An efficient clustering approach in electrical energy consumption patterns

Dine Tiara Kusuma[1,2], Norashikin Ahmad[2], Sharifah Sakinah Syed Ahmad[2], Iriansyah BM Sangadji[1], Yozika Arvio[1]
[1]Faculty of Energy Telematics, Institut Teknologi PLN, Jakarta, Indonesia
[2]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia

Article Info

Article history:

Received May 25, 2024
Revised Oct 3, 2024
Accepted Oct 17, 2024

Keywords:

Clustering
Davies-Bouldin index
Electrical
Fuzzy subtractive clustering
K-means

ABSTRACT

A comprehensive understanding of electrical energy consumption patterns is essential for strategizing and monitoring the use of energy resources. Industry and business customers of electrical have energy consumption patterns that vary widely depending on the type of industry, business size, and operating hours. This research uses clustering analysis to obtain electrical energy consumption patterns in industrial and business electricity customer groups by grouping data into similar groups. The variables used in this research are daytime, active power (kW), apparent (kVa), and power factor (PF). The objective of this research is to determine the efficacy and benefits of each clustering technique employed in load profile analysis. The clustering algorithm approach used in this research is k-means and fuzzy subtractive clustering (FSC). The trials carried out on these two approaches provide valuable knowledge regarding the effectiveness and superiority of each algorithm in producing significant clusters from the data used in this research. The evaluation conducted using the Davies-Bouldin index (DBI) indicates that the quality value for FSC is 0.25 for business customers and 0.31 for industrial customers. On the other hand, the quality value for k-means is 0.55 for business customers and 0.56 for industrial customers.

Corresponding Author:

Dine Tiara Kusuma
Faculty of Energy Telematics, Institut Teknologi PLN
Jakarta, Indonesia
Email: dinetiarakusuma@gmail.com

1. INTRODUCTION

The use of electrical energy as a medium of human activity is growing because of its very important role in advancing various fields of human civilization, including economic, technological, and socio-cultural. The power plant oversees shouldering the electricity load. In a practical scenario, an electrical load can be defined as a device that relies on electrical energy for its operation. The increasing demand for electrical energy presents a dilemma for electricity supply companies, whose services consist of distributing consumer electricity to ensure the multidisciplinary stability of society as a whole [1]. PT PLN (Persero), as a power supply company, must have a comprehensive understanding of customers' energy consumption patterns and manage electricity supply effectively to meet peak load needs. Peak load analysis allows a power service provider company to strategically plan its infrastructure, organize power plants effectively, and manage distribution networks efficiently to ensure consistent and reliable electricity supply over a given period. Electrical energy plays an important role both in the daily life of people, individuals and in the advancement of the industrial sector. The need for electrical energy is increasing along with the increasing population and human activities. Therefore, having a comprehensive understanding of electrical energy consumption patterns is very important in strategizing and monitoring the effective use of energy resources.

Analysis of electrical energy consumption patterns can be done by clustering. Clustering is a method that can be used to perform an analysis of electrical energy consumption patterns [2], [3]. Clustering is the process of grouping data into groups or clusters that have similar characteristics or patterns [4]. A few articles introduce cluster analysis as a technique for combining data sets [5], [6]. Grouping analysis with data sets can be performed using various categories, algorithms, and evaluation indicators. Industrial and business customers of electrical have widely varying energy consumption patterns depending on industry type, business size, and operating hours. Clustering helps identify groups of consumers with similar consumption patterns, enabling deeper analysis and more informed decision making in energy management. This is very important so that utility companies can improve the efficiency of energy capacity and distribution planning by grouping customer energy consumption based on time. To minimize operational costs, maximize infrastructure use, and prevent network overload.

Various clustering techniques have been used in various studies to cluster electrical load patterns in [7]-[12]. Research by Sharma and Singh [13] introduced a methodology to analyze electricity consumption patterns and evaluate load factors on various days within a given geographic area. This analysis was performed using the k-means clustering algorithm. Research on the analysis of electrical energy consumption patterns was also conducted by Alam and Ali [10], who used radius values to find clusters that are close to each other on household electrical energy consumption. The analysis was performed with the fuzzy subtractive clustering (FSC) algorithm.

