BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1123-1133
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1119

Clustering State Electricity Company (PLN) Customer Electricity Consumption Patterns Using K-Means for Operational Efficiency

Anang Putranto*, Imam Much Ibnu Subroto, Arief Marwanto

Faculty of Industrial Technology, Master of Electrical Engineering, Universitas Islam Sultan Agung, Semarang, Indonesia
Email: [1,*]putrantoanang@unissula.ac.id, [2]imam@unissula.ac.id, [3]arief@unissula.ac.id
Corresponding Author Email: putrantoanang@unissula.ac.id

Abstract−Electricity consumption patterns among PLN customers show diverse characteristics that require data-driven analysis to support accurate operational planning and service strategies. Differences in installed power, operating hours, and electricity usage intensity can indicate variations in customer behavior, load demand, and service needs. This study aims to analyze electricity consumption patterns of PLN customers in the South Semarang area using the K-Means Clustering method. The dataset consists of 6,622 active customers, with variables including installed power, operating hours, and average electricity consumption over the last eight months. The research stages include identifying the need for consumption pattern analysis, preparing data, cleaning incomplete or inconsistent values, normalizing variables, modeling with K-Means, and interpreting cluster results. The optimal number of clusters was determined using the Elbow Method and Silhouette Score to ensure meaningful grouping and good separation quality. The results show that the optimal number of clusters is three. The first cluster represents low consumption patterns with 1,881 customers (28.4%). The second cluster represents medium consumption patterns with 2,927 customers (44.2%). The third cluster represents high consumption patterns with 1,814 customers (27.4%). The WCSS value of 498.67 and Silhouette Score of 0.68 indicate good clustering performance and practical usefulness for decision-making by PLN in service quality improvement.

Keywords: Electricity Consumption Pattern; K-Means Clustering; PLN Customers; Operational Efficiency; Service Strategy

1. INTRODUCTION

The development of electricity demand requires utility companies to manage customer data more carefully. PT Perusahaan Listrik Negara (PLN), as an electricity service provider, has a customer base with highly diverse consumption characteristics, ranging from small household customers and medium-sized businesses to commercial and industrial customers. This diversity causes general service strategies to be less effective because the needs, usage intensity, and load contribution of each customer are not the same. Therefore, understanding electricity consumption patterns is an important aspect in supporting operational efficiency and designing customer services that better match the characteristics of each group [1].

Electricity consumption patterns can describe customer energy usage behavior based on technical variables such as installed power, operating hours, and average electricity consumption. Installed power indicates the electricity capacity registered for a customer, operating hours describe the intensity or duration of electricity usage, while average consumption indicates the amount of energy consumed within a certain period [2], [3]. These three variables can be used to quantitatively observe customer consumption tendencies. If these consumption patterns are analyzed systematically, PLN can obtain more structured information regarding low, medium, and high consumption customer groups [4], [5], [6].

The main problem addressed in this study is the suboptimal utilization of customer electricity consumption data as a basis for operational segmentation and service strategies. Large customer datasets are often used only for administrative data or routine reporting, even though these data have the potential to provide important information for decision making [7], [8]. Without clear grouping of consumption patterns, energy efficiency programs, load monitoring, and customer services risk being poorly targeted [7], [9], [10]. This condition may lead to less effective allocation of operational resources, difficulty in determining customer monitoring priorities, and less specific service communication strategies for customers [6], [11].

One approach that can be used to analyze electricity consumption patterns is data mining with clustering methods. Clustering is an unsupervised learning technique that aims to group data based on similarity of characteristics without requiring predefined class labels [12], [13]. Among various clustering algorithms, K-Means is widely used because it has a simple concept, is efficient for large datasets, and is easy to interpret. This algorithm works by dividing data into a number of clusters based on distance proximity to centroids so that members within the same cluster have relatively homogeneous characteristics [14]. Various K-Means developments have also continued to improve cluster separation quality, computational efficiency, and robustness to outliers [15], [16].

Related studies show that clustering can be applied to group electricity consumption and customer load profiles. Al-Wakeel et al. applied K-Means for load estimation based on household smart meter data so that electricity usage patterns could be mapped more clearly[17], [18]. McLoughlin et al. used a clustering approach to characterize domestic electricity load profiles based on smart metering data [3]. Palaniappan et al. categorized residential electricity consumption patterns using clustering and classification techniques [19]. Other studies also show that customer data and smart meter-based clustering can support service segmentation, consumption behavior analysis, and customer management strategies in the electric utility sector [20].

Although several studies have discussed the use of clustering in energy data, there remains room for development in the context of utilizing PLN customer electricity consumption patterns in the South Semarang area. Previous studies have tended to focus on algorithm implementation or general segmentation, while this study emphasizes the relationship

Copyright © 2026 The Author, Page 1123
This Journal is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1123-1133
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1119

between consumption pattern clustering results and operational implications as well as customer service strategies. Thus, the clustering results do not stop at cluster formation but are translated into a basis for recommendations that can be used for load management, energy efficiency education, customer monitoring, and segment-based service design [4], [5], [6].

This study aims to analyze PLN customer electricity consumption patterns using the K-Means Clustering method based on installed power, operating hours, and average electricity consumption variables. This study also aims to determine the optimal number of clusters, interpret the characteristics of each cluster, and formulate operational recommendations and service strategies based on the resulting consumption patterns. The contribution of this study lies in presenting a systematic analysis flow from data understanding, modeling, evaluation, interpretation, to the use of clustering results to support operational efficiency and improve PLN service quality in a more targeted manner [9], [10], [21].

