



## Analysis of Household Electricity Consumption Patterns Using K-Means Clustering

Fika Saputri<sup>1\*</sup>, Asminar<sup>2</sup>, Tambi<sup>3</sup>, Mustarum Musaruddin<sup>4</sup>, Muhammad Nadzirin Anshari Nur<sup>5</sup>, Adhi Setiawan Samsul<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Department of Electrical Engineering, Universitas Halu Oleo, Indonesia

Correspondence E-mail: [fikasaputri29@gmail.com](mailto:fikasaputri29@gmail.com)

### ABSTRACT

The continuous increase in household electricity consumption and irregular usage patterns pose challenges for efficient energy management. Irregular electricity load patterns can lead to energy waste, increased costs, and disruptions to the distribution system. Therefore, a method capable of accurately identifying electricity consumption patterns is needed. This study aims to analyze household electricity consumption patterns using the K-Means Clustering algorithm. The data used comes from PLN UP3 Kendari in the form of historical data for the 2023–2025 period and real-time data processed into daily parameters. The K-Means method was used to cluster the data based on the similarity of electricity consumption characteristics. The determination of the optimal number of clusters was performed using the Elbow Method and the Silhouette Score. The results of the study indicate that the optimal number of clusters is four, with a Silhouette Score of 0.5122 and a Davies-Bouldin Index of 0.6881, indicating that the cluster quality is quite good. Each cluster represents a distinct consumption pattern: low, stable medium, high, and fluctuating. Thus, the K-Means Clustering algorithm is effective for identifying household electricity consumption patterns.

### ARTICLE INFO

#### Article History:

Submitted/Received 31 Mar 2026

First Revised 07 Oct 2026

Accepted 20 Oct 2026

First Available online 22 Oct 2026

Publication Date 22 Oct 2026

#### Keyword:

K-Means Clustering,  
Electricity Consumption,  
Electricity Load Patterns,  
Load Profiling,  
Data Mining

## 1. INTRODUCTION

Residential electricity consumption continues to rise in tandem with population growth and technological advancements. This increase poses challenges for efficient energy management, including energy waste, rising operational costs, and environmental impacts resulting from emissions from fossil-fuel-based power plants [1][2].

In the Kendari region, electricity consumption patterns are influenced by diverse community activities, resulting in fluctuating and unpredictable electricity load patterns [1][3]. These irregular electricity load patterns can cause problems in the power distribution system, such as voltage drops, increased power losses, and network stability disturbances. Therefore, a method is needed that can identify and group electricity consumption patterns.

Previous studies have employed clustering methods to identify patterns in electricity consumption [4], one of which is the K-Means Clustering algorithm, known for its simplicity and efficiency in grouping data based on similar characteristics [2][5][6]. Furthermore, cluster quality evaluation is generally performed using the Silhouette Score and the Davies-Bouldin Index to ensure optimal clustering results [7][8]. However, the application of this method to residential electricity consumption data, taking into account daily load characteristics and distribution system operational parameters, still requires further study to obtain representative clustering results.

Based on these issues, this study aims to analyze household electricity consumption patterns using the K-Means clustering algorithm. The results of this study are expected to provide an overview of the characteristics of electricity consumption patterns and serve as a basis for decision-making to improve the efficiency and reliability of the electricity distribution system. Clustering approaches are also continuously evolving and are used in various fields of data analysis to improve the accuracy of data grouping [9].

## 2. METHODS

This study uses household electricity consumption data obtained from PLN UP3 Kendari, consisting of historical data for the 2023–2025 period and real-time data, which were processed into daily parameters. The variables used include date, feeder name, voltage, current, active power, active energy, power factor, system frequency, peak load, load factor, peak load duration, and number of customers, which are used as parameters to describe electricity consumption characteristics. The research stages were conducted systematically, as shown in **Figure 1**.

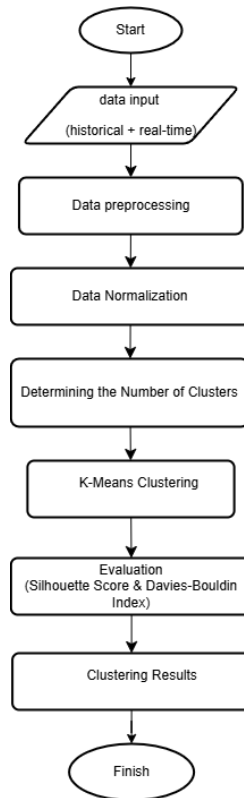


**Figure 1.** Research flowchart.

Based on **Figure 1**, the research stages are described as follows:

