

Analisis K-Means Clustering pada Wireless Sensor Network Berdasarkan Residual Energy dan Packet Delivery Ratio

Indri Theresia Manik^{1*}, Nadieta Eunike Sitorus², Sasta Monika Marbun³, Sri yuni Utari Sihombing⁴.

^{1,2,3,4}Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, Indonesia

Email Corresponding : indrimanik52@gmail.com.

Article Info

Article history:

Received: 02 February 2026

Revised: 01 May 2026

Accepted: 25 May 2026

Keywords:

*K-Means Clustering,
Wireless Sensor Network,
Residual Energy,
Packet Delivery Ratio,
Silhouette Analysis.*

ABSTRACT

Wireless Sensor Networks (WSNs) rely on nodes with constrained energy resources and efficient data transmission. This study uses the K-Means clustering algorithm to classify nodes based on residual energy and Packet Delivery Ratio (PDR) using a publicly available dataset containing 1,000 observations. PDR was derived from the number of packets sent and received, with values exceeding 100% limited to 100%. Both variables were then standardized using z-score normalization. The optimal number of clusters was evaluated using the Elbow method and silhouette analysis for values of k ranging from 1 to 10. The analysis identified k = 2 as the optimal choice, yielding the highest average silhouette score of 0.47015. The resulting clusters consisted of Cluster 1, comprising 262 observations (26.2%), and Cluster 2, containing 738 observations (73.8%). The mean PDR in Cluster 1 was lower at 43.21%, compared with 96.59% in Cluster 2. In contrast, the average residual energy was relatively comparable between the two clusters, measuring 1.464 and 1.447, respectively. These findings suggest that PDR was the primary variable distinguishing the two groups. The average silhouette scores for Cluster 1 and Cluster 2 were 0.399 and 0.495, respectively, indicating a moderately distinct clustering structure, although the separation was not entirely strong. The findings are exploratory and do not imply causal relationships. Furthermore, the results should be interpreted cautiously because 578 observations with PDR values exceeding 100% were capped at 100%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. INTRODUCTION

Wireless Sensor Networks (WSNs) are widely used in environmental monitoring, industrial automation, smart agriculture, and other applications that require distributed data collection and wireless communication. A WSN consists of sensor nodes that sense environmental conditions, process information, and transmit data to a sink or base station. Despite their flexibility, WSNs face challenges related to limited energy resources, communication reliability, and network lifetime. Because sensor nodes often operate with constrained power supplies, energy consumption and data transmission performance are important considerations in WSN design and evaluation [1], [2].

Residual energy and Packet Delivery Ratio (PDR) provide complementary information about WSN conditions. Residual energy represents the energy remaining at a node after network operations, whereas PDR is the proportion of transmitted packets that are successfully received at the destination. Residual energy is relevant to understanding the remaining capacity of nodes to operate, while PDR reflects the success of data delivery. Research on WSN clustering has shown that energy-related characteristics and communication performance are important factors in organizing nodes and managing network operation [3], [6]. Considering residual energy and PDR together can therefore support an exploratory assessment of differences among observations. However, these variables describe different dimensions of network performance, and their association should not automatically be interpreted as a causal relationship.

Clustering is commonly used in WSNs to organize nodes into groups and support network management. Low-Energy Adaptive Clustering Hierarchy (LEACH) introduced a hierarchical clustering approach in which nodes form clusters and cluster heads coordinate data transmission [7]. Subsequent research has investigated modifications to clustering and cluster-head selection to reduce energy consumption, balance node workloads, and extend network lifetime [4], [10]. K-Means is also used in WSN studies because it groups observations according to similarity in selected features. Existing work has applied K-Means and its variants to energy-efficient clustering, routing, localization, and cluster-head selection [3], [6], [8], [11]. These studies provide a basis for using clustering to explore differences in node characteristics.

An important decision in K-Means clustering is the selection of the number of clusters. The Elbow method examines the change in within-cluster sum of squares (WSS) as the number of clusters increases and helps identify a point at which additional clusters provide diminishing reductions in within-cluster variation [12]. Silhouette analysis provides a complementary measure by comparing the similarity of an observation to its own cluster with its similarity to neighboring clusters [13]. A silhouette value closer to 1 indicates that an observation is well matched to its assigned cluster, a value around 0 suggests that it lies near a cluster boundary, and a negative value may indicate a potentially unsuitable assignment. Using both approaches can provide more informed support for choosing and evaluating a clustering configuration than selecting the number of clusters arbitrarily.