2. RESEARCH METHODOLOGY

2.1 Research Stages

This study uses a quantitative approach based on data mining with an unsupervised learning method. The main method used is K-Means Clustering because the study aims to group customers based on the similarity of electricity consumption patterns without prior class labels. The research stages are designed to answer the main problem, namely how customer electricity consumption data can be processed into meaningful consumption pattern groups and used to support PLN operational efficiency and customer service strategies [5].

The first process is identifying the need for electricity consumption pattern analysis. At this stage, the problem is formulated based on the condition that PLN customers have variations in power, operating hours, and electricity usage. These different characteristics require a clustering approach so that PLN can understand customers in a more structured way. The output of this process is the formulation of research objectives, data scope, variables used, and modeling needs directed at forming low, medium, and high consumption clusters [6], [11].

The second process is customer data preparation and processing. Active customer data are selected according to variables relevant to electricity consumption patterns. This stage includes checking data completeness, validating numeric data types, checking duplicates, identifying extreme values, selecting features, and normalizing data. Normalization is required because the variables of installed power, operating hours, and electricity consumption have different units and value ranges. The output of this process is a clean and standardized dataset ready to be used in clustering modeling [9], [10]

The third process is consumption pattern modeling using the K-Means algorithm. At this stage, several k values are tested to observe changes in the Within-Cluster Sum of Squares (WCSS). The optimal number of clusters is determined using the Elbow Method and validated with the Silhouette Score. After the number of clusters is determined, the K-Means algorithm is executed to produce a cluster label for each customer. The output of this process is the division of customers into electricity consumption pattern groups based on the proximity of their characteristics to each cluster centroid [22], [23], [24].

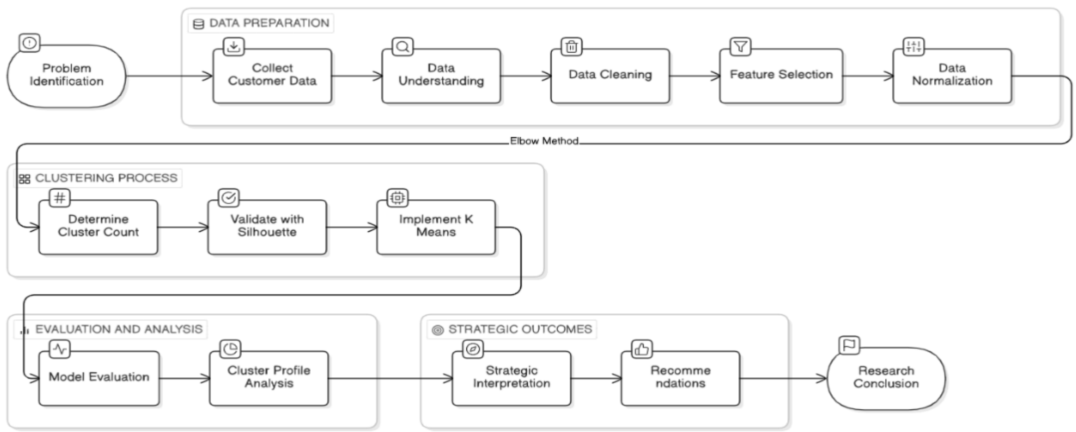

Figure 1. Research stages of PLN customer electricity consumption pattern analysis

The fourth process is result interpretation and recommendation formulation. The resulting clusters are analyzed based on the number of members, percentages, centroids, and average values of each variable. This interpretation is used to assign meaning to each cluster, such as low, medium, and high consumption patterns. The interpretation results are then linked to PLN problems so that operational recommendations can be formulated, such as load monitoring, energy efficiency education, and customer service strategies based on electricity consumption characteristics [10], [21].

Figure 1 shows the research flow consisting of four main processes. These four processes are interconnected, beginning with the identification of analysis needs and ending with the translation of clustering results into operational

This Journal is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1123-1133
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1119

recommendations and service strategies. This flow ensures that the research process does not only produce clusters technically, but also addresses the problem-solving needs raised in the study [13].

2.2 Data Sources, Unit of Analysis, and Data Scope

The research dataset was obtained from active PLN customers in the South Semarang area. The unit of analysis in this study is the individual customer, so each row of data represents one customer with specific electricity consumption characteristics. The data used include technical information directly related to electricity consumption, namely installed power, operating hours, and average electricity consumption over the last eight months [6], [11].

The data scope is limited to active customers who have complete numerical data for the analyzed variables. This limitation is applied so that the clustering process is not influenced by irrelevant or incomplete data. This study does not use customer satisfaction survey data, complaint data, or service quality data, so the term customer satisfaction in this article is positioned as a strategic implication of improved service accuracy, not as a directly measured variable [7].

Table 1. Summary of data sources and scope

Aspect	Description
Data source	Active PLN customer data in the South Semarang area
Unit of analysis	Individual customers
Final data size	6,622 active customers
Consumption period	Average electricity usage over the last 8 months
Main variables	Installed power, operating hours, and average electricity consumption
Data limitations	Does not include satisfaction surveys, complaint data, and service quality data

2.3 Variables and Operational Definitions

The research variables were selected based on their relationship to customer electricity consumption patterns. All variables used are numerical and measured on a ratio scale, so they can be processed using Euclidean distance in the K-Means algorithm. Variable selection also considers the ease of interpreting cluster results so that the analysis can be used as a basis for PLN operational recommendations.

Table 2. Research variables

Variable	Operational Definition	Scale
Installed Power	Registered customer electricity capacity in VA	Ratio
Operating Hours	Estimated duration of customer electricity usage in hours	Ratio
Average Consumption	Average customer electricity consumption in kWh over the last 8 months	Ratio

2.4 Data Processing Stages

Data processing stages are carried out to ensure that the dataset is suitable for clustering. Data cleaning is performed by checking duplicates, missing values, data type consistency, and extreme values that may affect centroid positions. Because K-Means is sensitive to distance and feature scale, data checking and variable scaling are important stages to prevent the resulting clusters from being biased by unrepresentative values [22], [23].