Based on the research that has been done, k-means and FSC are two of the clustering methods that can be used for peak load laterization in electricity customers [14]. K-means as one of the supervised clustering methods certainly has different cluster results from the clustering results in the FSC method which is an unsupervised clustering method. Comparison with these two methods can provide insight into the effectiveness and advantages of each algorithm in generating meaningful clusters from peak load time data. In addition, the performance of these two algorithms can be known by evaluating in terms of computational time and algorithm complexity. The main contributions of the study are the findings of this study offer practical suggestions regarding the most efficient clustering technique to employ while measuring electrical energy consumption and the results of this research provide valuable insights for the development of smarter, data-based energy management strategies, which can be widely applied to improve sustainability and effectiveness of energy use.

2. RELATED WORK

In addition to the studies conducted by Alam and Ali [10] and Sharma and Singh [13]. The latest study conducted by Wang *et al.* [11] which researched the analysis of user's power consumption behavior based on k-means resulting in a classification of users into four different types based on power consumption behavior based on the number of clusters from k-means. Similar research on electrical energy consumption has also been conducted by Zhang and Mo [12] researched about clustering analysis of electricity consumption behavior through various traditional clustering algorithms, such as k-means and fuzzy c-means, which results in electricity users can be classified into different categories based on their electricity consumption behavior. Oriona *et al.* [15] introduced two novel distances between ordinal time series and used them to construct fuzzy clustering procedures. Phamtoan *et al*. [16] presented a fuzzy behavior clustering approach that combines fuzzy techniques with a well-known clustering algorithm. These papers demonstrate the effectiveness of FSC for time series analysis. Based on the characteristics of the data in this research, customer groupings do not describe the use of electrical energy at a certain time, while trends in energy consumption patterns describe the characteristics of data based on time series, so that in this research FSC is one of the alternatives that will be used with the k-means algorithm for clustering of electrical energy consumption data based on consumption time.

3. METHOD

Section 3 explains the research method used in this study, which includes the stages of data collection, data preprocessing, data analysis using the k-means and FSC clustering approaches, as well as the evaluation results, the research method is shown in Figure 1. Based on Figure 1, to obtain electrical energy consumption patterns with the k-means and FSC algorithms, it is carried out in 4 stages, namely:

3.1. Data collection

The electrical energy consumption data obtained for the purposes of this study is electrical energy consumption data for the low voltage business and industry class in 2020 collected from the automatic meter reading (AMR) system which is recorded every 15 minutes, for 7 days with the attributes contained therein are time (day, date, hour, minutes, and second), voltage (volt), current (ampere), active power (kW), apparent (kVa), reactive power (Var), frequency (Hz), and power factor (PF). Data obtained from AMR does not require any additional instruments [17], so, in this study the data obtained can be forwarded directly into the data preprocessing process before the clustering process with k-means and FSC. To classify the time of electrical energy consumption in this study, the attributes used are time, active power (kW), and apparent (kVa), because these factors greatly affect the efficiency of the electrical system which is closely related to PF. PF is generated from a comparison between active power and apparent power. The PF value is limited between 0 and 1. A low PF (close to 0) can result in higher reactive power usage resulting in reduced power utilization which can increase electricity costs [18].

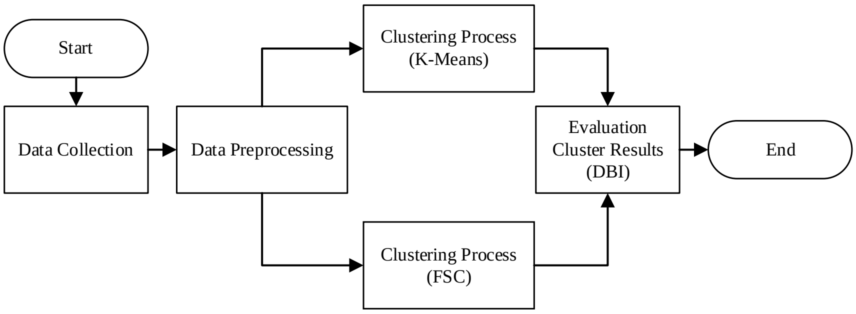


Figure 1. Research method

3.2. Data preprocessing

The preprocessing of this study's data was carried out in 3 stages, namely:

a. Data cleaning is done by cleaning raw data that has been downloaded from AMR. This data cleaning process is carried out by filling in the missing data parts caused by blackouts during data recording, unstable telecommunications networks when sending data from electricity meters to the AMR system by calculating the replacement value (imputation) using the average value function in the missing attribute.
b. Data transformation, done by normalizing data is done by changing data into a regular scale. The normalization carried out in this study is to use the min–max scale with (1) [7], [19], [20]:

$$Xnorm = \frac{X - Xmin}{Xmax - Xmin} \quad (1)$$

Where, Xnormis the value of the data attribute searched; X is the old value of the attribute in the data; Xmin is the minimum absolute value of all data; and Xmax is maximum absolute value of all data.
c. Data discretization, done by sorting data into smaller intervals, the interval of recording discredited data in this study is every 1 hour, so that the number of data records used is as many as 168 data records.