This study applies K-Means clustering to a publicly available Wireless Sensor Network dataset containing 1,000 observations and 18 attributes. The analysis focuses on residual energy and PDR, calculated from the available energy and packet-count attributes. The procedure includes data inspection, PDR calculation, preprocessing, standardization, determination of the number of clusters using the Elbow method and silhouette analysis, and interpretation of the resulting clusters. Standardization is used to reduce the influence of differences in variable scales on distance calculations. Cluster sizes, average values, and visual distributions are then examined to describe the characteristics of the groups. The dataset source and analysis procedure are documented in the research method section [14].

Data quality is an important consideration in interpreting the results. In the dataset used in this study, some observations have received-packet counts that exceed sent-packet counts, producing calculated PDR values above 100%. The current analysis caps values above 100% before standardization and clustering. This adjustment keeps PDR within its intended percentage range, but it can also compress differences among observations with raw calculated values above the threshold. Consequently, the resulting clusters should be interpreted in light of this preprocessing decision, and the original packet-count records warrant further examination. The study does not treat the clustering output as proof of causal relationships or as direct evidence of improved network lifetime.

The contribution of this study is an empirical characterization of WSN observations using the combined variables residual energy and PDR, with the number of clusters assessed using the Elbow method and silhouette analysis. Unlike studies that propose new routing protocols or optimize cluster-head selection, this study focuses on describing the groups formed in the available dataset and determining whether their main differences are associated with residual energy, PDR, or both. This is an exploratory contribution rather than a claim that these variables have not previously been used in WSN research. Accordingly, the objective is to identify a suitable cluster configuration and describe the characteristics of the resulting groups based on residual energy and PDR.

2. METHOD

This study uses a quantitative, unsupervised approach to group Wireless Sensor Network (WSN) nodes by energy efficiency and data delivery. K-Means clustering is widely used for this purpose [2], [3]. The research has seven stages, shown in Figure 1.

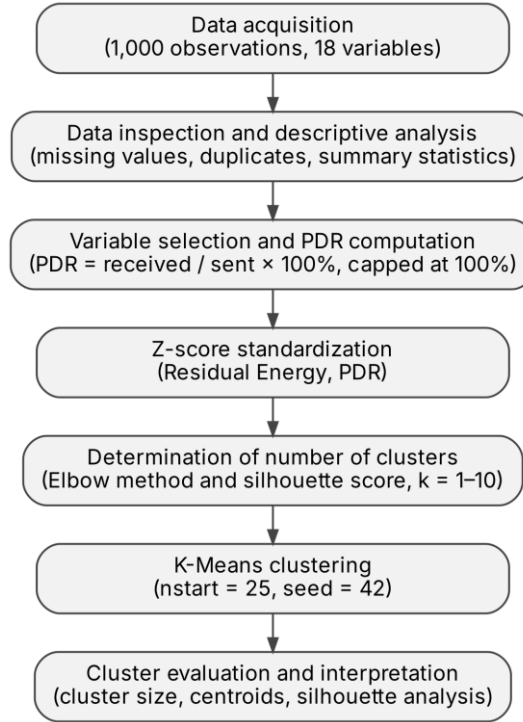


Figure 1. Research flowchart

2.1. Research Design and Data Acquisition

The data come from the public dataset "WSN: Dataset for Secure and Energy-Efficient" on Kaggle [7]. It contains 1,000 node records with 18 attributes, covering energy, authentication, node role, and packet counts. Three attributes were used: residual_energy, data_packets_sent, and data_packets_received. The residual energy indicates energy efficiency, and the packet counts give the delivery performance. The dataset's own cluster labels were not used. The data were checked for missing values and duplicate rows.