1. Data Input  
The data collection process, consisting of historical and real-time data, which is used as the basis for analyzing electricity consumption patterns.
2. Cleaning Data  
Removing invalid data and duplicates, as well as handling missing data, to ensure data quality is maintained.
3. Data Normalization  
Standardizing the scale across variables to prevent any single variable from dominating the clustering process.
4. Determining the Number of Clusters  
The optimal number of clusters is determined using the Elbow Method and Silhouette Score to identify the most representative number of clusters.
5. K-Means Clustering  
The clustering process is performed using the K-Means algorithm, which groups data [10][11] based on their distance from the centroids. The distance between data points is calculated using Euclidean distance, which is defined as follows [12]:

$$d(x, \mu) = \sqrt{\sum_{i=1}^n (x_i - \mu_i)^2}$$

Where  $x_i$  is the data value on the  $i$ -th dimension,  $\mu_i$  is the cluster centroid on the  $i$ -th dimension, and  $n$  is the number of variables.

6. Evaluation

The clustering results were evaluated using the Silhouette Score [7][13] and the Davies-Bouldin Index (DBI) [8][14] to assess cluster quality. The Silhouette Score is defined as follows [7]:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Where  $a(i)$  is the average distance of the data points from the same cluster, while  $b(i)$  is the average distance to the nearest cluster. A value close to 1 indicates good cluster quality. In addition to the Silhouette Score, the Davies-Bouldin Index (DBI) is used, which is defined as follows [8]:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{S_i + S_j}{M_{ij}} \right)$$

Where  $k$  is the number of clusters,  $S_i$  and  $S_j$  are the average distances between data points within the- $i$  and  $k-j$  clusters, respectively, and  $M_{ij}$  is the distance between the centroids of the- $i$  and the- $j$  clusters.

### 7. Clustering Results

The final step involves interpreting the results of the data clustering based on the observed characteristics of electricity consumption.

## 3. RESULTS AND DISCUSSION

The household electricity consumption data used in this study were obtained from PLN UP3 Kendari for the period 2023–2025. After data cleaning, a total of 5,305 data points were used in the analysis.

The number of clusters was determined using the Elbow Method and the Silhouette Score. Based on the analysis results, the optimal number of clusters was determined to be four ( $k=4$ ). A Silhouette Score of 0.5122 indicates that the cluster quality is in the fairly good category [15], while a Davies-Bouldin Index (DBI) value of 0.6881 indicates that the separation between clusters is quite clear.

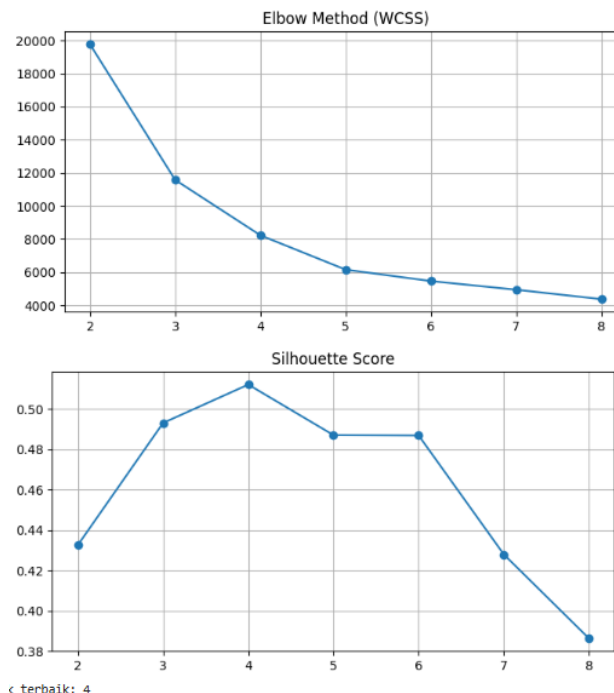


**Figure 2.** Determining the optimal number of clusters.

Based on **Figure 2**, the results of clustering using the K-Means algorithm show that household electricity consumption data can be grouped into four clusters with distinct characteristics. These differences indicate variations in electricity consumption patterns among residential customers. Clustering methods have been widely used in electricity consumption analysis to identify energy usage patterns [16][17][18][19]. To clarify the characteristics of each cluster, a summary of the cluster profiles is presented in Table 1.