3.3. K-means clustering process

The k-means clustering algorithm is a supervised clustering method that can be used to partition data into K clusters. The sequence of stages in the k-means clustering algorithm is as follows [2]:

a. K objects are randomly selected to serve as the center of the initial cluster, referred to as centroids. K objects selected in this study are K=3, which means that the data will be clustered into 3-time groups, namely time groups - times of high, medium, and low electrical energy use.
b. Each object is allocated to the center of the cluster closest to it by the Euclidian distance formulation which is determined by calculating the distance between each object and each centroid. The following is the Euclidian distance formulation used in this study to find the closest distance value between data and the centroid [21]:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (xi - yi)^2} \quad (2)$$

where, *d* is distance between x and y; *x* is cluster data center; *y* is data on the attribute; *i* is any data; *n* is amount of data; *Xi* is data at cluster center to *I*; and *Yi* is data on each data to *i*. The center of the cluster and the objects assigned to it collectively represent a class. This procedure will be repeated until the necessary conditions for termination are met.

3.4. Fuzzy subtractive clustering process

The FSC algorithm is an unsupervised clustering method that can be used to partition data based on proximity values between data (radius) [22]. The sequence of stages in the FSC algorithm is as follows:

a. The subtractive clustering (SC) algorithm determines fuzzy clusters by assigning a potential value of pi, for each data point based on its likelihood according to Gaussian size:

$$pi = \sum_j e^{-a||xi-xj||^2} \tag{3}$$

where, a=4/ra^2 and *ra* represents the cluster radius.

b. The data point with maximum potential is selected as the initial cluster center, and all data points within a radius distance of the cluster center are connected to the first cluster. The center of the second cluster is determined using a similar method, but this time the data points associated with the first cluster are excluded. This process is repeated until all data points are within the central radius of the cluster.

3.5. Evaluation

The Davies-Bouldin index (DBI) is used to assess clustering effectiveness by measuring and analyzing data set features. Maximizing the distance between two clusters and minimizing the distance between members of one cluster is the goal of this approach. To calculate the DBI value, (4) and (5) can be used [9], [23]:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j = 1, i \neq 1} Rij \tag{4}$$

$$Ri = \frac{var\,(Ci)+var(Cj)}{|Ci-Cj|} \tag{5}$$

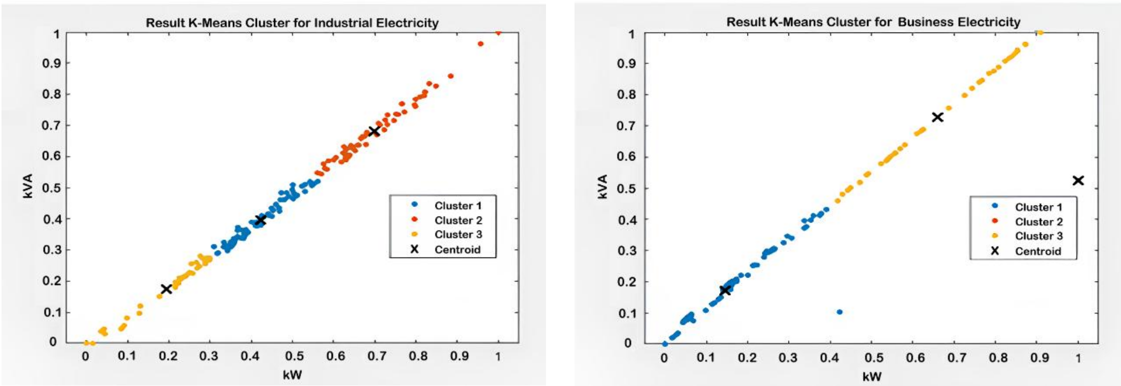
where *Ci*, *Cj*: central cluster I, and central cluster. The smaller the DBI, the better the clustering quality [24], [25].