2.2. Data Preprocessing and Standardization

The Packet Delivery Ratio (PDR) of node i is computed as in (1). A PDR above 100% is invalid because received packets cannot exceed sent packets. Such values were capped at 100% as in (2) before standardization.

$$\text{PDR}_i = \frac{\text{Received}_i}{\text{Sent}_i} \times 100\% \quad (1)$$

$$\text{PDR}_i = \min(\text{PDR}_i, 100) \quad (2)$$

Residual Energy and PDR were then standardized with the z-score in (3), where x is the value, μ the mean, and σ the standard deviation. This keeps either variable from dominating the Euclidean distance.

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

2.3. Determination of the Number of Clusters

The number of clusters k was tested from 1 to 10 using the Elbow method and silhouette analysis, each with 20 random starts ($\text{nstart} = 20$). The Elbow method uses the total within-cluster sum of squares (WSS) in (4), where C_j is cluster j and μ_j is its centroid. The suitable k is where the decrease in WSS levels off [1].

$$\text{WSS}(k) = \sum_{j=1}^k \sum_{x \in C_j} \|x - \mu_j\|^2 \quad (4)$$

Because the elbow can be ambiguous, the average silhouette score was also used [6], [8]. The silhouette of observation i is given in (5), where $a(i)$ is the mean distance to its own cluster and $b(i)$ is the lowest mean distance to another cluster. It ranges from -1 to 1 , and higher values mean a better fit [4]. The k with the highest average silhouette was selected.

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (5)$$

2.4. K-Means Clustering

K-Means assigns each observation to the nearest of k centroids and updates the centroids to minimize the WSS in (4), as outlined in Algorithm 1. The default Hartigan–Wong algorithm in R was used [9]. The final model used $k = 2$, 25 random starts ($nstart = 25$), and a random seed of 42 for reproducibility. All analysis was run in R version 4.5 [10], using the base functions `read.csv`, `scale`, `dist`, and `kmeans`, and the cluster package [11] for silhouette values. Plots were made with base R graphics.

Algorithm 1. K-Means clustering of WSN nodes

Input: standardized data X ($n = 1,000$; Residual Energy, PDR), k , $nstart$, seed = 42

Output: cluster label of each observation, centroids

1. Set the random seed to 42
2. for $r = 1$ to $nstart$ do
 3. Select k observations randomly as initial centroids
 4. repeat
 5. Assign each observation to the nearest centroid (Euclidean distance)
 6. Recompute each centroid as the mean of its assigned observations
 7. until assignments no longer change
 8. Compute the WSS of the solution
9. end for
10. Return the solution with the lowest WSS

2.5. Cluster Evaluation and Interpretation

The results were evaluated in four steps: (i) cluster sizes and percentages; (ii) mean Residual Energy and PDR per cluster on the original scale, and centroids on the standardized scale; (iii) average silhouette per cluster and the distribution of silhouette values [4], [5]; and (iv) a scatter plot of Residual Energy against PDR. The clusters are read as groups with similar characteristics, not as evidence of a causal relationship.

3. RESULTS AND DISCUSSION

In this section, the results of the Wireless Sensor Network (WSN) data analysis are presented based on the steps of descriptive analysis, data preprocessing and normalization, determination of the number of clusters, and the application of the K-Means algorithm. The obtained results are then further analyzed to understand the network clustering characteristics based on the variables of Residual Energy and Packet Delivery Ratio (PDR).

3.1. Descriptive Analysis

Descriptive analysis was conducted to understand the characteristics of the Residual Energy and PDR variables within the WSN dataset. The dataset analyzed comprises 1,000 observations and 18 variables. Preliminary inspection revealed no missing values or duplicate rows. Descriptive statistics for both key variables are presented in Table 1.

Table 1. Descriptive Statistics for Residual Energy and PDR

Variable	Minimum	First Quartile (Q1)	Median	Mean	Third quartile (Q3)	Maximum	standard deviation
Residual Energy	0,570	1,170	1,450	1,451	1,722	2,330	0,379
PDR	11,11	67,01	100,00	82,60	100,00	100,00	25,683

Based on Table 1, the Residual Energy variable has a mean of 1.451 and a standard deviation of 0.379. Values range from 0.570 to 2.330, indicating variation in residual energy across WSN nodes. Meanwhile, the PDR has a mean of 82.60% and a standard deviation of 25.683. Both the median and the third quartile of the PDR reach 100%, indicating that many observations are at the maximum limit following the capping of PDR values.