**Table 1.** Cluster profile.

| cluster | tegangan_rata2 | arus_rata2 | daya_aktif | energi_aktif | faktor_daya | frekuensi_sistem | beban_puncak | load_factor | durasi_puncak_jam | jumlah_pelanggan |
|---------|----------------|------------|------------|--------------|-------------|------------------|--------------|-------------|-------------------|------------------|
| 0       | 21000.0        | 41.229113  | 1.261050   | 30.265203    | 0.85        | 50.0             | 1.774525     | 71.751309   | 17.220314         | 9822.030784      |
| 1       | 21000.0        | 181.449879 | 5.549899   | 133.197566   | 0.85        | 50.0             | 7.614046     | 73.279439   | 17.587065         | 22889.244032     |
| 2       | 21000.0        | 116.586466 | 3.565960   | 85.583047    | 0.85        | 50.0             | 4.462256     | 80.142676   | 19.234242         | 15368.667759     |
| 3       | 21000.0        | 87.809874  | 2.685788   | 64.458910    | 0.85        | 50.0             | 5.623405     | 47.711655   | 11.450797         | 19707.785311     |

Based on **Table 1**, the clustering results using the K-Means algorithm show that the data is divided into four clusters with distinct characteristics. Cluster 0 exhibits the lowest values for current, active power, and active energy, and is therefore categorized as a cluster with low energy consumption. Conversely, Cluster 1 has the highest parameter values and the largest number of customers, indicating the highest level of electricity consumption.

Cluster 2 represents medium-level energy consumption with parameter values falling between the low and high clusters, and has a relatively high load factor. Meanwhile, Cluster 3 exhibits a more fluctuating consumption pattern, characterized by a lower load factor and

a shorter peak load duration. To reinforce the results of the analysis, representative feeder profiles for each cluster were visualized, as shown in **Figure 3**.

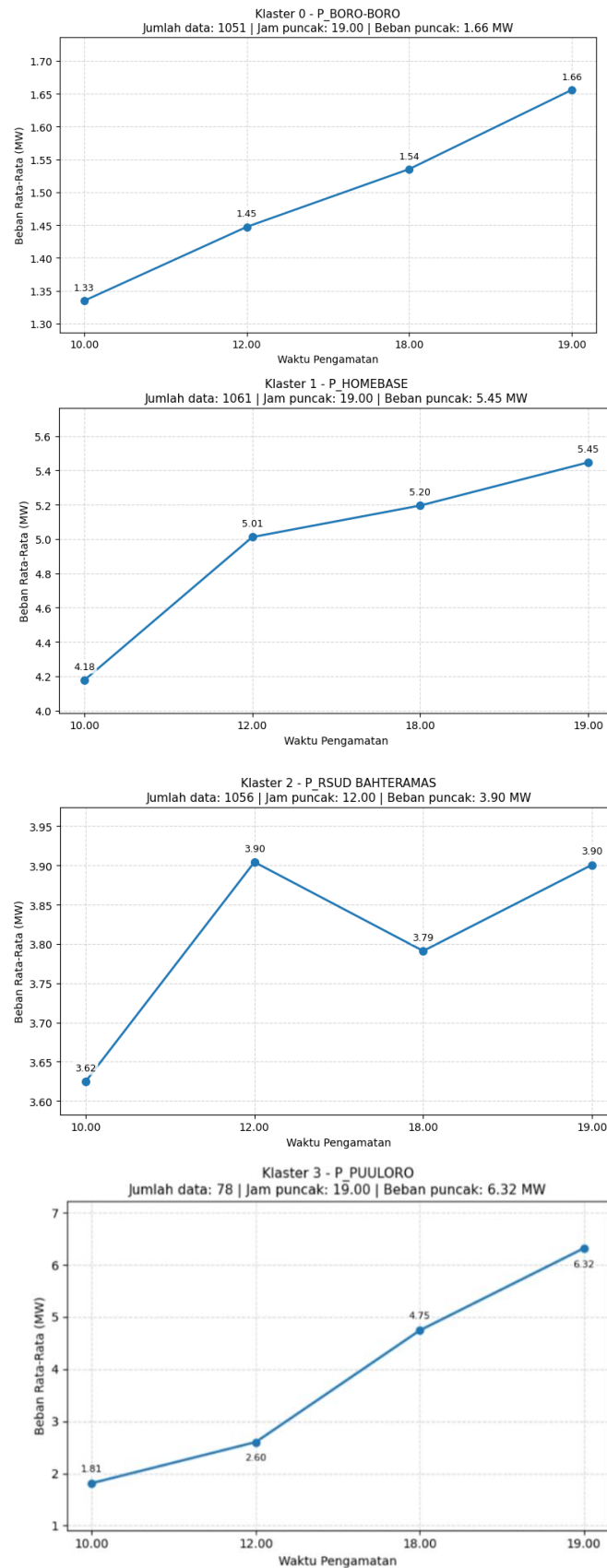


**Figure 3.** Profile of pepresentative distributors for each cluster.

**Figure 3**, shows the average load profile of representative feeders in each cluster based on observation time. In general, all clusters exhibit a trend of increasing load toward peak times, though with different characteristics.

