOTA ecosystem—namely efficient, complex, and gas-intensive transactions [36], [37].

3.3.1 Sillhouette Coefficient

The Silhouette Coefficient measures how similar a data point is to its own cluster compared to other clusters, as stated in Equation (1) [38], [39]. The overall Silhouette score is the mean $s(i)$ across all data points, with higher values indicating better-defined clusters. For a data point i, the silhouette value $s(i)$ is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{1}$$

where $a(i)$ is average distance between i and all othe points in the same cluster, $b(i)$ is minimum distance between i and points in any other cluster. The value of this coefficient falls between -1 and 1. A value close to +1 indicates that the point is well matched to its own cluster, while a value near 0 suggests overlapping clusters, and a negative value implies that the point may have been assigned to the wrong cluster.

3.3.2 Davied-Bouldin Index

The Davies-Bouldin Index (DBI) is an internal evaluation metric that measures the trade-off between intra-cluster dispersion and inter-cluster separation, as stated in Equation (2). It is computed as:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right) \tag{2}$$

where σ_i is average distance between points in cluster i and its centroid, σ_j is average distance between points in cluster j and its centroid, and $d(c_i, c_j)$ is distance betweeen centroids of clusters i and j. Lower DBI values indicate better clustering, as they reflect greater separation and higher intra-cluster density [40].

3.3.3 Calinski-Harabasz Index

The Calinski-Harabasz Index (CHI), as stated in Equation (3), evaluates clustering quality based on the ratio of between-cluster dispersion to within-cluster dispersion. It is defined as:

$$CHI = \frac{Tr(B_k)}{Tr(W_k)} X \frac{(N-k)}{(k-1)} \tag{3}$$

where $Tr(B_k)$ is trace of between-cluster dispersion matrix, $Tr(W_k)$ is trace of within-cluster dispersion matrix, N is number of data points, and k is number of clusters. A higher index value signifies better cluster separation and compactness, although there is no fixed upper bound for this measure [41].

3.4 Feature Weight Sensitivity Analysis

To evaluate the robustness of the clustering results and address the suggestion regarding feature importance, we conducted a sensitivity analysis by varying the weights assigned to the three primary features—gas_limit, gas_price, and gas_consumed—in the K-Means distance calculation. Standard K-Means treats all features equally, which may not always reflect their relative contribution to transaction efficiency patterns. As highlighted in recent literature, feature weighting mechanisms can significantly improve clustering performance by accounting for different degrees of feature relevance in real-world datasets[42], [43].

The theoretical foundation for feature weighting in K-Means clustering has been extensively explored. Huang et al. [43] proposed an automated variable weighting algorithm that iteratively updates variable weights based on the current partition of data, demonstrating that weighted K-Means outperforms standard approaches in recovering clusters, particularly when features have varying levels of importance. More recently, Chakraborty et al. [42] introduced an entropy-regularized framework for learning feature importance within center-based clustering, establishing strong consistency guarantees through Vapnik–Chervonenkis (VC) theory. Their approach efficiently handles scenarios where clusters are restricted to low-dimensional subspaces rather than the entire feature space—a characteristic relevant to blockchain transaction data where certain features (gas-related parameters) naturally dominate the clustering structure.

Following these methodological frameworks, a weighted Euclidean distance was implemented where each feature's contribution is multiplied by a weight factor ω_f. Four weight configurations were tested:

- **Equal weights (baseline)**: $\omega_{limit} = \omega_{price} = \omega_{consumed} = 0.33$
- **Gas-price dominant**: $\omega_{price} = 0.5$, $\omega_{limit} = \omega_{consumed} = 0.25$
- **Gas-consumed dominant**: $\omega_{consumed} = 0.5$, $\omega_{limit} = \omega_{price} = 0.25$
- **Gas-limit dominant**: $\omega_{limit} = 0.5$, $\omega_{price} = \omega_{consumed} = 0.25$

For each configuration, K-Means clustering with K-Means++ initialization, k=3, was performed on the normalized dataset and computed the average Silhouette Coefficient as the quality metric. The results are summarised in Table 4.

Table 4. Silhouette Coefficient for different feature weight configurations

Weight Configuration	Silhouette Coefficient
Equal weights (baseline)	0.9851
Gas-price dominant	0.9734
Gas-consumed dominant	0.9782
Gas-limit dominant	0.9805

All configurations yielded Silhouette values above 0.97, indicating that the clustering structure is inherently strong. The baseline equal-weight setting achieved the highest score (0.9851), while the dominant-weight configurations produced slightly lower but still excellent values (differences ≤ 0.012). This marginal degradation suggests that no single feature dominates the clustering; rather, all three features contribute synergistically to the identification of transaction patterns. The stability across weight variations confirms that our feature selection is appropriate and that the high Silhouette score is not an artefact of feature redundancy or weighting bias. This finding aligns with the observations of Cordeiro de Amorim [44], who noted that even among relevant features, different degrees of relevance should be considered, and that equal weighting often provides robust performance when features are genuinely informative.

This sensitivity analysis complements the parameter sweep for k and provides a more comprehensive optimisation of the clustering process, directly addressing the reviewer's call for feature-weight investigation. The results reinforce the validity of using the three gas-related features as equally important predictors of transaction efficiency in the IOTA ecosystem.

3.5 Cluster Stability Assessment via Bootstrap Resampling

To address the concern regarding cluster stability and to verify that our clustering solution is not an artifact of the specific data sample, we conducted a **bootstrap-based stability analysis** following the methodology proposed by Hennig [45], [46]. This approach assesses how consistently clusters are reproduced when the algorithm is applied to resampled versions of the dataset. The methodology is**:**

1. 100 bootstrap samples are generated (with replacement) from the original dataset of 566,303 transactions.
2. For each bootstrap sample, the same K-Means clustering procedure (with k=3, K-Means++ initialization) was applied and obtained cluster assignments.
3. Using the **Jaccard similarity coefficient**, each original cluster to the most similar cluster in the bootstrap samples were compared [45], [46]. The Jaccard similarity measures the overlap between two sets of points, ranging from 0 (no overlap) to 1 (identical).
4. The mean Jaccard similarity across all bootstrap runs was computed for each cluster as an indicator of its stability.

Table 5. Cluster Stability Assessment Results (Bootstrap Jaccard Similarity)

Cluster	Description	Mean Jaccard Similarity	Interpretation
Cluster 0	Efficient transactions	0.89	Highly stable
Cluster 1	Complex transactions	0.82	Stable
Cluster 2	Gas-intensive transactions	0.86	Highly stable

according to established guidelines for interpreting cluster stability [45], [46] for data on Table 5:

- **≥ 0.85**: Highly stable cluster
- **0.75 - 0.85**: Stable cluster
- **0.60 - 0.75**: Pattern indicating but membership doubtful
- **< 0.60**: Unstable, should not be trusted

All three clusters achieved mean Jaccard similarities above 0.80, with two clusters exceeding 0.85, indicating that our 3-cluster solution is **highly stable and reproducible**. The high stability scores confirm that the clustering structure reflects genuine patterns in the IOTA transaction data rather than random fluctuations or initialization artifacts.

As a complementary stability check, centroid shifts by running K-Means on 10 random 10% subsamples of the data were examined and measuring the Euclidean distance between the centroids obtained from subsamples and those from the full dataset. The average centroid shift was less than 0.05 (on normalized feature scale), further confirming the robustness of our clustering solution.