To visualize the distribution patterns of the two variables, histograms of Residual Energy and PDR were applied.

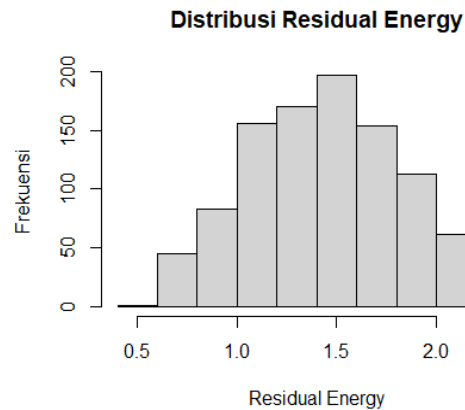


Figure 2. Residual Energy Distribution

The Residual Energy Histogram is used to display the distribution of residual energy values across observations. This diagram is useful for showing variations in value and data density which are then used as one of the foundations for grouping WSN nodes.

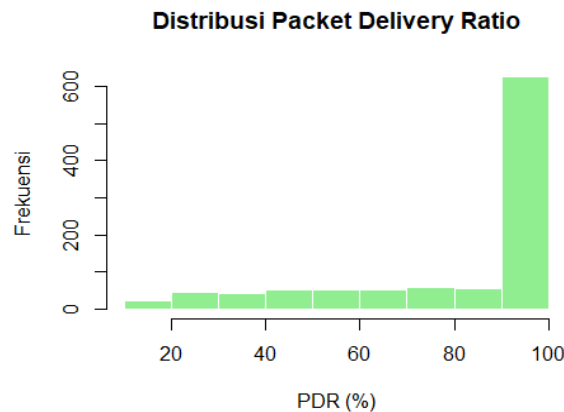


Figure 3. Packet Delivery Ratio Distribution

The PDR histogram illustrates the distribution of packet delivery ratios observed in the analysis. The concentration of values at 100% warrants attention because, during data processing, any PDR value exceeding 100% is capped at 100%. Consequently, the distribution shown in Figure 3 represents the PDR after this capping has been applied, rather than the raw PDR distribution.

3.2. Data Preprocessing and Standardization

Data preparation was conducted prior to the clustering process to ensure the data met the analysis criteria. The variables used for clustering were residual energy and PDR. PDR was calculated by determining the ratio of received packets to sent packets and multiplying the result by 100%.

The analysis revealed 578 observations where the number of received packets exceeded the number of sent packets, resulting in raw PDR values greater than 100%. In this analysis, PDR values exceeding 100% were capped at 100% using the `pmin()` function. Following this adjustment, the PDR showed a mean of 82.60364% and a standard deviation of 25.68323%.

Table 2. Summary of Variable Standardization

Variable	Pre-average	prior standard deviation	Post-test standard deviation	Post-test mean
Residual Energy	1,45139	0,37881	1,0000	≈ 0
PDR (%)	82,60364	25,68323	1,0000	≈ 0

Table 2 indicates that both variables have a mean of approximately zero and a standard deviation of one following standardization. Consequently, the scale difference between residual energy and PDR can be minimized prior to distance measurement. The standardized data are subsequently used to determine the number of clusters and in the K-Means clustering process.

3.3. Determination of the Number of Clusters

The number of clusters was determined using the Elbow method and silhouette analysis. The Elbow method was applied by comparing the Total Within-Cluster Sum of Squares (WSS) values for cluster counts ranging from $k = 1$ to $k = 10$. The number of clusters was decided based on the decrease in WSS as the number of clusters increased.