4. RESULTS AND DISCUSSION

Clustering is carried out using 168 data records, with the variables used being daytime, active power (kW), and apparent (kVa) on industrial and business electricity customer data. This data is obtained from recording electrical energy consumption for 24 hours in 7 days. As the results shown in Figures 2 and 3, the use of kW and kVA variables shows the total capacity required of the electrical system to operate the load and provides an overview of the energy consumption used by the electrical load to operate. The following are the results of clustering with the k-means and FSC algorithms.

4.1. K-means

The k-means algorithm in this study was used to form three clusters, namely the time of use of high, medium, and low electrical energy. Figure 2 displays the clustering results using the k-means algorithm Figure 2(a) for industry and Figure 2(b) business customers electricity. Table 1 and (Table 2 see in Appendix) display result of the number of clusters formed from industry and business data, along with the members in each cluster using the k-means algorithm.



(a) (b)

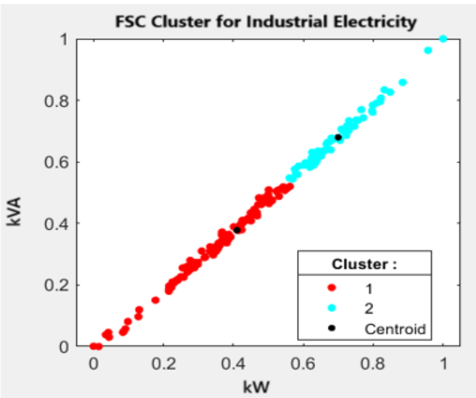
Figure 2. K-means cluster for; (a) industrial and (b) business electricity

Table 1. Cluster result in industrial electricity customer group using k-means

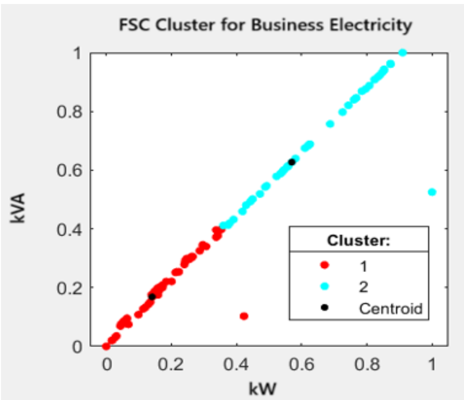
Number of clusters	DayTime	kW	kVa	*PF*	Member of cluster
37	Saturday, 07:00	0.17	0.19	0.89	Wednesday 00:00, Wednesday 01:00, Wednesday 02:00, Wednesday 03:00, Wednesday 04:00, Wednesday 05:00, Wednesday 10:00, Wednesday 13:00, Wednesday 18:00, Wednesday 19:00, Wednesday 20:00, Wednesday 21:00, Wednesday 22:00, Wednesday 23:00, Thursday 00:00, Thursday 01:00, Thursday 02:00, Thursday 03:00, Thursday 04:00, Thursday 05:00, Thursday 08:00, Thursday 09:00, Thursday 11:00, Thursday 19:00, Thursday 22:00, Thursday 23:00, Friday 00:00, Friday 02:00, Friday 02:00, Friday 04:00, Friday 05:00, Friday 06:00, Friday 08:00, Friday 09:00, Friday 16:00, Saturday 01:00, Saturday 09:00, Saturday 10:00, Saturday 11:00, Saturday 13:00, Saturday 14:00, Saturday 15:00, Saturday 16:00, Saturday 18:00, Saturday 19:00, Saturday 22:00, Saturday 23:00, Sunday 01:00, Sunday 02:00, Sunday 03:00, Sunday 06:00, Sunday 07:00, Sunday 11:00, Sunday 14:00,,Sunday 19:00, Sunday 21:00, Monday 00:00, Monday 01:00, Monday 02:00, Monday 03:00, Monday 04:00, Monday 10:00, Monday 11:00, Monday 13:00, Monday 19:00, Monday 20:00, Monday 21:00, Tuesday 00:00, Tuesday 01:00, Tuesday 02:00, Tuesday 03:00, Tuesday 04:00, Tuesday 05:00, Tuesday 08:00, Tuesday 14:00,Tuesday 18:00.
56	Thursday, 06:00	0.68	0.69	0.97	Wednesday 06:00, Wednesday 07:00, Wednesday 09:00, Wednesday 11:00, Wednesday 12:00, Wednesday 14:00, Wednesday 15:00, Wednesday 16:00, Wednesday 17:00, Thursday 06:00, Thursday 07:00, Thursday 10:00, Thursday 12:00, Thursday 13:00, Thursday 14:00, Thursday 15:00, Thursday 16:00, Thursday 17:00, Thursday 18:00, Thursday 20:00, Thursday 21:00, Friday 01:00, Friday 17:00, Friday 18:00, Friday 19:00, Friday 20:00, Friday 21:00, Friday 22:00, Friday 23:00, Saturday 00:00, Saturday 08:00, Saturday 12:00, Saturday 17:00, Saturday 20:00, Saturday 21:00, Sunday 00:00, Sunday 04:00, Sunday 05:00, Sunday 22:00, Sunday 23:00, Monday 06:00, Monday 07:00, Monday 14:00, Monday 15:00, Monday 16:00, Monday 17:00, Monday 18:00, Monday 22:00, Monday 23:00, Tuesday 06:00, Tuesday 07:00 Tuesday 19:00, Tuesday 20:00, Tuesday 21:00, Tuesday 22:00, Tuesday 23:00.
75	Monday, 01:00	0.39	0.42	0.93	Wednesday 08:00, Friday 03:00, Friday 07:00, Friday 10:00, Friday 11:00, Friday 12:00, Friday 13:00, Friday 14:00, Friday 15:00, Saturday 02:00, Saturday 03:00, Saturday 04:00, Saturday 05:00, Saturday 06:00, Saturday 07:00, Sunday 08:00, Sunday 09:00, Sunday 10:00, Sunday 12:00, Sunday 13:00, Sunday 15:00, Sunday 16:00, Sunday 17:00, Sunday 18:00, Sunday 20:00, Monday 05:00, Monday 08:00, Monday 09:00, Monday 12:00, Tuesday 09:00, Tuesday 10:00, Tuesday 11:00, Tuesday 12:00, Tuesday 13:00, Tuesday 15:00, Tuesday 16:00, Tuesday 17:00.