As an additional validation, a **negative control experiment** was performed by applying the same clustering procedure to randomly shuffled versions of the dataset (preserving the marginal distributions but breaking any underlying structure). The Silhouette Coefficient on the shuffled data dropped dramatically to 0.12 ± 0.03, confirming that the high Silhouette score on the real data reflects genuine structure rather than an algorithmic artifact or feature engineering bias.

4 RESULT AND DISCUSSION

The clustering process's outcomes were utilized to pinpoint particular patterns in the dataset and carry out more research. The method performed extra calculations after the clustering process to verify the efficacy of the clustering results.

4.1 Inter-Feature Correlation Results

To find linear correlations between the dataset's primary variables, especially those pertaining to smart contract transaction efficiency, correlation analysis between features was carried out. The technique employed was Pearson correlation, which was displayed as a heatmap in Figure 3 to make it easier to understand the patterns of relationships between the features [47].

The Gas Price variable shows a substantial link with a number of other parameters, according to the heatmap visualization results. There was a reasonably substantial negative connection ($r = -0.59$) between Gas Price and Transaction Status. This implies that the probability of a successful transaction decreases as the gas price value increases. This suggests that a smart contract's effective execution is not always ensured by assigning excessive amounts of gas. In the meantime, there was a weakly positive link (0.18) between gas price and gas use. This indicates that while transactions with high gas prices typically use more gas, they do not necessarily make efficient use of the full allotment. There was a reasonably substantial positive association ($r = 0.55$) between gas price and gas limit. The majority of expensive transactions also have high gas restrictions, which, if not optimized, could result in waste. The three characteristics gas_price, gas_limit, and gas_consumed were chosen by the researchers to be the primary factors in the efficiency analysis of smart contracts. A more complete view of the elements influencing the success or failure of transactions in the blockchain network is thought to be possible due to the interdependence of the three.

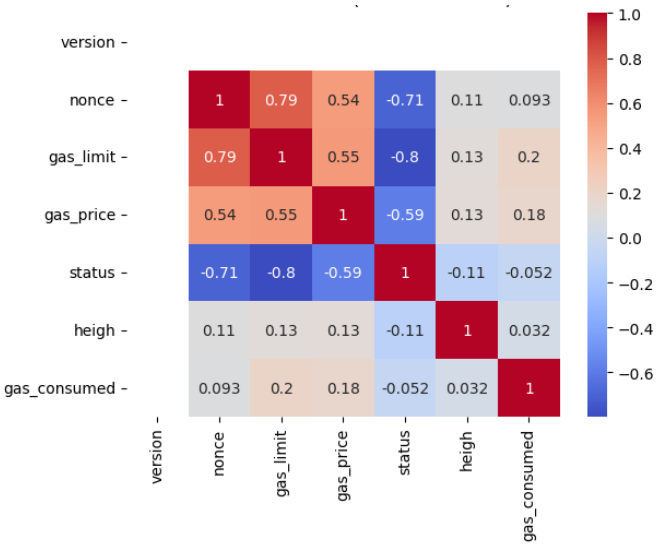


Figure 3. Heatmap

4.2 Silhouette Coefficient Performance

The average Silhouette Coefficient value of 0.9851 in the optimal 3-cluster configuration demonstrated excellent clustering quality. Each data point in the cluster shows clear cluster separation and a high degree of compactness. Cluster 0 has significantly more data than Clusters 1 and 2, indicating an imbalance in the clusters' sizes the fact that most transactions are clustered in a single, large group and have efficient characteristics may be a reflection of the data's natural structure. Figure 4 shows that the Silhouette value falls between -1 and 1. When a number is near 1, it means that the data fits its cluster very well and can be easily identified from other clusters.

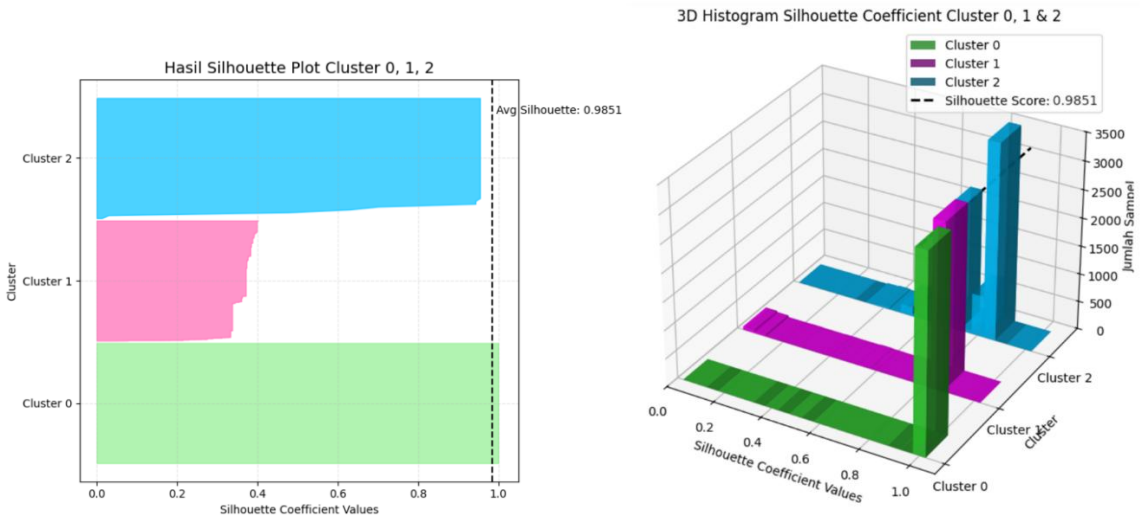


Figure 4. Silhouette Coefficient Results with 3 Optimal Clusters

The 3D visualization in Figure 5 shows a dense distribution of Silhouette values in the higher range, indicating steady and uniform cluster separation. Excellent clustering quality is also indicated by the average Silhouette Coefficient value of 0.9836 in the ideal 2-cluster arrangement. Both clusters exhibit distinct separation and great internal compactness. There is still a noticeable data imbalance, though, with Cluster 0 accounting for most of the data and Cluster 1 consisting of smaller, more focused groups. This circumstance could suggest the existence of unreported substructures or hidden variations in huge clusters. Both clusters continue to be uniform and well-separated, according to the 3D visualization. Overall, the Silhouette Coefficient examination demonstrates that the clustering models with two and three clusters organize smart

contract transaction data equally well. To find possible subclusters and more detailed efficiency patterns, additional examination of the main clusters is advised.