After the data are declared suitable, feature selection is performed to select variables relevant to the research objective. This study uses three main variables, namely installed power, operating hours, and average consumption [9], [14], [20]. These three variables are selected because they directly represent the intensity and capacity of customer electricity consumption. The next stage is normalization using Min-Max Scaling so that all variables are within the same range, from 0 to 1.

$$X' = (X - Xmin) / (Xmax - Xmin) \quad (1)$$

In Equation (1), X is the original value, Xmin is the minimum value, Xmax is the maximum value, and X' is the normalized value. Normalization is performed so that variables with large value ranges, such as installed power and electricity consumption, do not dominate other variables in Euclidean distance calculations. Feature scaling in K-Means is important, especially when data have different units and value ranges [16].

2.5 Modeling Using the K-Means Algorithm

K-Means Clustering is used to divide customers into several clusters based on the similarity of electricity consumption patterns. This algorithm begins by determining the number of clusters k, selecting initial centroids, calculating the distance of each data point to the centroids, assigning data to the nearest cluster, and updating the centroids until the centroid positions become stable or the maximum iteration limit is reached [22], [25], [26]. In this study, K-Means is selected because its results are easy to interpret and suitable for operational segmentation objectives, although several modern K-Means developments can be used to improve performance on more complex data [27], [28], [29].

The main objective of K-Means is to minimize data variation within clusters. Thus, customers within the same cluster are expected to have relatively similar consumption characteristics. The K-Means objective function is written as follows.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1123-1133
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1119

$$J = \Sigma(i=1 \text{ to } k)\ \Sigma(x_j \in C_i)\ ||x_j - \mu_i||^2 \quad (2)$$

In Equation (2), Ci is the i-th cluster, μi is the centroid of the i-th cluster, xj is the j-th customer data point, and $||x_j - \mu_i||^2$ is the squared Euclidean distance between the customer data and the centroid. The smaller the objective function value, the more homogeneous the data within the formed cluster.

2.6 Determining the Optimal Number of Clusters

The optimal number of clusters is determined using the Elbow Method and Silhouette Score. The Elbow Method analyzes changes in WCSS values for several k values. The number of clusters is selected at the point where the decrease in WCSS begins to slow significantly, so adding the next cluster no longer provides a major improvement in clustering quality [30].

$$WCSS = \Sigma(i=1 \text{ to } k)\ \Sigma(x_j \in C_i)\ ||x_j - \mu_i||^2 \quad (3)$$

WCSS measures the total variation of data within clusters. A smaller WCSS value indicates that the data are closer to their respective cluster centroids. However, WCSS always decreases as the number of clusters increases, so additional validation using the Silhouette Score is required.

$$s(i) = (b(i) - a(i)) / \max\{a(i), b(i)\} \quad (4)$$

In Equation (4), a(i) is the average distance of the i-th data point to other data points in the same cluster, while b(i) is the nearest average distance of the i-th data point to another cluster. A Silhouette Score close to 1 indicates that the data point is 1 on electricity consumption characteristics and linked to appropriate operational recommendations. In this way, clustering results can be used as a basis for more applicable strategy formulation [25], [26], [27].

3. RESULTS AND DISCUSSION

The results and discussion are organized according to the flow of solving the research problem. The discussion begins with an overview of the dataset, normalization results, determination of the optimal number of clusters, clustering results, cluster distribution, characteristics of each consumption pattern, and the implications of the results for operational ezfficiency and customer service strategies. This flow is used so that the research results do not only present algorithm outputs but also explain how the results can be used to address PLN problems.

3.1 Dataset Overview

The research dataset consists of 6,622 active PLN customers in the South Semarang area. The data used focus on three main variables: installed power, operating hours, and average electricity consumption over the last eight months. These three variables were selected because they can describe the capacity, intensity, and volume of customer electricity consumption. By using eight-month average data, the analyzed consumption patterns are expected to be more stable than those based on only one month of data.

Installed power describes the maximum capacity that customers can use according to their service contract. Operating hours indicate the intensity of electricity usage, while average consumption indicates the amount of energy used. The combination of these three variables allows a more complete picture of consumption patterns to be formed. For example, customers with high installed power do not necessarily have high consumption if their operating hours are low, while customers with medium installed power may show high consumption if their usage intensity is high.

Table 3. Sample raw data of PLN customers

No.	Power (VA)	Operating Hours	Average Consumption (kWh)
1	33000	180	5921.00
2	33000	101	3202.75
3	23000	218	5013.75
4	23000	232	4660.00
5	23000	159	3200.00
...	...	...	...
6622	450	2	16.38

The sample data in Table 3 show that there is considerable variation among customers, especially in installed power and average electricity consumption. This variation indicates the need for grouping so that customers with similar characteristics can be analyzed as a single group. Without clustering, low- and high-consumption customers would be treated generally, causing the strategic information contained in the data to be less visible.

Table 4. Summary of analysis strengthening

Analysis Aspect	Description	Implication for the Study
Dataset size	6,622 active customers	Sufficiently large for consumption pattern-based clustering analysis

Copyright © 2026 The Author, Page 1126
This Journal is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1123-1133
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1119

Analysis Aspect	Description	Implication for the Study
Data type	Numerical and ratio-scale	Suitable for K-Means, which uses distance calculations
Main variables	Power, operating hours, and average consumption	Represents the capacity, intensity, and volume of electricity consumption
Variable limitation	Does not include satisfaction survey or complaint data	Results are directed toward service strategies, not direct satisfaction measurement

3.2 Data Normalization Results

Before clustering is performed, the data are normalized using the Min-Max Scaling method. Normalization is important because installed power, operating hours, and average consumption variables have different value ranges. If the data are not normalized, the variable with the largest values may dominate the distance calculation and influence cluster formation [24].