4.2. Fuzzy subtractive clustering

The FSC algorithm in this study used cluster radius values set using a value of 0.42 [26], with a squash factor of 1.5 [19], [27], [28], accept ratio of 0.5 and reject ratio of 0.15 [29]. Figure 3 displays the results clustering with the FSC algorithm for Figure 3(a) industry and Figure 3(b) business customers electricity. Table 3 and (Table 4 in Appendix) display result of the number of clusters formed from industry and business data, along with the members in each cluster using the FSC algorithm:



(a)



(b)

Figure 3. FSC cluster for; (a) industrial and (b) business electricity

Table 3. Cluster result in industrial electricity customer group using FSC

Number of clusters	DayTime	kW	kVa	*PF*	Member of cluster
112	Tuesday 01:00	0.41	0.37	0.92	Wednesday 00:00, Wednesday 01:00, Wednesday 02:00, Wednesday 03:00, Wednesday 04:00, Wednesday 05:00, Wednesday 08:00, Wednesday 10:00, Wednesday 13:00, Wednesday 18:00, Wednesday 19:00, Wednesday 20:00, Wednesday 21:00, Wednesday 22:00, Wednesday 23:00, Thursday 00:00, Thursday 01:00, Thursday 02:00, Thursday 03:00, Thursday 04:00, Thursday 05:00, Thursday 08:00, Thursday 09:00, Thursday 11:00, Thursday 19:00, Thursday 22:00, Thursday 23:00, Friday 00:00, Friday 02:00, Friday 03:00, Friday 04:00, Friday 05:00, Friday 06:00, Friday 07:00, Friday 08:00, Friday 09:00, Friday 10:00, Friday 11:00, Friday 12:00, Friday 13:00, Friday 14:00, Friday 15:00, Friday 16:00, Saturday 01:00, Saturday 02:00, Saturday 03:00, Saturday 04:00, Saturday 05:00, Saturday 06:00, Saturday 07:00, Saturday 09:00, Saturday 10:00, Saturday 11:00, Saturday 13:00, Saturday 14:00, Saturday 15:00, Saturday 16:00, Saturday 18:00, Saturday 19:00, Saturday 22:00, Saturday 23:00, Sunday 01:00, Sunday 02:00, Sunday 03:00, Sunday 06:00, Sunday 07:00, Sunday 08:00, Sunday 09:00, Sunday 10:00, Sunday 11:00, Sunday 12:00, Sunday 13:00, Sunday 14:00, Sunday 15:00, Sunday 16:00, Sunday 17:00, Sunday 18:00, Sunday 19:00, Sunday 20:00, Sunday 21:00, Monday 00:00, Monday 01:00, Monday 02:00, Monday 03:00, Monday 04:00, Monday 05:00, Monday 08:00, Monday 09:00, Monday 10:00, Monday 