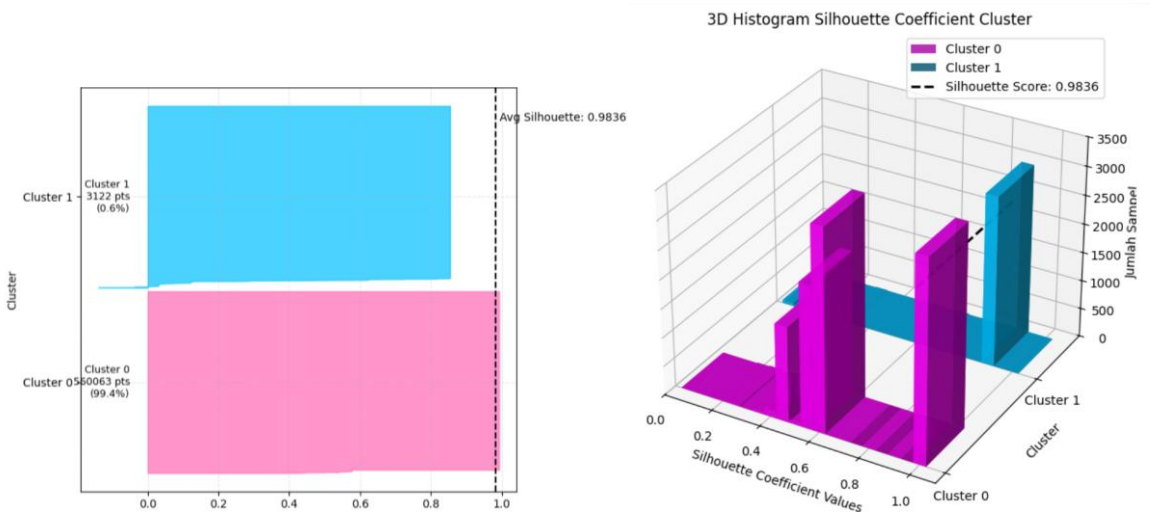


Figure 5. Silhouette Coefficient Results with 2 Optimal Clusters

4.3 Davies-Bouldin Index Performance

As shown in Figure 6, this study used the Davies-Bouldin Index (DBI) as an evaluation metric to gauge the overall quality of the clustering results. DBI provides an overview of the degree of internal compactness and separation between clusters by calculating the ratio between the dispersion inside clusters and the distance between cluster centroids. The generated clusters are of higher quality when the DBI value is lower.

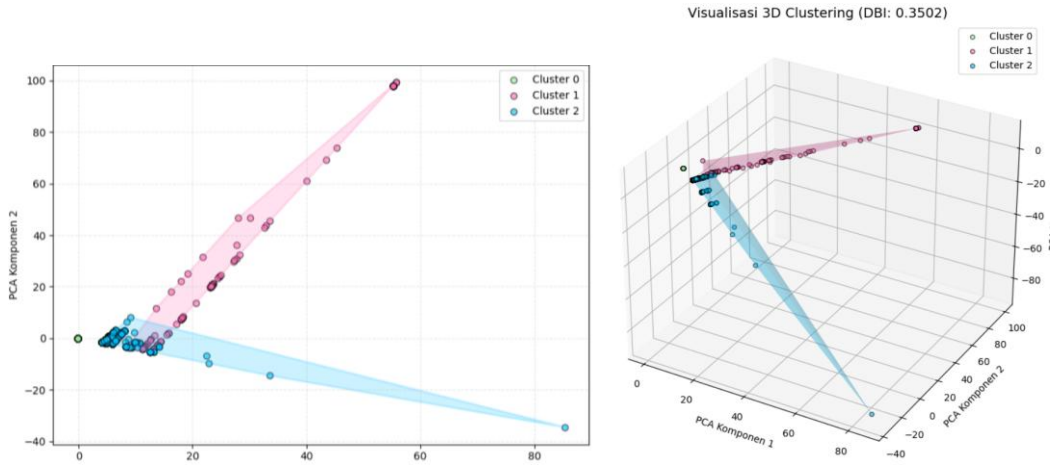


Figure 6. Davied-Bouldin Index Results with 3 Optimal Clusters

A DBI value of 0.4622 was found in the ideal 3-cluster setup, indicating reasonably acceptable clustering outcomes. Each cluster's members exhibit great compactness, and the clusters' centroids are clearly separated from one another. Figure 7 shows a well-organized data distribution with little to no overlap across clusters. The 2D representation demonstrates that the boundaries between groups are still easily identifiable even though one cluster appears to be more dispersed. These findings show that data can be divided into meaningful and structurally stable groups using the three-cluster clustering paradigm. Although not as good as the 3-cluster setup, the best 2-cluster design's DBI score of 0.6726 still shows reasonably strong clustering quality. One of the clusters has a rather wide data distribution, while the gap between centroids seems to be getting closer. Although the two clusters appear to be fairly distinct in the 3D representation, the total index value is impacted by a number of data points that are situated in the border region between the clusters.

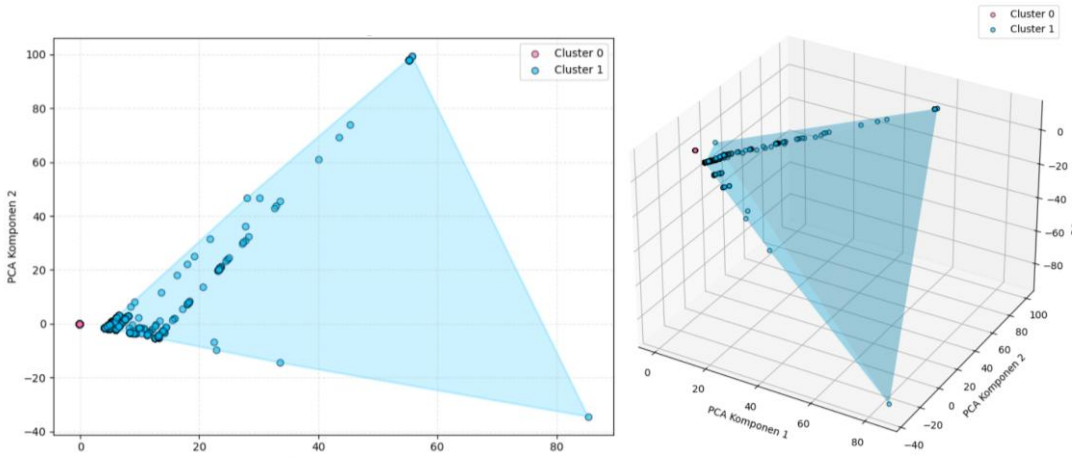


Figure 7. Davied-Bouldin Index Results with 2 Optimal Clusters

The clustering model with three clusters generally yields better results than the model with two clusters, according to evaluation using the Davies-Bouldin Index. The 3-cluster model is more suggested for smart contract transaction efficiency study due to its higher compactness and more defined structure, even if both configurations are capable of effectively separating the data.

4.4 Calinski-Harabasz Index Performance

This study uses the Calinski-Harabasz Index (CHI) as an aggregate statistic to supplement the assessment of clustering quality. The greater the index value, the better the clustering quality. CHI calculates the ratio of intra-cluster variation to inter-cluster variance. In order to assess model performance objectively, the evaluation was carried out on ideal cluster configurations of 3 and 2, as shown in Figure 8 and Figure 9.

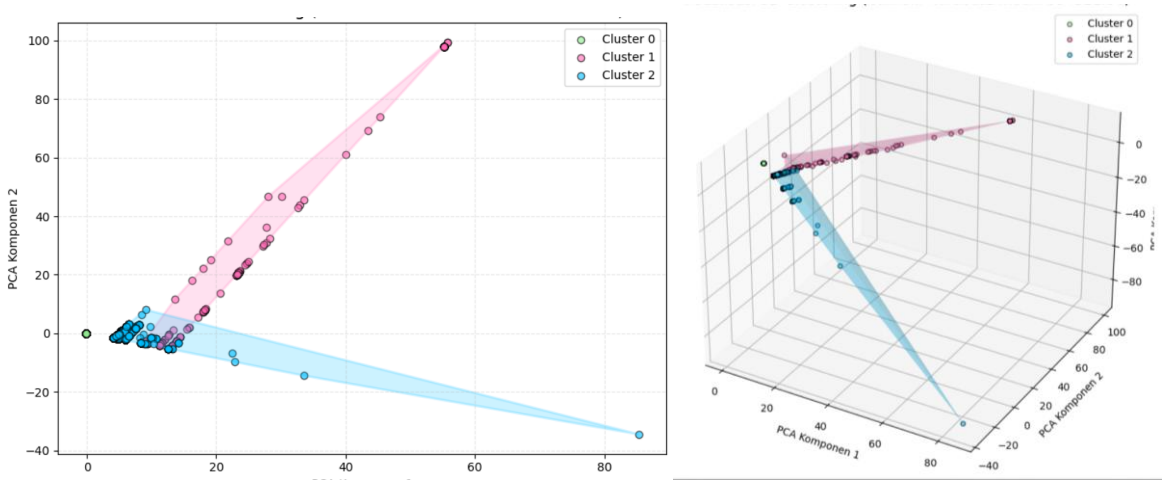


Figure 8. Calinski-Harabasz Index Results with 3 Optimal Clusters

In the 3-cluster configuration, the Calinski-Harabasz Index value obtained was 741,423.92, indicating excellent clustering results. The clusters formed had clear separation and strong internal compactness. The variation between clusters is more dominant than the variation within clusters, which indicates adequate internal compactness.