The normalization results transform all variable values into the range of 0 to 1. Thus, each variable has a more proportional contribution in the clustering process. Table 5 shows examples of normalization results for several customer records.

Table 5. Min-Max normalization results

No.	Power (norm)	Operating Hours (norm)	Consumption (norm)
1	1.0000	0.2532	0.9823
2	1.0000	0.1391	0.5305
3	0.6969	0.3066	0.8304
4	0.6969	0.3263	0.7720
5	0.6969	0.2236	0.5301
...	...	...	...
6622	0.0000	0.0000	0.0015

Table 5 shows that installed power, operating hours, and consumption values are already on the same scale. This condition makes the K-Means process fairer because the distance between customers is calculated based on the relative contribution of each variable. Normalization also helps reduce the risk of bias due to differences in measurement units [31].

3.3 Determining the Optimal Number of Clusters

The optimal number of clusters is determined through the Elbow Method by comparing WCSS values for several k values. Testing was performed from k=2 to k=8. The determination of the number of clusters does not only consider the smallest WCSS value, because this value will continue to decrease as the number of clusters increases. Therefore, the point at which a significant decline begins to flatten becomes the main consideration in selecting the number of clusters [23].

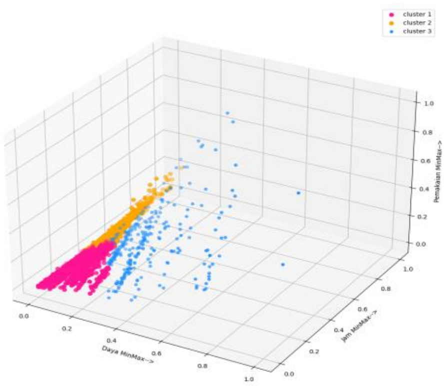

Figure 2. Supporting visualization for cluster determination and results

Table 6. WCSS calculation results

Number of Clusters (k)	WCSS
2	1847.32
3	498.67
4	412.85
5	358.24
6	321.45
7	295.78
8	274.32

Copyright © 2026 The Author, Page 1127
This Journal is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1123-1133
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1119

Based on Table 6, the WCSS value decreases from 1,847.32 at k=2 to 498.67 at k=3. A decrease of approximately 73% indicates that adding a cluster from two to three provides a highly significant improvement in clustering quality. After k=3, the WCSS decrease becomes more gradual, for example from 498.67 at k=3 to 412.85 at k=4. This condition indicates that adding subsequent clusters does not provide an improvement in quality as large as the previous change.

The Elbow result is then strengthened using the Silhouette Score. The Silhouette Score value of 0.68 at k=3 indicates that the formed clusters have good separation. This value indicates that most customers are fairly close to members of their own cluster and sufficiently far from other clusters. Thus, the optimal number of clusters in this study is set at three clusters [17], [18].

3.4 Clustering Results of Electricity Consumption Patterns

After the optimal number of clusters is determined, the K-Means algorithm is run with k=3. The clustering results show that PLN customers can be grouped into three electricity consumption patterns: low consumption, medium consumption, and high consumption. Cluster naming is based on the average values of installed power, operating hours, and average electricity consumption in each cluster.

The visualization of clustering results shows relatively clear separation among groups. Customers with low installed power, operating hours, and consumption tend to gather in one area, while customers with medium and high characteristics form different cluster areas. This indicates that the variables used are sufficiently capable of distinguishing customer consumption patterns.

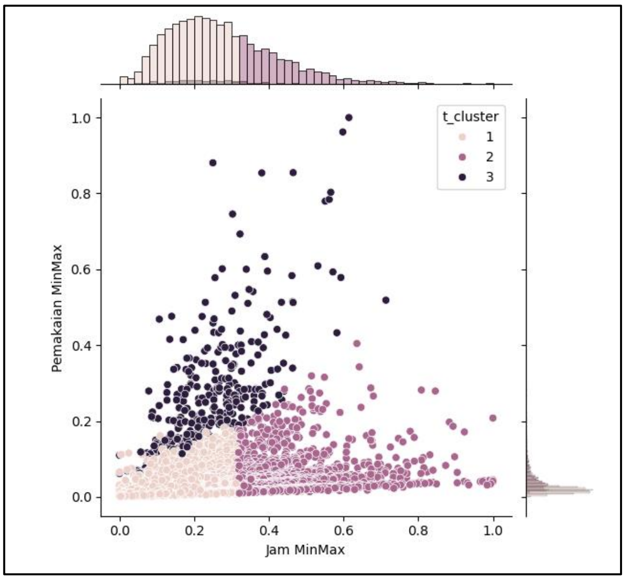

Figure 3. Relationship between average consumption and operating hours

Figure 3 shows a tendency that customers with higher operating hours generally have greater average electricity consumption. This pattern strengthens the interpretation that usage intensity is one of the important factors in distinguishing electricity consumption clusters.