Clusters 0 and 1 show a relatively stable increase until the evening peak, while Cluster 2 exhibits a more dynamic pattern with a peak during the day. Meanwhile, Cluster 3 shows the most significant increase with the highest peak load.

These differences indicate variations in electricity consumption patterns within each cluster, both in terms of load levels and peak times. This aligns with previous research showing that clustering methods can effectively identify electricity consumption patterns [20].

#### 4. CONCLUSION

This study successfully grouped household electricity consumption patterns using the K-Means Clustering algorithm into four clusters with distinct characteristics. The evaluation results showed a Silhouette Score of 0.5122 and a Davies-Bouldin Index of 0.6881, indicating that the quality of the clustering falls into the “fairly good” category.

Each cluster represents a distinct consumption pattern—low, medium, high, and fluctuating—with differences in load levels and the timing of peak loads.

Overall, the K-Means method is effective in identifying electricity consumption patterns and can serve as a foundation for load management and electricity distribution planning.

#### 5. ACKNOWLEDGMENT

The author would like to thank PT PLN (Persero) UP3 Kendari for its support and for providing data for this study. The author also extends gratitude to the academic advisor for providing guidance throughout the research process. Furthermore, the author would like to thank Halu Oleo University for its academic support.

#### 6. AUTHORS' NOTE

The authors declare that there are no conflicts of interest regarding the publication of this article. The authors affirm that this journal is free of plagiarism

#### 7. REFERENCES

- [1] F. Haykal, R. R. Hariadi, and K. Ghazali, “Rancang Bangun Sistem Pemantauan dan Forecasting Konsumsi Energi Listrik Menggunakan Internet of Things dan Algoritma Seasonal Time Series,” *Jurnal Teknik ITS*, vol. 11, no. 2, pp. A103–A109, 2022.
- [2] A. N. A. Maulidhia, I. I. Widyastuti, F. I. Sukarno, R. B. S. Tsany, and T. Brian, “Implementation of the K-Means Algorithm on Household Electric Load,” *Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem dan Komputer*, vol. 5, no. 1, pp. 17–24, 2025.
- [3] G. Okereke et al., “K-means clustering of electricity consumers using smart meter data,” *Journal of Electrical Systems and Information Technology*, 2023.
- [4] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: A review,” *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
- [5] G. Chicco, “Overview and performance assessment of the clustering methods for electrical load pattern grouping,” *Energy*, vol. 42, no. 1, pp. 68–80, 2012.
- [6] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2011.

- [7] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [8] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.
- [9] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 136–144, 2020.
- [10] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Proc. Fifth Berkeley Symp. Mathematical Statistics and Probability*, 1967, pp. 281–297.
- [11] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.
- [12] F. Falanda, R. Gustriansyah, and H. Hartini, "Penentuan Objek Wisata, Objek Kuliner Serta Akomodasi Disekitar Pengguna Dikota Palembang Dengan Menggunakan Algoritma Euclidean Distance," *Jurnal Ilmiah Informatika Global*, vol. 7, no. 2, Dec. 2016.
- [13] Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan DBSCAN," *J. Inform. Tek. Elektro Terap.*, vol. 12, no. 3S1, 2024.
- [14] L. Hanum, "Pengelompokan Gaya Belajar Mahasiswa Menggunakan Metode K-Means dan Validasi Menggunakan Davies Bouldin Index," *Jurnal Manajemen, Pendidikan dan Ilmu Komputer*, vol. 2, no. 2, pp. 37–47, 2025.
- [15] Y. Zhang, "Modified silhouette score with generalized mean and trimmed mean," Ph.D. dissertation, 2023.
- [16] R. Wu, "Behavioral analysis of electricity consumption characteristics for customer groups using the k-means algorithm," *Syst. Soft Comput.*, vol. 6, p. 200143, 2024.
- [17] H. Yu et al., "Corn leaf diseases diagnosis based on K-means clustering and deep learning," *IEEE Access*, vol. 9, pp. 143824–143835, 2021, doi: 10.1109/ACCESS.2021.3120379
- [18] S. Noertjahjani, D. P. Setyawan, and A. Kiswanto, "Analysis of Household Electricity Consumption Segmentation using K-means Clustering," *Int. J. Inform. Comput.*, vol. 7, no. 2, pp. 449–461, 2025.
- [19] G. Chicco, R. Napoli, and F. Piglion, "Comparisons among clustering techniques for electricity customer classification," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 933–940, 2006.
- [20] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420–430, 2014.