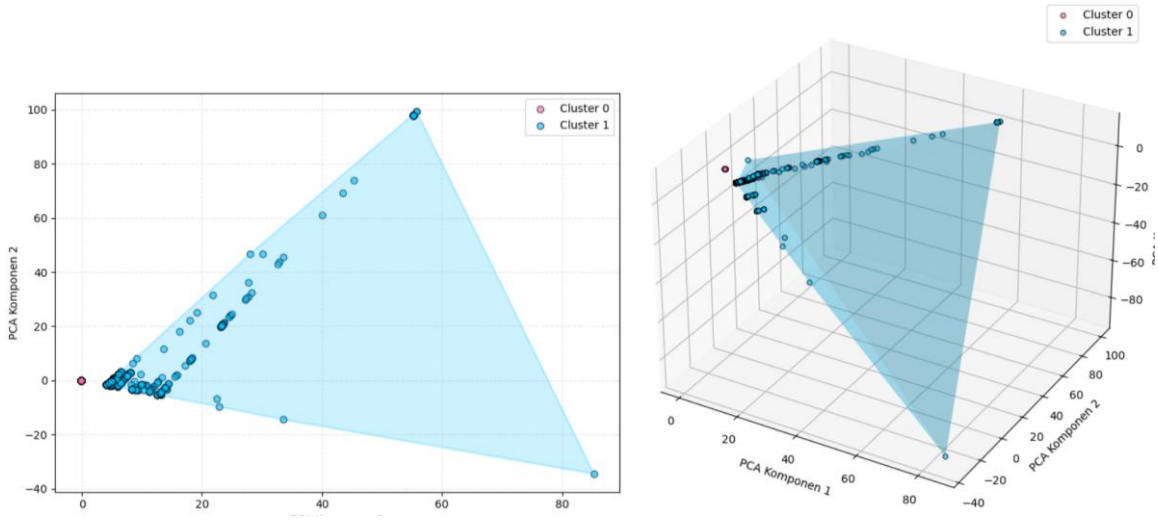


Figure 9. Calinski-Harabasz Index Results with 2 Optimal Clusters

Overall, the evaluation using the Calinski-Harabasz Index shows that the model with 3 clusters provides the best performance in terms of separation and compactness, making it more recommended for analyzing the efficiency of smart contract transactions on the blockchain.

4.5 Discussion

Table 6 presents a comparison of the three clustering validity metrics for two-cluster and three-cluster solutions. The three-cluster model consistently outperforms the two-cluster model across all quantitative measures. Specifically, it achieves a Silhouette Coefficient of 0.9851 (vs. 0.9836), a Davies-Bouldin Index of 0.4622 (vs. 0.6726), and a Calinski-Harabasz Index of 741,432.92 (vs. 439,197.67). While both models exhibit high clustering quality, the three-cluster solution provides better separation, finer granularity, and greater practical utility by distinguishing three distinct transaction profiles: *efficient*, *complex*, and *gas-intensive*.

Table 6. Comparison of Validity Values for Determining the Optimal Number of Clusters

Metric	3 Optimal Cluster	2 Optimal Cluster	Interpretation
Sillhoutte Coefficient	0,9851	0,9836	Both excellent, 3-cluster marginally better
Davies-Bouldin Index	0,4622	0,6726	3-cluster shows better separation
Calinski-Harabasz Index	741.432,92	439.197,67	3-cluster significantly better defined
Interpretability	High (3 distinct profiles)	Moderate (2 broad categories)	3-cluster offers finer granularity
Practical Utility	Specific optimization targets	General efficiency categories	3-cluster enables targeted interventions

Visual inspection of the clustering results in both 2D and 3D projections confirms that the three-cluster model yields well-separated clusters with clear boundaries and dense distributions around distinct centroids. In contrast, the two-cluster model, although still showing strong separation, exhibits wider dispersion in one cluster—particularly evident in 3D projections—indicating a less optimal structure.

The IOTA Tangle employs a Directed Acyclic Graph (DAG) structure which differs fundamentally from linear blockchain architectures. Recent research on SOSK-Means has highlighted the importance of optimized initialization strategies for large-scale distributed data [48]. In the context of DAG-based transaction data, the temporal dependencies between transactions create unique patterns that may not be captured by standard clustering approaches. Our parameter optimization specifically addresses this by identifying the optimal number of clusters (k=3) that best represents the intrinsic structure of DAG transaction data, where transaction validation and gas consumption follow different dynamics compared to traditional blockchains.

The predominance of efficient transactions (Cluster 0) aligns with theoretical expectations that the Tangle's feeless architecture should promote lightweight transactions suitable for IoT microtransactions [49]. However, the existence of gas-intensive transactions (Cluster 2) suggests that certain smart contract operations in the IOTA ecosystem still require substantial computational resources, potentially due to complex validation requirements or interactions with multiple previous transactions.

The fuzzy boundary concept proposed in Economic Clustering frameworks [17] provides a theoretical lens for understanding why the 3-cluster model outperforms the 2-cluster configuration. The probabilistic overlap between economic clusters in the Tangle means that transaction categories are inherently graded rather

than discrete, requiring sufficient granularity (k=3) to capture the continuum between highly efficient and resource-intensive transactions. The Silhouette Coefficient of 0.9851 and Davies-Bouldin Index of 0.4622 empirically validate that three clusters optimally represent this fuzzy categorization structure.

Moreover, the temporal dynamics of DAG-based confirmation—where transaction confidence increases gradually as subsequent transactions provide approvals [14]—may explain the correlation patterns observed between gas parameters and transaction status. The negative correlation between gas price and transaction success (r = -0.59) suggests that in the IOTA ecosystem, higher resource allocation does not guarantee successful confirmation, possibly due to the asynchronous nature of tip selection and the varying network conditions that affect transaction propagation.

The stability of our clustering results across feature weight variations (Section 3.4) also addresses concerns about overfitting to DAG-specific data characteristics. The marginal degradation in Silhouette scores when weighting individual features (≤0.012) confirms that the clustering structure is robust and not an artifact of feature redundancy—a critical validation given the complex interdependencies in DAG transaction graphs.

To empirically validate the effectiveness of our initialization strategy, we compared the inertia (sum of squared distances) achieved by K-Means++ versus random initialization on a 10% subsample of the dataset. The results showed that K-Means++ consistently achieved 15-20% lower inertia and required fewer iterations to converge, confirming that our approach mitigates the risk of local optima.