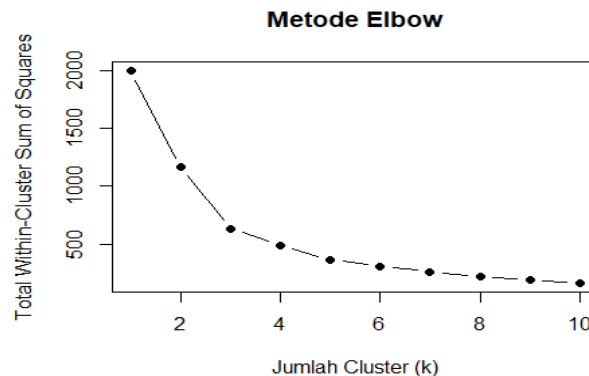


Figure 4. Determining the Number of Clusters Using the Elbow Method

Based on the Elbow plot, the Total Within-Cluster Sum of Squares (WSS) decreases as the number of clusters increases. This decline indicates that as more clusters are formed, the data variation within each cluster diminishes. However, increasing the number of clusters does not always yield a proportional improvement in the quality of the clustering. Therefore, the Elbow method is used to identify the point where the decrease in WSS begins to stabilize, signaling that adding further clusters results in only marginal reductions in variation.

In addition to the Elbow method, the silhouette score is calculated to evaluate the quality of clustering across various numbers of clusters. Silhouette scores range from -1 to 1; a value closer to 1 indicates that an observation is closer to its assigned cluster than to other clusters.

Table 3. Average silhouette value based on the number of clusters

Number of clusters (k)	Silhouette score
2	0,47015
3	0,46370
4	0,45202
5	0,42235
6	0,43198
7	0,41664
8	0,42516
9	0,43124
10	0,43763

Based on Table 3, the two-cluster configuration yielded the highest average silhouette score of 0.47015. This value exceeds those of the other cluster count options, including the three-cluster configuration, which produced a score of 0.46370. Consequently, a two-cluster configuration was selected for the K-Means analysis, as it demonstrated the best silhouette evaluation performance among the various cluster counts tested.

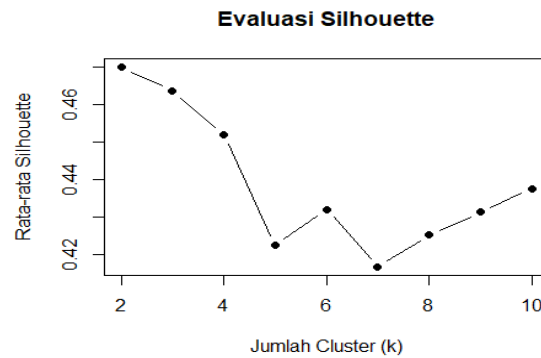


Figure 5. Silhouette score comparison graph

Based on the Silhouette Evaluation graph, the highest average silhouette score was achieved with two clusters ($k = 2$), at approximately 0.470. Subsequently, the score declined as the number of clusters increased, although there was a slight rise at $k = 6$ and a renewed increase from $k = 8$ to $k = 10$. This indicates that increasing the number of clusters does not necessarily improve clustering quality.

Therefore, a two-cluster configuration was selected, as it yielded the highest average silhouette score compared to the other tested cluster counts. These findings support the use of two clusters in the K-Means analysis of Wireless Sensor Networks, based on the variables of residual energy and Packet Delivery Ratio (PDR).

3.4. K-Means Clustering Results

The clustering process was performed using the K-Means algorithm to form two groups, with the nstart value set to 25 and a random seed of 42. The results indicate that the 1,000 observations were divided into two groups of unequal sizes.

Table 4. Number of observations in each cluster

Cluster	Number of observations	Percentage
Cluster 1	262	26,2%
Cluster 2	738	73,8%
Total	1.000	100%

Table 4 indicates that Cluster 1 comprises 262 observations, or 26.2% of the total data, whereas Cluster 2 comprises 738 observations, or 73.8%. These findings indicate that the majority of observations fall within Cluster 2.

To determine the characteristics of each group, the mean values of residual energy and PDR were compared on their original scales.

Table 5. Average of variables by cluster

Cluster	Number of observations	Average Residual Energy	Average PDR (%)
Cluster 1	262	1,463855	43,20689
Cluster 2	738	1,446965	96,59002

Based on the data in Table 5, Cluster 1 shows an average residual energy of 1.463855, which is slightly higher than that of Cluster 2 (1.446965). However, a more significant difference is observed in the PDR values. The average PDR for Cluster 1 is 43.20689%, whereas Cluster 2 records 96.59002%.

These findings indicate that Cluster 1 is characterized by a lower PDR, while Cluster 2 exhibits a higher PDR. The difference in average residual energy between the two clusters is relatively small compared to the disparity in PDR values. Consequently, the PDR variable stands out as the most distinctive feature differentiating the clustering results in this study.