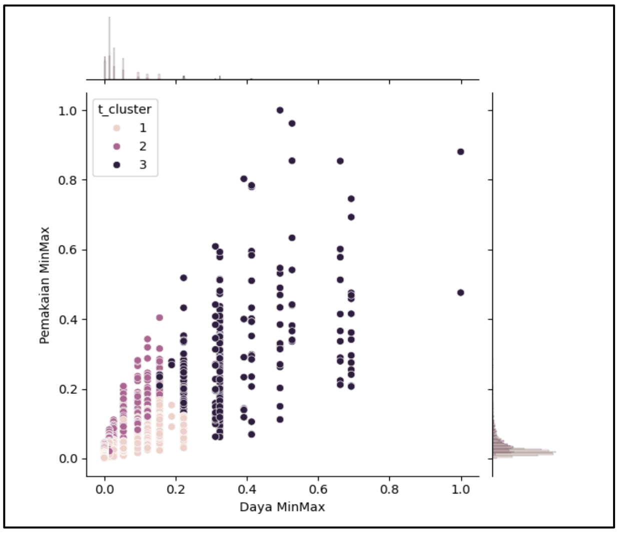

Figure 4. Relationship between average consumption and installed power

Copyright © 2026 The Author, Page 1128
This Journal is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1123-1133
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1119

Figure 4 shows that customers with high installed power tend to have greater electricity consumption potential. However, variations among data points also show that high installed power does not automatically mean high consumption, so its combination with operating hours remains necessary.

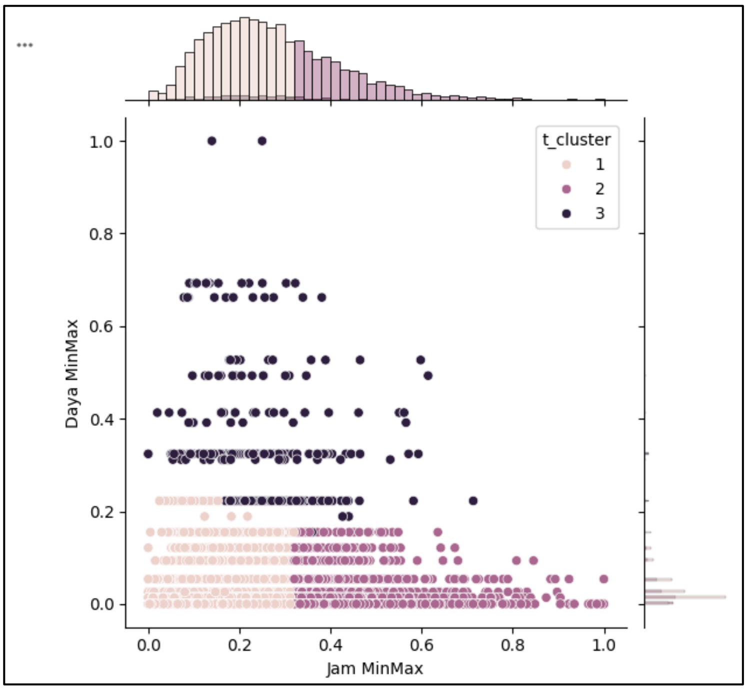

Figure 5. Relationship between installed power and operating hours

Figure 5 shows the relationship between power capacity and usage intensity. Clusters with low installed power and operating hours are located in different areas from clusters with high installed power and operating hours, making the separation of consumption patterns easier to interpret.

3.5 Cluster Distribution

Customer distribution in each cluster shows relatively balanced proportions. No cluster is overly dominant, although the medium consumption cluster has the largest number of members. This distribution is important because it shows that the clustering results do not only divide a small portion of customers, but can identify groups with sufficiently significant sizes for further analysis.

Table 7. Distribution of electricity consumption pattern clusters

Cluster	Number of Customers	Percentage	Interpretation
Cluster 1	1,881	28.4%	Low consumption pattern
Cluster 2	2,927	44.2%	Medium consumption pattern
Cluster 3	1,814	27.4%	High consumption pattern
Total	6,622	100.0%	-

Cluster 2 is the largest group with 2,927 customers or 44.2% of the total data. This shows that the majority of customers are in the medium consumption pattern. Cluster 1 consists of 1,881 customers or 28.4%, while Cluster 3 consists of 1,814 customers or 27.4%. The proportion of Cluster 3, which reaches 27.4%, is an important concern because customers in this cluster have the potential to contribute significantly to system load and operational monitoring needs.

3.6 Characteristics of Each Consumption Pattern

The characteristics of each cluster are analyzed based on the final centroid or the average variable values within the cluster. Centroids help explain the main profile of each group so that clustering results can be more easily translated into recommendations. Table 8 shows a summary of the characteristics of each cluster based on average installed power, operating hours, and electricity consumption [10], [14].

Table 8. Final centroids and cluster interpretation

Cluster	Average Power	Average Operating Hours	Average Consumption	Interpretation
Cluster 1	1,320 VA	45 hours	78 kWh	Low consumption pattern
Cluster 2	6,800 VA	128 hours	1,245 kWh	Medium consumption pattern
Cluster 3	18,500 VA	198 hours	4,567 kWh	High consumption pattern

This Journal is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1123-1133
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1119

3.6.1 Low Consumption Pattern

Figure 6. Profile of low consumption pattern cluster

Cluster 1 consists of 1,881 customers or 28.4% of the total data. This group has an average installed power of 1,320 VA, average operating hours of 45 hours, and average consumption of 78 kWh. These characteristics indicate that customers in this cluster have relatively low electricity usage intensity. This group may consist of small household customers or customers with limited electricity usage.

The operational implication of this cluster is the need for service strategies focused on energy efficiency education and the provision of easily understandable consumption information. Although customer consumption is low, the number of customers in this group remains significant, so appropriate communication strategies can help maintain service quality. PLN can also use this group as a target for safe and efficient electricity usage literacy programs.