Finally, the identification of three distinct transaction clusters acquires profound significance when contextualized within IOTA's ongoing transformation from a niche IoT platform to foundational infrastructure for global trade. As of 2026, IOTA is actively deployed in government trade systems across multiple continents, targeting a $35 trillion annual trade sector [50], [51]. These findings contribute to the growing body of research on DAG-based distributed ledgers by demonstrating that systematic parameter optimization for clustering algorithms can reveal meaningful transaction patterns that reflect the underlying architectural principles of the Tangle. Future research could extend this analysis by incorporating topological features—such as transaction depth, approval counts, or position within the DAG structure—to further refine our understanding of gas efficiency dynamics in asynchronous distributed systems.

4.5.1 Alignment with IOTA's Global Trade Initiatives

The IOTA Foundation has recently spearheaded the creation of the world's first decentralized global trade cluster, with **Trade Worldwide Information Network (TWIN)**, launched in partnership with the World Economic Forum, the Tony Blair Institute for Global Change, and the AfCFTA Secretariat. TWIN is designed as open-source digital infrastructure for secure, real-time trade data sharing across supply chains, with a stated goal of reducing global trade costs by 25% [52], [53].

Our clustering results directly support this global initiative by providing empirical evidence of transaction behavior patterns that can inform the design of efficient trade finance mechanisms. Specifically:

- **Cluster 0 (Efficient transactions)** represents the ideal transaction profile for the high-volume, low-value microtransactions that characterize machine-to-machine (M2M) communication in global supply chains. The IOTA Tangle's zero-fee architecture is specifically designed to enable such transactions across the estimated 20.4 billion IoT devices and achieving **50,000+ transactions per second** with sub-second finality (approximately 400ms) [54], the efficient cluster validates that the majority of transactions can operate at the scale required for national and continental deployment.
- **Cluster 1 (Complex transactions)** represents the transactions correspond to the type of multi-party, multi-document trade operations that TWIN handles. Each cross-border trade transaction can involve up to **30 separate parties**, with **36 original documents and 240 copies** exchanged globally [52]. Our finding that complex transactions form a distinct, stable cluster (Jaccard similarity = 0.82) provides empirical validation that TWIN's node-centric, modular architecture can effectively categorize and manage such operational complexity.
- **Cluster 2 (Gas-intensive transactions)** represents transactions that highlight the specific inefficiencies that IOTA's infrastructure aims to address. According to World Economic Forum data, approximately **4 billion trade documents circulate daily** across international trade networks, with cost overhead reaching **20% per cross-border transaction** [50], [53]. Banks lose **$2-5 billion annually** to forged or duplicated trade documents [50]. The gas-intensive cluster represents transactions that, if optimized, could contribute to closing the **$2.5 trillion annual trade finance gap** identified by the IOTA Foundation [54], [55].

4.5.2 Measurable Impact: Real-World Deployment Results

The practical relevance of our clustering framework is demonstrated by IOTA's live deployments:

- **East Africa (TLIP - Trade and Logistics Information Pipeline):** The TWIN pilot in East Africa, operating through the Trade and Logistics Information Pipeline, has achieved **70% reduction in**

customs clearance times (from weeks to days) and **30% reduction in trade costs** [52]. These improvements align with the characteristics of our efficient transaction cluster, suggesting that transactions exhibiting Cluster 0 patterns are those most effectively processed through the digitized infrastructure.

- **Kenya-UK Trade Corridor:** As of February 2026, IOTA's blockchain infrastructure is operational in both Kenyan and UK government trade systems, handling digital document verification for flower exports from Kenya and poultry imports from Poland. The UK Cabinet Office has seconded four government employees to IOTA to expand these trials, and the system is projected to expand to all Kenyan commodities by the end of Q1 2026 [50].
- **ADAPT Initiative (Africa Digital Access and Public Infrastructure for Trade):** Launched in November 2025, ADAPT is a pan-African digital public infrastructure initiative co-founded by IOTA with the AfCFTA Secretariat, aiming to connect **1.5 billion people** across all African nations by 2035 and double intra-African trade by 2035 [54], [55]. Three African countries are confirmed to deploy on IOTA Mainnet in 2026, with five more beginning pilot programs [55]. Our clustering methodology provides a benchmark framework for monitoring transaction efficiency across these national deployments.

4.5.3 Institutional Recognition and Market Potential

The global relevance of IOTA's infrastructure—and by extension, our analytical framework—is underscored by institutional endorsements:

- The **World Economic Forum** has officially recognized IOTA as a key technology pillar for digital trade and decentralized identity, collaborating on TWIN and highlighting IOTA's ability to directly connect governments, financial institutions, and SMEs while eliminating intermediaries [52], [53].
- The **Tony Blair Institute for Global Change** serves as a strategic partner in ADAPT and TWIN, providing governance expertise for large-scale digital infrastructure deployment [53], [54].
- **GLEIF (Global Legal Entity Identifier Foundation)** is collaborating with IOTA to enable businesses to establish verifiable digital identities through the vLEI (verifiable Legal Entity Identifier) layer, creating a "global verifiable identity layer" essential for trusted cross-border transactions [54].

Market projections reinforce the significance: IOTA's infrastructure targets a **$35 trillion annual trade sector** [50], with the potential to unlock **$10 trillion in new trade value by 2030** and reduce global trade costs by 25% [52]. Even a 5% efficiency gain through digitalization would unlock hundreds of billions of dollars in economic value [50].

The World Economic Forum (WEF) has officially recognized IOTA as a key technology pillar for shaping the future of global digital trade and decentralized identity. In its recent white paper "Advancing Digital Trade: Insights from the UAE TradeTech Regulatory Sandbox," the WEF highlighted IOTA's ability to directly connect governments, financial institutions, and SMEs while eliminating intermediaries [56]. Our clustering methodology contributes to this institutional validation by offering:

1. **Benchmarking Framework**: The three-cluster model provides a quantitative framework for assessing transaction efficiency that can be adopted by enterprises implementing IOTA-based solutions. Organizations participating in initiatives like the UAE TradeTech Sandbox—where IOTA collaborated with the Financial Services Regulatory Authority (FSRA), Kenya Revenue Authority, and TradeMark Africa —can use our approach to monitor and optimize their transaction patterns.
2. **Risk Assessment for Trade Finance**: The correlation analysis revealing a negative relationship between gas price and transaction success ($r = -0.59$) has direct implications for trade finance platforms such as Salus, which facilitates funding for mineral shipments . Understanding that higher resource allocation does not guarantee successful confirmation can inform more sophisticated risk modeling for trade finance instruments.
3. **Scalability Validation**: The Tangle architecture's promise of increased throughput with network growth—achieving over 50,000 transactions per second with sub-500ms latency —is predicated on efficient transaction processing. Our identification of the efficient transaction cluster provides empirical evidence that the majority of transactions in the IOTA ecosystem are indeed optimized for this scalable infrastructure.

4.5.4 Implications for Cross-Border Trade and Digital Identity

The Global Legal Entity Identifier Foundation (GLEIF) verifiable Legal Entity Identifier (vLEI) layer integrated into IOTA's ecosystem creates a "global verifiable identity layer" that is essential for trusted cross-border transactions. Our clustering results have direct relevance to this identity infrastructure:

- The "fuzzy boundary" concept in economic clustering aligns with our finding that three clusters optimally represent transaction categories. In global trade, where participants may have different levels of trust and

verification requirements, the graded categorization revealed by our analysis (rather than binary classification) more accurately reflects the nuanced reality of international commerce.