Table 6. Cluster centers on standardized data

Cluster	Residual energy	PDR
Cluster 1	0,03291	-1,53395

Cluster	Residual energy	PDR
Cluster 2	-0,01168	0,54457

Table 6 shows that the centroid of Cluster 1 for the PDR variable has a value of -1.53395 , whereas the centroid of Cluster 2 has a value of 0.54457 . This indicates that the PDR characteristics in Cluster 1 fall below the mean of the standardized data, while those in Cluster 2 are above the mean. On the other hand, the residual energy centroid values for both clusters are quite close to each other.

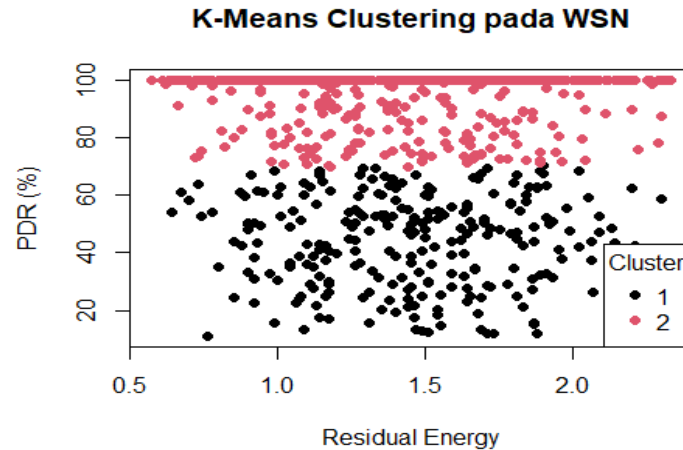


Figure 6. Visualization of K-Means clustering results based on Residual Energy and PDR.

Based on the K-Means visual analysis, two clusters emerged from the Residual Energy and Packet Delivery Ratio (PDR) variables. Cluster 1, represented by black dots, tends to exhibit lower PDR values, whereas Cluster 2, marked by pink dots, tends to show higher PDR values, with a significant portion of the data distributed within the 90–100% range.

When analyzed in terms of the Residual Energy variable, the value ranges of the two clusters overlap. This indicates that the most significant distinction between the clusters is evident in their PDR values rather than their Residual Energy levels. This finding aligns with the average PDR figures—43.21% for Cluster 1 and 96.59% for Cluster 2—while the average Residual Energy for both clusters remains quite similar.

In other words, the K-Means clustering technique successfully identified differences in WSN observations primarily based on packet delivery success rates. Cluster 1 represents a group with a relatively low PDR, while Cluster 2 represents a group with a relatively high PDR. However, this visual analysis does not imply that Residual Energy is the primary cause of the PDR differences; the graph merely illustrates the clustering patterns derived from these two variables.

3.5. Evaluation and Discussion of Clustering Result

The quality of the clustering was evaluated using silhouette analysis. The results indicate an overall average silhouette value of 0.47015. The average silhouette value for Cluster 1 was 0.39910, while for Cluster 2 it was 0.49538. Thus, based on these average values, Cluster 2 demonstrates better internal separation quality than Cluster 1.

The positive overall silhouette value suggests that the clustering results reveal a reasonably clear cluster structure, although the separation between clusters is not entirely robust. The minimum silhouette value of -0.01668 also indicates the presence of some observations that may lie close to the boundaries between groups or might be better suited to a different cluster. Therefore, the clustering results should be interpreted as groupings based on shared characteristics rather than as a perfect separation.

Table 7. Silhouette evaluation summary

Minimum	First quartile (Q1)	Median	Third quartile (Q3)	Maximum
-0,01668	0,38826	0,51708	0,59005	0,63165

The difference in average silhouette values between the two clusters indicates that observations in Cluster 2 fit their group better than those in Cluster 1. This aligns with the average PDR characteristics, where

Cluster 2 recorded a PDR of 96.59002%, whereas Cluster 1 reached only 43.20689%. In other words, the difference in PDR values serves as the primary distinguishing feature between the two groups in this clustering analysis.