3.6.2 Medium Consumption Pattern

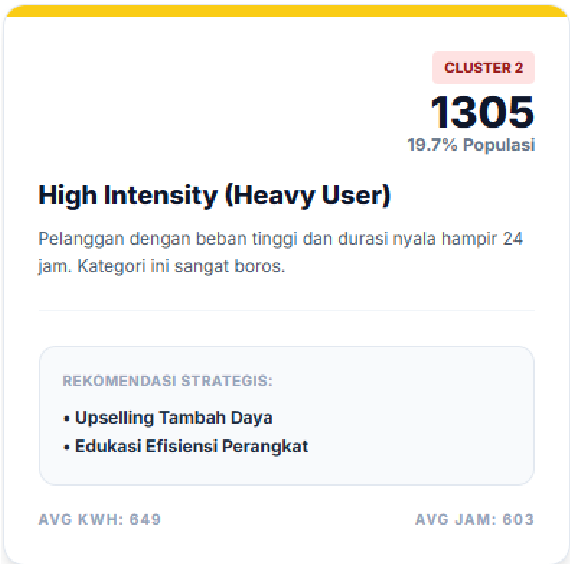

Figure 7. Profile of medium consumption pattern cluster

Cluster 2 is the largest group with 2,927 customers or 44.2% of the total data. This group has an average installed power of 6,800 VA, average operating hours of 128 hours, and average consumption of 1,245 kWh. These characteristics indicate that customers in this cluster have a stable consumption pattern at a medium level.

The dominance of the medium consumption cluster shows that most PLN customers in the South Semarang area are in the fairly active consumption category. This group is important for PLN because its size is large and it has the potential to significantly affect total electricity consumption. Strategies that can be implemented include periodic consumption monitoring, provision of monthly usage information, and loyalty-based service programs or energy efficiency reminders.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1123-1133
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1119

3.6.3 High Consumption Pattern

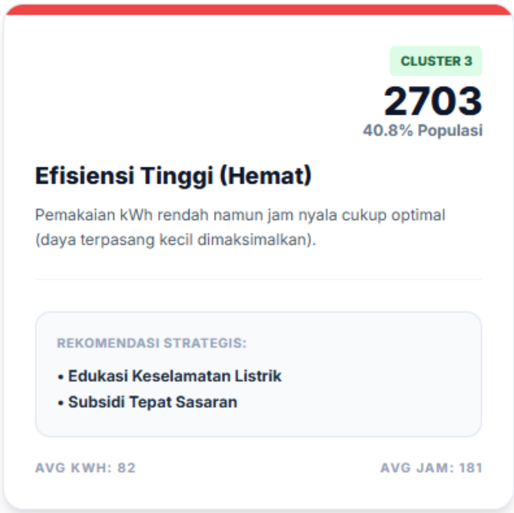

Figure 8. Profile of high consumption pattern cluster

Cluster 3 consists of 1,814 customers or 27.4% of the total data. This group has an average installed power of 18,500 VA, average operating hours of 198 hours, and average consumption of 4,567 kWh. These characteristics indicate that customers in this cluster have high energy needs and the potential to contribute significantly to electricity load.

The high consumption cluster needs to be a main concern in PLN operational efficiency strategies. Customers in this group can be targeted for peak load monitoring, communication related to energy efficiency, and priority services to ensure supply reliability. Because this group has high intensity and volume of usage, small changes in consumption behavior can have a greater impact on system efficiency compared with the low consumption group.

3.7 Implications for Operational Efficiency and Service Strategy

The clustering results provide a more concrete basis for PLN in developing data-driven strategies. Before clustering, customers appear as a heterogeneous dataset. After clustering, customers can be understood as three consumption patterns with different characteristics and needs. This difference is important because operational and service strategies should not be made uniform for all customers.

For the low consumption cluster, strategic focus can be directed toward energy-saving education, bill understanding, and basic information services. For the medium consumption cluster, strategies can be directed toward periodic usage monitoring and strengthening service communication because this group represents the majority of customers. For the high consumption cluster, strategies can be directed toward load monitoring, priority services, and more intensive energy efficiency approaches. Thus, clustering results can help PLN allocate resources more appropriately.

Table 9. Relationship between problems, segmentation results, and operational recommendations

PLN Problem	Consumption Pattern Analysis Result	Recommendation
Customers have heterogeneous consumption characteristics	Three consumption patterns are formed: low, medium, and high	Use service strategies based on consumption segments
Efficiency programs are not yet fully targeted	Low and medium clusters have different usage characteristics	Develop energy efficiency education according to cluster characteristics
Large customer loads need more specific monitoring	The high cluster has the largest average power and consumption	Conduct load monitoring and priority services for high-consumption customers
Service strategies can still be made more personalized	Each cluster has different service needs	Develop service communication based on customer consumption patterns

Table 9 shows that clustering results can be translated into recommendations relevant to PLN problems. In other words, K-Means is not only used to produce cluster labels, but also as a decision-support tool. The analysis results can help PLN design more targeted services and support operational efficiency through an understanding of customer consumption patterns [2], [6].

3.8 Discussion of Research Results

The main finding of this study shows that PLN customer electricity consumption patterns can be clearly grouped into three clusters. The medium cluster is the largest group, indicating that the majority of customers are at a stable and non-extreme consumption level. This condition is important because the medium group can become the main target of service strategies based on routine monitoring. If this largest group is managed properly, its impact on service efficiency and customer communication can be significant.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Although the high consumption cluster is not the largest group, it has the highest average installed power, operating hours, and consumption values. This group is important from an operational perspective because it is related to greater energy needs and potential contribution to peak load. Therefore, the strategy for this cluster should not be limited to general services but should include more proactive approaches such as consumption monitoring, usage pattern notifications, and service coordination when consumption anomalies occur.

The low consumption cluster also has strategic significance. The number of customers in this cluster reaches 28.4%, making it a fairly large group. Service strategies for this group can be directed toward safe electricity usage education, billing information, and prevention of usage discrepancies. Thus, each cluster has a different role in supporting operational efficiency and customer service strategies.