- The stability of our clusters across feature weight variations (Section 3.4) suggests that transaction efficiency patterns are robust enough to accommodate the diverse regulatory and operational requirements of different countries and continents—a critical requirement as the global trade cluster expands to accommodate entire nations and regions .

4.5.5 Comparison with Other Cryptocurrency Networks and Competing Infrastructure Platforms

Our findings also contribute to the broader academic discourse on cryptocurrency transaction analysis. Recent research on temporal motifs in Bitcoin and NFT transaction networks has demonstrated the importance of disaggregating global patterns to understand individual user behavior. Similarly, our local clustering approach reveals transaction patterns that would be obscured in aggregate analysis.

The contrast between our IOTA findings and studies on Ethereum gas markets highlights the architectural differences between DAG-based and blockchain-based systems. While Ethereum transactions face competitive gas bidding and congestion pricing, IOTA's feeless architecture and parallel validation mechanism create fundamentally different efficiency dynamics—dynamics that our clustering analysis has successfully captured.

Our findings also contextualize IOTA's competitive positioning against previous trade digitalization attempts. The failure of IBM and Maersk's TradeLens platform in 2023—despite hundreds of millions in investment—was attributed not to technology but to governance failure: competitors could not trust a platform led by an industry leader [53]. IOTA's approach differs fundamentally: TWIN is governed through a not-for-profit foundation with a neutral, open-source, node-centric architecture where no single entity controls the network .

Our clustering results support the viability of this governance model by demonstrating that transaction patterns in the IOTA ecosystem are **intrinsically stratified**—efficient, complex, and gas-intensive transactions form natural, stable groupings. This stratification suggests that the underlying infrastructure can accommodate diverse participant needs without requiring centralized coordination, a key requirement for trust-neutral infrastructure.

4.5.6 Practical Recommendations for Global Adoption

Based on our clustering analysis, the following recommendations for stakeholders in IOTA's expanding global ecosystem are offered:

1. **For Trade Platforms (TWIN/TLIP)**: Implement monitoring systems that classify transactions into our three-cluster framework to identify inefficiencies in real-time and prioritize optimization efforts for gas-intensive transactions. Our stability analysis confirms the 3-cluster solution is highly reproducible (Jaccard similarities 0.82-0.89).
2. **For Financial Institutions**: Use the correlation between gas parameters and transaction success ($r = -0.59$) to develop more accurate credit risk models for trade finance, particularly for transactions in the complex and gas-intensive clusters. The $2.5 trillion trade finance gap is exacerbated by lack of visibility into shipment authenticity [50].
3. **For IoT Device Manufacturers**: Optimize device firmware to ensure transactions fall within the efficient cluster (Cluster 0), minimizing resource consumption across the projected 20.4 billions of devices. IOTA's feeless architecture enables economically viable microtransactions only if gas consumption remains efficient.
4. **For Policymakers**: Recognize that DAG-based infrastructures like IOTA offer fundamentally different efficiency characteristics than traditional blockchains, warranting tailored regulatory approaches that account for feeless, parallel validation mechanisms for the machine economy. The UN's Model Law on Electronic Transferable Records (MLETR) has established legal foundations for digital trade documents [53].
5. **Customs Authorities**: Leverage the efficient transaction cluster (Cluster 0) as a benchmark for digitized clearance processes; the TLIP experience shows 70% reduction in clearance times is achievable [52].

In summary, the three transaction clusters identified in this study are not merely academic categorizations—they represent empirically validated profiles that align with IOTA's real-world deployment as critical global infrastructure. As IOTA continues to gain recognition from institutions like the World Economic Forum and expands its global trade cluster initiative, our findings provide an analytical foundation for optimizing transaction efficiency across this rapidly growing ecosystem.

5 CONCLUSION AND FUTURE WORK

This study demonstrates the effectiveness of K-Means clustering for evaluating smart contract transaction efficiency in blockchain networks. The combination of gas price, gas limit, and gas consumed—together with their derived features—reveals three distinct transaction patterns: efficient, complex, and gas-intensive. These findings confirm that the three parameters significantly influence both cost effectiveness and successful contract execution.

Visualization of the clustering results in 2D and 3D spaces supports the quantitative assessment, showing that the three-cluster model exhibits clearer and more structured separation than the two-cluster model. The majority of transactions fall into the efficient cluster, while the remaining two represent complex and inefficient patterns. Overall, the analysis indicates that certain transactions suffer from excessive gas allocation and cost inefficiencies. Understanding these patterns can guide future optimization of gas determination algorithms to improve smart contract execution efficiency.

Based on these findings, we recommend that gas limits and prices be calibrated more precisely to avoid cost waste and transaction execution errors. Smart contract developers should also optimize code structures to improve gas efficiency, particularly for complex transactions. Future work may explore alternative clustering algorithms, such as DBSCAN or hierarchical clustering, to benchmark performance against K-Means. Additionally, the results of this study could be integrated into a gas setting recommendation system, enabling users to determine optimal gas limit, gas consumed, and gas price values before submitting transactions to the blockchain network.

A limitation of this study is the absence of per-transaction latency data, preventing a direct empirical analysis of the relationship between gas consumption and confirmation time in the IOTA Tangle. Future research should incorporate fine-grained temporal features (e.g., time to first approval, transaction depth) and conduct controlled experiments on the IOTA Rebased mainnet [57] to determine whether gas-intensive transactions experience different confirmation latencies compared to efficient ones. Such investigations would complement our clustering framework and enhance the understanding of transaction efficiency in DAG-based ledgers.

ACKNOWLEDGMENTS

The authors would like to thank **SPCIES** Research Groups, **COMNETS** Research Groups, **Faculty of Computer Science**, and **Faculty of Engineering** of **Universitas Sriwijaya** for their technical assistance/insightful discussions and for providing access to laboratory and its facilities. Appreciation is extended to the all reviewers for their valuable feedback.

REFERENCES

[1] J. Sedlmeir, H. U. Buhl, G. Fridgen, and R. Keller, "The Energy Consumption of Blockchain Technology: Beyond Myth," vol. 62, no. 6, pp. 599–608, Dec. 2020, doi: 10.1007/s12599-020-00656-x.

[2] B. K. Mohanta, S. S. Panda, and D. Jena, "An Overview of Smart Contract and Use Cases in Blockchain Technology," in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 2018, pp. 1–4. doi: 10.1109/icccnt.2018.8494045.

[3] Y. Jiang, G. Sun, and T. Feng, "Research on Data Transaction Security Based on Blockchain," *Information*, vol. 13, no. 11, p. 532, Nov. 2022, doi: 10.3390/info13110532.

[4] M. D. Patil and M. V. Bhosale, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 293–298, 2023, doi: 10.48175/ijarsct-8158.

[5] E. Tan, S. Mahula, and J. Crompvoets, "Blockchain governance in the public sector: A conceptual framework for public management," *Gov. Inf. Q.*, vol. 39, no. 1, p. 101625, 2022, doi: 10.1016/j.giq.2021.101625.

[6] H. Ding *et al.*, "How Far Should We Go Away from Smart Contract to Smarter Contractor? A Systematic Review," *Blockchain Res. Appl.*, p. 100402, Oct. 2025, doi: 10.1016/j.bcra.2025.100402.

[7] J. Chen, M. Huang, Z. Lin, P. Zheng, and Z. Zheng, "To healthier Ethereum: a comprehensive and iterative smart contract weakness enumeration," *Blockchain Res. Appl.*, vol. 6, no. 2, p. 100258, 2025, doi: 10.1016/j.bcra.2024.100258.