In the context of Wireless Sensor Network applications, these findings demonstrate that clustering based on residual energy and PDR can help identify groups of observations with distinct packet delivery characteristics. Groups with low average PDR are a focal point for data delivery performance assessment, while groups with high average PDR exhibit superior packet reception rates. However, given the minimal difference in average residual energy between the two clusters, these results do not indicate a significant disparity in energy conditions between them.

It should be noted that there were 578 observations where the number of received packets exceeded the number of sent packets. In this study, PDR values were capped at a maximum of 100% prior to standardization and clustering; therefore, the interpretation of PDR results must take this adjustment into account. Further examination of the data regarding sent and received packet counts is required to ensure that the PDR values accurately reflect network conditions.

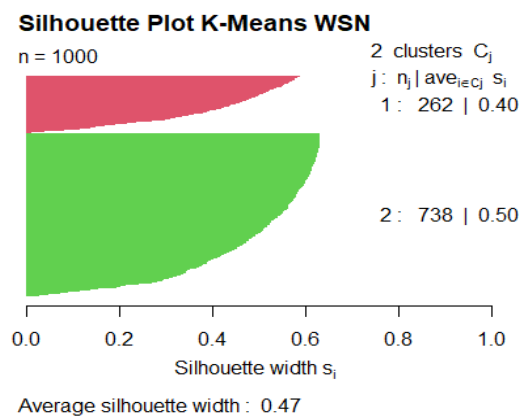


Figure 7. Final silhouette plot

In general, the K-Means approach using two clusters yields a grouping primarily differentiated by PDR levels. An average silhouette score of 0.47015 supports the decision to select two clusters over the other tested cluster counts. However, the evaluation also reveals limitations in the quality of inter-cluster separation; consequently, the clustering results should be viewed as a representation of data characteristics rather than as evidence that all nodes within a given cluster share identical network conditions.

4. CONCLUSION

This study applied K-Means clustering to 1,000 Wireless Sensor Network observations using residual energy and Packet Delivery Ratio (PDR). The Elbow method and silhouette analysis supported a two-cluster configuration, which gave the highest average silhouette value of 0.47015.

The two clusters consist of 262 observations (26.2%) with a lower average PDR of 43.21% and 738 observations (73.8%) with a higher average PDR of 96.59%, while their average residual energy is similar (1.464 and 1.447). PDR is therefore the main feature that distinguishes the clusters. The cluster structure is reasonably clear but not fully robust, as shown by average silhouette values of 0.39910 and 0.49538 for the two clusters and a minimum silhouette value of -0.01668 . The clusters describe groups of observations with similar characteristics and should not be read as evidence of causal relationships.

Because 578 observations had PDR values above 100% that were capped before standardization, the sent and received packet counts should be examined further before the PDR results are used to assess network conditions. Future work may examine the original packet-count records and additional network attributes.

ACKNOWLEDGMENTS

The authors would like to thank all those who helped complete this article.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT (mandatory) (10 PT)

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Indri Theresia Manik	✓	✓			✓	✓	✓	✓	✓	✓		✓		
Nadicta Eunike Sitorus		✓			✓		✓		✓					
Sasta Monika Marbun			✓	✓	✓				✓		✓			
Sri Yuni Utari Sihombing						✓	✓		✓					

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT (mandatory) (10 PT)

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.

INFORMED CONSENT (if applicable) (10 PT)

Not applicable. This study used a publicly available secondary dataset of wireless sensor network node records and did not involve human participants or personal data.

ETHICAL APPROVAL (if applicable) (10 PT)

Not applicable. This study did not involve human participants or animals.

DATA AVAILABILITY (mandatory) (10 PT)

The data that support the findings of this study are openly available in Kaggle at <https://www.kaggle.com/datasets/ziya07/wsn-dataset-for-secure-and-energy-efficient>.

REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, Mar. 2002, doi: 10.1016/S1389-1286(01)00302-4.
- [2] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Comput. Netw.*, vol. 52, no. 12, pp. 2292–2330, Aug. 2008, doi: 10.1016/j.comnet.2008.04.002.
- [3] R. Zuhdianto and F. S. Mukti, "A clustering optimization for energy efficiency in wireless sensor network using K-means algorithm," *J. Tek. Inform. (JUTIF)*, vol. 4, no. 1, 2023, doi: 10.52436/1.jutif.2023.4.1.523.
- [4] W. Ahmad, A. Singh, S. Kumar, B. M. Krishna, A. Pandey, and Madhumita, "Optimizing energy efficiency in wireless sensor networks using enhanced K-means cluster head selection," *Int. J. Commun. Netw. Inf. Secur.*, vol. 16, no. 3, pp. 565–573, 2024.
- [5] B. Zeng, S. Li, and X. Gao, "Threshold-driven K-means sector clustering algorithm for wireless sensor networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2024, Art. no. 68, 2024, doi: 10.1186/s13638-024-02403-2.
- [6] A. Hilmani, Y. Sabri, A. Maizate, S. Aouad, and M. Koundi, "Integration of K-Means and Silhouette score for energy efficiency of wireless sensor networks," *Int. J. Reconfigurable Embedded Syst.*, vol. 14, no. 1, pp. 26–34, 2025, doi: 10.11591/ijres.v14.i1.pp26-34.
- [7] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Annu. Hawaii Int. Conf. Syst. Sci. (HICSS)*, 2000, pp. 1–10, doi: 10.1109/HICSS.2000.926982.
- [8] S. Periyasamy, S. Khara, and S. Thangavelu, "Balanced cluster head selection based on modified k-Means in a distributed wireless sensor network," *Int. J. Distrib. Sens. Netw.*, vol. 12, no. 3, 2016, doi: 10.1155/2016/5040475.
- [9] A. Ray and D. De, "Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network," *IET Wireless Sens. Syst.*, vol. 6, no. 6, pp. 181–191, 2016, doi: 10.1049/iet-wss.2015.0087.
- [10] M. Abdou, H. M. Amer, M. M. Abdelsalam, and A. T. Khalil, "EVRP: A novel geometrical based energy efficient eye vision routing protocol for wireless sensor networks based on the k-means algorithm," *Ad Hoc Netw.*, vol. 160, Art. no. 103528, 2024, doi: 10.1016/j.adhoc.2024.103528.
- [11] S. El Khediri, W. Fakhet, T. Moulahi, R. Khan, A. Thaljaoui, and A. Kachouri, "Improved node localization using K-means clustering for Wireless Sensor Networks," *Comput. Sci. Rev.*, vol. 37, Art. no. 100284, 2020, doi: 10.1016/j.cosrev.2020.100284.
- [12] R. L. Thorndike, "Who belongs in the family?" *Psychometrika*, vol. 18, no. 4, pp. 267–276, Dec. 1953, doi: 10.1007/BF02289263.

-
- [13] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
 - [14] Ziya07, "WSN: Dataset for Secure and Energy-Efficient," Kaggle. Accessed: Oct. 10, 2026. [Online]. Available: <https://www.kaggle.com/datasets/ziya07/wsn-dataset-for-secure-and-energy-efficient>
 - [15] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, 1967, pp. 281–297.
 - [16] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," *J. Roy. Statist. Soc. C, Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979, doi: 10.2307/2346830.
 - [17] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.
 - [18] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
 - [19] A. Shahraki, A. Taherkordi, Ø. Haugen, and F. Eliassen, "Clustering objectives in wireless sensor networks: A survey and research direction analysis," *Comput. Netw.*, vol. 180, Art. no. 107376, 2020, doi: 10.1016/j.comnet.2020.107376.
 - [20] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *Proc. IEEE INFOCOM*, vol. 3, 2003, pp. 1713–1723, doi: 10.1109/INFCOM.2003.1209194.
 - [21] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 366–379, Oct.–Dec. 2004, doi: 10.1109/TMC.2004.41.
 - [22] M. Chatterjee, S. K. Das, and D. Turgut, "WCA: A weighted clustering algorithm for mobile ad hoc networks," *Cluster Comput.*, vol. 5, pp. 193–204, 2002, doi: 10.1023/A:1013941929408.
 - [23] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining (KDD)*, 1996, pp. 226–231.
 - [24] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2026. [Online]. Available: <https://www.R-project.org/>
 - [25] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, "cluster: Cluster Analysis Basics and Extensions," R package, 2025. [Online]. Available: <https://CRAN.R-project.org/package=cluster>
-