Compared with previous studies that emphasized general K-Means implementation, this study strengthens the interpretation of clustering results in the PLN operational context. Cluster results are not only classified as low, medium, and high, but are also linked to operational recommendations. This approach makes the research findings more applicable because they can be used as a basis for determining service priorities and energy efficiency programs [3][32][9].

However, this study does not measure customer satisfaction directly because it does not use survey, complaint, or service response data. Therefore, the relationship between clustering results and customer satisfaction should be understood as an indirect contribution. This means that understanding consumption patterns can help PLN design more relevant services, and more relevant services may support improved customer satisfaction. To measure the direct effect on satisfaction, future studies need to add customer satisfaction data and service quality indicators [10], [14].

4. CONCLUSION

This study successfully analyzed the electricity consumption patterns of PLN customers in the South Semarang area using the K-Means Clustering method. Based on 6,622 active customer records with installed power, operating hours, and average electricity consumption variables over the last eight months, three optimal clusters were obtained. The number of clusters was determined using the Elbow Method and Silhouette Score, with a WCSS value of 498.67 and a Silhouette Score of 0.68. These results indicate that the grouping has good separation quality and can be used to interpret customer consumption patterns. The three consumption patterns formed are low consumption with 1,881 customers or 28.4%, medium consumption with 2,927 customers or 44.2%, and high consumption with 1,814 customers or 27.4%. The low consumption cluster has the smallest average installed power, operating hours, and consumption values, making it suitable for energy-saving education strategies and basic information services. The medium consumption cluster is the largest group, making it important for periodic monitoring and consistent service communication. The high consumption cluster has the largest average consumption value, so it needs to be prioritized in load monitoring and operational services. The results of this study show that K-Means-based electricity consumption pattern analysis can help PLN understand customer characteristics in a more structured way. This understanding can be used to support operational efficiency, monitoring priority formulation, energy use education, and customer service strategies based on consumption segments. This study has limitations because it only uses data from the South Semarang area and does not include customer satisfaction data, complaints, service quality, or external data such as weather and geographic location. Future studies can develop the model by adding these variables and comparing K-Means with other clustering methods such as K-Medoids, DBSCAN, or hierarchical clustering so that the segmentation results become more comprehensive and robust [33].

REFERENCES

[1] K. Kurniawan, "Governance, Risks, and Compliance in Fulfilling the New and Renewable Energy Mix at the State Electricity Company (PLN)," *J. Ecohumanism*, vol. 3, no. 8, hal. 484–492, 2024, doi: 10.62754/joe.v3i8.4749.

[2] M. P. A. Mailoor, "Evaluating key factors for successful continuous auditing implementation: Insights from an energy and electricity company," *Int. J. Digit. Account. Res.*, vol. 25, hal. 1–33, 2025, doi: 10.4192/1577-8517-v25_1.

[3] S. A. Antikasari, "Indonesia State Electricity Company's Self-Service Technology Adoption: New PLN Mobile," *2023 8th International Conference on Business and Industrial Research Icbir 2023 Proceedings*. hal. 871–876, 2023. doi: 10.1109/ICBIR57571.2023.10147456.

[4] Y. Maraden, "Enhancing Electricity Theft Detection through K-Nearest Neighbors and Logistic Regression Algorithms with Synthetic Minority Oversampling Technique: A Case Study on State Electricity Company (PLN) Customer Data," *Energies*, vol. 16, no. 14, 2023, doi: 10.3390/en16145405.

[5] M. Idhom, "Time Series Regression: Prediction of Electricity Consumption Based on Number of Consumers at National Electricity Supply Company," *TEM J.*, vol. 12, no. 3, hal. 1575–1581, 2023, doi: 10.18421/TEM123-39.

[6] A. P. Taruna, "Electricity Theft Detection Using Machine Learning in Traditional Meter Postpaid Residential Customers: A Case Study on State Electricity Company (PLN) Indonesia," *IEEE Access*, vol. 13, hal. 7167–7191, 2025, doi: 10.1109/ACCESS.2025.3526764.

[7] S. Boukrouh, "Worldwide camel meat and products: An extensive analysis of production, consumption patterns, market evolution, and supply chain effectiveness," *Meat Science*, vol. 228. 2025. doi: 10.1016/j.meatsci.2025.109882.

[8] T. A. Anyasi, "Edible insects as an alternative protein source: Nutritional composition and global consumption patterns," *Future Foods*, vol. 12. 2025. doi: 10.1016/j.fufo.2025.100699.

[9] S. I. O. Herrera, "The role of environmental attitudes and consumption patterns in consumers' preferences for sustainable food from circular farming system: a six EU case studies," *Agric. Food Econ.*, vol. 13, no. 1, 2025, doi: 10.1186/s40100-025-00350-0.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