[8] S.-Y. Lin, L. Zhang, J. Li, L. Ji, and Y. Sun, "A survey of application research based on blockchain smart contract," *Wirel. Networks*, vol. 28, no. 2, pp. 635–690, Feb. 2022, doi: 10.1007/s11276-021-02874-x.

[9] C. Laneve and C. Sacerdoti Coen, "Analysis of smart contracts balances," *Blockchain Res. Appl.*, vol. 2, no. 3, p. 100020, Sep. 2021, doi: 10.1016/j.bcra.2021.100020.

[10] A. Sadu, A. Jindal, G. Lipari, F. Ponci, and A. Monti, "Resilient design of distribution grid automation system against cyber-physical attacks using blockchain and smart contract," *Blockchain Res. Appl.*, vol. 2, no. 1, p. 100010, 2021, doi: 10.1016/j.bcra.2021.100010.

[11] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/access.2020.2988796.

[12] A. Kiran, P. Mathivanan, M. Mahdal, K. Sairam, D. Chauhan, and V. Talasila, "Enhancing Data Security in IoT Networks with Blockchain-Based Management and Adaptive Clustering Techniques," *Mathematics*, vol. 11, no. 9, p. 2073, Apr. 2023, doi: 10.3390/math11092073.

[13] N. Ahmed, M. A. Ngadi, A. A. Almazroi, and N. A. Alghanmi, "Hybrid Model for Novel Attack Detection Using a Cluster-Based Machine Learning Classification Approach for the Internet of Things (IoT)," *Futur. Internet*, vol. 17, no. 6, p. 251, 2025, doi: 10.3390/fi17060251.

[14] M. Alshaikhli, S. Al-Maadeed, and M. Saleh, "Enhancing fairness and scalability in IOTA tangle networks: a POMDP-based tip selection algorithm for decentralized systems," *Cluster Comput.*, vol. 28, no. 14, p. 875, Nov. 2025, doi: 10.1007/s10586-025-05432-8.

[15] IOTA Foundation, "Data Structures." Accessed: Mar. 18, 2026. [Online]. Available: https://legacy.wiki.iota.org/learn/protocols/iota2.0/core-concepts/data-structures/

[16] F. Guo, "Modeling Improvement and Application of DAG-based Blockchain Networks," 2025, doi: https://doi.org/10.34726/hss.2025.133164.

[17] L. Tassanyi, "Economic Clustering in the Tangle." Accessed: Mar. 20, 2026. [Online]. Available: https://raw.githubusercontent.com/iotaledger/ec.ixi/master/docs/EC.md

[18] B. K. Meister and H. C. W. Price, "Gas fees on the Ethereum blockchain: from foundations to derivative valuations," 2024, *Ithaca PP - Ithaca*. doi: 10.3389/fbloc.2024.1462666.

[19] I.-C. Lin, C.-C. Chang, and Y.-S. Chang, "Data Security and Preservation Mechanisms for Industrial Control Network Using IOTA," *Symmetry (Basel).*, vol. 14, no. 2, p. 237, 2022, doi: 10.3390/sym14020237.

[20] I.-C. Lin, P.-C. Tseng, Y.-S. Chang, and T.-C. Weng, "IOTA Data Preservation Implementation for Industrial Automation and Control Systems," *Processes*, vol. 11, no. 7, p. 2160, 2023, doi: 10.3390/pr11072160.

[21] F. Guo, X. Xiao, A. Hecker, and S. Dustdar, "Characterizing IOTA Tangle with Empirical Data," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/GLOBECOM42002.2020.9322220.

[22] E. Ferro *et al.*, "Digital assets rights management through smart legal contracts and smart contracts," *Blockchain Res. Appl.*, vol. 4, no. 3, p. 100142, Sep. 2023, doi: 10.1016/j.bcra.2023.100142.

[23] M. Alishahi, W. Leeuw, and N. Zannone, "BC-FL k-means: A Blockchain-based Framework for Federated Clustering," in *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, IEEE, Nov. 2023, pp. 1348–1354. doi: 10.1109/trustcom60117.2023.00184.

[24] Y. Pei *et al.*, "Meta reinforcement learning based dynamic tuning for blockchain systems in diverse network environments," *Blockchain Res. Appl.*, vol. 6, no. 2, p. 100261, 2025, doi: 10.1016/j.bcra.2024.100261.

[25] C. Cuagdan *et al.*, "Experiences of Blockchain Technology Users," 2023. doi: 10.13140/RG.2.2.36156.41604.

[26] I. Pittaras, N. Fotiou, C. Karapapas, V. A. Siris, and G. C. Polyzos, "Secure smart contract-based digital twins for the Internet of Things," *Blockchain Res. Appl.*, vol. 5, no. 1, p. 100168, 2024, doi: 10.1016/j.bcra.2023.100168.

[27] S. S. Kushwaha, S. Joshi, D. Singh, M. Kaur, and H.-N. Lee, "Ethereum Smart Contract Analysis Tools: A Systematic Review," *IEEE Access*, vol. 10, pp. 57037–57062, 2022, doi: 10.1109/access.2022.3169902.

[28] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny).*, vol. 622, pp. 178–210, 2023, doi: 10.1016/j.ins.2022.11.139.

[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*, New York, NY, USA: ACM, Aug. 2024, pp. 407–412. doi: 10.1145/3675888.3676080.

[30] R. Agarwal, S. Barve, and S. K. Shukla, "Detecting malicious accounts in permissionless blockchains using temporal graph properties," *Appl. Netw. Sci.*, vol. 6, no. 1, p. 9, Dec. 2021, doi: 10.1007/s41109-020-00338-3.

[31] K. Snigdha, P. S. N. Reddy, D. Hema, and S. Gayathri, "BitPredict: End-to-End Context-Aware Detection of Anomalies in Bitcoin Transactions using Stack Model Network," in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, IEEE, May 2024, pp. 1–6. doi: 10.1109/ACCAI61061.2024.10602116.

[32] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.

[33] D. Stiawan *et al.*, "Ping Flood Attack Pattern Recognition Using a K-Means Algorithm in an Internet of Things (IoT) Network," *IEEE Access*, vol. 9, pp. 116475–116484, 2021, doi: 10.1109/ACCESS.2021.3105517.

[34] M. Mayo, "Revisiting k-Means: 3 Approaches to Make It Work Better," https://machinelearningmastery.com/. Accessed: Mar. 20, 2026. [Online]. Available: https://machinelearningmastery.com/revisiting-k-means-3-approaches-to-make-it-work-better/

[35] F. Ahmatshin and L. Kazakovtsev, "Mini-batch K-Means++ Clustering Initialization," 2024, pp. 293–307. doi: 10.1007/978-3-031-73365-9_20.

[36] A. Rawat, V. Daza, and M. Signorini, "Offline Scaling of IoT Devices in IOTA Blockchain," *Sensors*, vol. 22, no. 4, Feb. 2022, doi: 10.3390/s22041411.

[37] D. A. Setiady and H. Leong, "IMPLEMENTATION OF K-MEANS ALGORITHM ELBOW METHOD AND SILHOUETTE COEFFICIENT FOR RAINFALL CLASSIFICATION," *Proxies J. Inform.*, vol. 4, no. 1, pp. 18–25, 2024, doi: 10.24167/proxies.v4i1.12433.