[10] N. Shahsavari-Pour, “Building electrical consumption patterns forecasting based on a novel hybrid deep learning model,” *Results Eng.*, vol. 26, 2025, doi: 10.1016/j.rineng.2025.104522.
[11] Y. Astuti, “Sentiment Analysis of Electricity Company Service Quality Using Naïve Bayes,” *J. Resti*, vol. 7, no. 2, hal. 389–396, 2023, doi: 10.29207/resti.v7i2.4627.
[12] Y. Dai, “Building-related electric vehicle charging behaviors and energy consumption patterns: An urban-scale analysis,” *Transp. Res. Part D Transp. Environ.*, vol. 141, 2025, doi: 10.1016/j.trd.2025.104663.
[13] H. Lee, “Long-term trends and patterns in ultra-processed food consumption among Korean adults from 1998 to 2022,” *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-88489-0.
[14] Q. Yu, “Modeling electric vehicle behavior: Insights from long-term charging and energy consumption patterns through empirical trajectory data,” *Appl. Energy*, vol. 380, 2025, doi: 10.1016/j.apenergy.2024.125066.
[15] L. C. Chen, “Workload Balancing for Photolithography Machines in Semiconductor Manufacturing via Estimation of Distribution Algorithm Integrating Kmeans Clustering,” *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 55, no. 8, hal. 5565–5580, 2025, doi: 10.1109/TSMC.2025.3572370.
[16] M. A. Cotrina-Teatino, “KMeans-Riemannian model for classification mineral resources in a copper deposit in Peru,” *Int. J. Min. Reclam. Environ.*, 2025, doi: 10.1080/17480930.2025.2518987.
[17] J. Pang, “A New Method of Quality Evaluation for Electromagnet Products Based on SOM-Kmeans and Enhanced TODIM,” *Qual. Reliab. Eng. Int.*, vol. 41, no. 4, hal. 1235–1249, 2025, doi: 10.1002/qre.3725.
[18] Y. Sun, “Enhanced Ant Colony Optimization-KMeans and Reinforcement Learning Framework for Optimizing Tennis Training Programs,” *5th IEEE International Conference on Mobile Networks and Wireless Communications Icmnwc 2025*. 2025. doi: 10.1109/ICMNWC66779.2025.11354417.
[19] S. Palaniappan, S. Karuppannan, dan D. Velusamy, “Categorization of Indian residential consumers electrical energy consumption pattern using clustering and classification techniques,” *Energy*, vol. 289, hal. 129992, 2024, doi: 10.1016/j.energy.2023.129992.
[20] M. S. Talukder, “The role of online information sources in enhancing circular consumption behaviour: Fostering sustainable consumption patterns in the digital age,” *Bus. Strateg. Environ.*, vol. 34, no. 1, hal. 1419–1439, 2025, doi: 10.1002/bse.4053.
[21] L. D. O. Comini, “The Effects of Subsidies for Healthy Foods on Food Purchasing Behaviors, Consumption Patterns, and Obesity/Overweight: A Systematic Review,” *Nutrition Reviews*, vol. 83, no. 7. 2025. doi: 10.1093/nutrit/nuae153.
[22] R. Kwon, “Optimization of State Clustering and Safety Verification in Deep Reinforcement Learning Using KMeans++ and Probabilistic Model Checking,” *IEEE Access*, vol. 13, hal. 28085–28097, 2025, doi: 10.1109/ACCESS.2025.3540428.
[23] K. Elavarasi, “Employee Retention and Layoff Prediction Using Machine Learning: A Comparative Study of XGBoost and KMeans,” *International Conference on Advanced Computing Technologies Icoact 2025*. 2025. doi: 10.1109/ICoACT63339.2025.11005063.
[24] Y. Li, “Empirical analysis of the impact of financial development on the income gap between urban and rural residents in the context of large data using fuzzy Kmeans clustering algorithm,” *Int. J. Electr. Eng. Educ.*, vol. 62, no. 4, hal. 354–369, 2025, doi: 10.1177/0020720920936837.
[25] B. Banyuls, “Convergence in R&D Expenditure in the European Union: A Club Convergence and KMeans Clustering Analysis,” *J. Knowl. Econ.*, vol. 16, no. 3, hal. 13223–13251, 2025, doi: 10.1007/s13132-024-02496-6.
[26] D. K. R. Basani, “An IoMT-Enabled Surgical Monitoring System Utilizing Robotics and AI With E2ARiA-RESNET-50 and MI-KMEANS,” *Trans. Emerg. Telecommun. Technol.*, vol. 36, no. 4, 2025, doi: 10.1002/ett.70082.
[27] B. Qi, “A risk assessment model for the entire rice processing chain based on Kmeans++ and extreme learning machine,” *Lwt*, vol. 223, 2025, doi: 10.1016/j.lwt.2025.117803.
[28] J. Dong, “A Multi-Stage Artificial Intelligence Framework for Building Carbon Emission Prediction Integrating AHP, KMeans, and XGBoost,” *Advances in Transdisciplinary Engineering*, vol. 75. hal. 1925–1934, 2025. doi: 10.3233/ATDE250943.
[29] J. Sun, “Behavior Spectrum-Based Pedestrian Risk Classification via YOLOv8–ByteTrack and CRITIC–Kmeans,” *Appl. Sci. Switz.*, vol. 15, no. 18, 2025, doi: 10.3390/app151810008.
[30] M. K. Zuhanda, “Hybrid Deep Fixed K-Means (HDF-KMeans),” *Int. J. Eng. Sci. Inf. Technol.*, vol. 5, no. 3, hal. 103–111, 2025, doi: 10.52088/ijesty.v5i3.913.
[31] A. A. Fikri, “Sensor Fusion of Laser and Inertial Units with Kalman-KMeans-Fuzzy Framework for Real-Time Railway Geometry Monitoring,” *Bul. Ilm. Sarj. Tek. Elektro*, vol. 7, no. 3, hal. 572–594, 2025, doi: 10.12928/biste.v7i3.13780.
[32] M. Sohaib, “The role of renewable energy in mitigating carbon emissions: Insights from China’s energy consumption patterns,” *Energy Strateg. Rev.*, vol. 61, 2025, doi: 10.1016/j.esr.2025.101860.
[33] A. Riansyah, M. Qomaruddin, M. Indriastuti, dan M. Sagaf, “Clustering Digital Transformation of Small and Medium Enterprises (SMEs) Using Fuzzy K-means Method,” in *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2023, hal. 540–544. doi: 10.1109/EECSI59885.2023.10295664.

This Journal is licensed under a Creative Commons Attribution 4.0 International License