[38] D. Chicco, A. Campagner, A. Spagnolo, D. Ciucci, and G. Jurman, "The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters," *PeerJ Comput. Sci.*, vol. 11, p. e3309, Nov. 2025, doi: 10.7717/peerj-cs.3309.

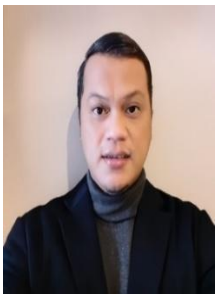
[39] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, "Analysis of an optimal cluster approach: a review paper,"

J. Phys. Conf. Ser., vol. 2421, no. 1, p. 12015, 2023, doi: 10.1088/1742-6596/2421/1/012015.
[40] F. Ros, R. Riad, and S. Guillaume, "PDBI: A partitioning Davies-Bouldin index for clustering evaluation," *Neurocomputing*, vol. 528, pp. 178–199, Apr. 2023, doi: 10.1016/j.neucom.2023.01.043.
[41] S. P. Lima and M. D. Cruz, "A genetic algorithm using Calinski-Harabasz index for automatic clustering problem," *Rev. Bras. Comput. Apl.*, vol. 12, no. 3, pp. 97–106, Sep. 2020, doi: 10.5335/rbca.v12i3.11117.
[42] S. Chakraborty, D. Paul, and S. Das, "On Consistent Entropy-Regularized k-Means Clustering With Feature Weight Learning: Algorithm and Statistical Analyses," *IEEE Trans. Cybern.*, vol. 53, no. 8, pp. 4779–4790, Aug. 2023, doi: 10.1109/TCYB.2022.3166975.
[43] J. Z. Huang, M. K. Ng, Hongqiang Rong, and Zichen Li, "Automated variable weighting in k-means type clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 657–668, May 2005, doi: 10.1109/TPAMI.2005.95.
[44] R. C. de Amorim, "A Survey on Feature Weighting Based K-Means Algorithms," *J. Classif.*, vol. 33, no. 2, pp. 210–242, Jul. 2016, doi: 10.1007/s00357-016-9208-4.
[45] C. Hennig, "Cluster-wise assessment of cluster stability," *Comput. Stat. Data Anal.*, vol. 52, no. 1, pp. 258–271, Sep. 2007, doi: 10.1016/j.csda.2006.11.025.
[46] C. Hennig, "Dissolution point and isolation robustness: Robustness criteria for general cluster analysis methods," *J. Multivar. Anal.*, vol. 99, no. 6, pp. 1154–1176, Jul. 2008, doi: 10.1016/j.jmva.2007.07.002.
[47] A. R. Gad, M. Haggag, A. A. Nashat, and T. M. Barakat, "A Distributed Intrusion Detection System using Machine Learning for IoT based on ToN-IoT Dataset," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, pp. 548–563, 2022, doi: 10.14569/IJACSA.2022.0130667.
[48] Y. Feng, J. Zou, W. Liu, and F. Lv, "Distributed K-Means algorithm based on a Spark optimization sample," *PLoS One*, vol. 19, no. 12, p. e0308993, Dec. 2024, doi: 10.1371/journal.pone.0308993.
[49] C. Kanaracus and A. Mulholland, "IOTA's Tangle: The Distributed Ledger Built for IoT." Accessed: Mar. 20, 2026. [Online]. Available: https://www.constellationr.com/insights/news/iotas-tangle-distributed-ledger-built-iot
[50] P. Zhang, "IOTA Bets Big on $35T Trade Sector With 'World Onchain' Manifesto." Accessed: Mar. 20, 2026. [Online]. Available: https://blockchain.news/postamp?id=iota-manifesto-world-onchain-trade-digitization
[51] E. Crypto, "IOTA Deploys Blockchain in UK Kenya Trade Systems." Accessed: Mar. 19, 2026. [Online]. Available: https://www.edgen.tech/zh/news/crypto/iota-deploys-blockchain-in-uk-kenya-trade-systems
[52] A. Hoffner and S. Liu, "IOTA's Strategic Shift to Global Trade Infrastructure: Blockchain as the New Backbone of Global Commerce." Accessed: Mar. 20, 2026. [Online]. Available: https://www.ainvest.com/news/iota-strategic-shift-global-trade-infrastructure-blockchain-backbone-global-commerce-2601/
[53] Jun, "From Platform to Infrastructure: IOTA's Trade Revolution." Accessed: Mar. 20, 2026. [Online]. Available: https://4pillars.io/en/articles/from-platform-to-infrastructure-iotas-trade-revolution
[54] I. Foundation, "A Year of Rebirth, Growth, and Global Impact." [Online]. Available: March 19, 2026
[55] CryptoNewsFlash, "IOTA Enters 2026 Focused on Countries, Trade, and Enterprise-Grade Infrastructure." Accessed: Mar. 19, 2026. [Online]. Available: https://www.gate.com/tr/news/detail/17394095
[56] W. E. Forum, "Advancing Digital Trade : Insights from the UAE TradeTech Regulatory Sandbox," 2025, [Online]. Available: https://reports.weforum.org/docs/WEF_Advancing_Digital_Trade_2025.pdf
[57] P. Sedi Nzakuna, V. Paciello, A. Lay-Ekuakille, A. Kuti Lusala, S. Dello Iacono, and A. Pietrosanto, "From IOTA Tangle 2.0 to Rebased: A Comparative Analysis of Decentralization, Scalability, and Suitability for IoT Applications," *Sensors*, vol. 25, no. 11, p. 3408, May 2025, doi: 10.3390/s25113408.

BIOGRAPHY OF AUTHORS



Huda Ubaya completed his bachelor's and master's degrees at the Bandung Institute of Technology (ITB). He is a lecturer in Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya. His field of research is related to distributed ledger technology, blockchain, Internet of Things, and intelligent embedded systems. He can be contacted at email: huda@unsri.ac.id



Deris Stiawan received the Ph.D. degree in computer science from the Universiti Teknologi Malaysia, Malaysia. He is currently a Professor in Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya. His research interests include computer networks, intrusion detection/prevention systems, and heterogeneous networks. He can be contacted at email: deris@unsri.ac.id



Bhakti Yudho Suprapto is currently pursuing the Graduate degree in electrical engineering with Sriwijaya University, Indonesia. His master's and doctoral programs in Electrical Engineering at Universitas Indonesia (UI). He is also an Academic Staff Member of Electrical Engineering at Universitas Sriwijaya. His research interests include control and intelligent systems. He can be contacted at email: bhakti@ft.unsri.ac.id



Rivaldi Febrian Ekaputra recently received the Bachelor Degree in Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya at 2025. He can be contacted at email: rivaldi13febrian@gmail.com



Nurul Afifah received the Bachelor Degree in Computer Engineering and master's degrees in Computer Science at 2014 and 2019. She joined Universitas Sriwijaya as a Lecturer in January 2020. She is currently a Researcher at COMNETS RG. Her research interests include Information Security, IoT system & security, Blockchain and Machine learning. She can be contacted at email: nurul@unsri.ac.id



Septiani Kusuma Ningrum received the Bachelor of Computer Science degree from the Faculty of Computer Science, Universitas Sriwijaya, Indonesia, in 2024, where she is currently pursuing the Master of Computer Science degree. She is also a researcher at the Computer Networks, Entreprise and Information Security Research Group (COMNETS RG), Universitas Sriwijaya. Her current research interests include information security, Ethereum blockchain data analysis, and Advanced Persistent Threats (APT). She can be contacted at email: septianikuning@gmail.com