11:00, Monday 12:00, Monday 13:00, Monday 19:00, Monday 20:00, Monday 21:00, Tuesday 00:00, Tuesday 01:00, Tuesday 02:00, Tuesday 03:00, Tuesday 04:00, Tuesday 05:00, Tuesday 08:00, Tuesday 09:00, Tuesday 10:00, Tuesday 11:00, Tuesday 12:00, Tuesday 13:00, Tuesday 14:00, Tuesday 15:00, Tuesday 16:00, Tuesday 17:00, Tuesday 18:00.
56	Thursday 06:00	0.6999	0.6796	0.97	Wednesday 06:00, Wednesday 07:00, Wednesday 09:00, Wednesday 11:00, Wednesday 12:00, Wednesday 14:00, Wednesday 15:00, Wednesday 16:00, Wednesday 17:00, Thursday 06:00, Thursday 07:00, Thursday 10:00, Thursday 12:00, Thursday 13:00, Thursday 14:00, Thursday 15:00, Thursday 16:00, Thursday 17:00, Thursday 18:00, Thursday 20:00, Thursday 21:00, Friday 01:00, Friday 17:00, Friday 18:00, Friday 19:00, Friday 20:00 Friday 21:00, Friday 22:00, Friday 23:00, Saturday 00:00 Saturday 08:00, Saturday 12:00, Saturday 17:00, Saturday 20:00 Saturday 21:00, Sunday 00:00, Sunday 04:00, Sunday 05:00, Sunday 22:00, Sunday 23:00, Monday 06:00, Monday 07:00, Monday 14:00, Monday 15:00, Monday 16:00, Monday 17:00, Monday 18:00, Monday 22:00, Monday 23:00, Tuesday 06:00, Tuesday 07:00, Tuesday 19:00, Tuesday 20:00, Tuesday 21:00, Tuesday 22:00, Tuesday 23:00.

4.3. Evaluation

Evaluation results obtained from tests conducted with the k-means algorithm and FSC are presented using the DBI in Table 5. Based on the study results, we found that FSC can generate superior clusters. The value of the evaluation results obtained through DBI on the FSC algorithm is comparatively lower than that of the alternative methods.

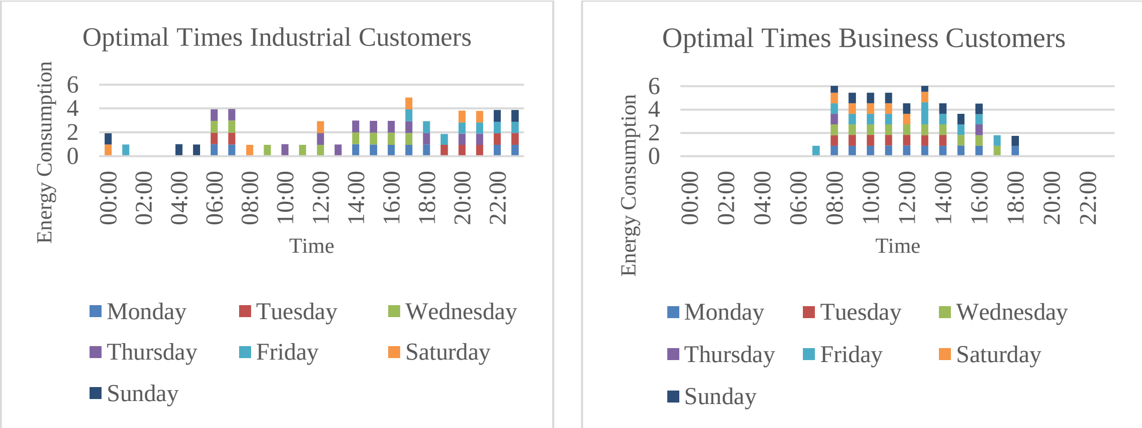
Table 5. Evaluation result

DBI	K-means	FSC
Industry	0.5667	0.31
Business	0.550	0.25

5. DISCUSSION

This discussion presents a graph of electrical energy consumption patterns in industrial and business customers based on the results of the FSC cluster. In each group of industrial and business customers, the best centroid PF is found in the 2nd cluster, namely 0.97 for industrial customer groups and 0.91 for business customer groups. When the PF approaches 1, it indicates that the electrical system is experiencing little or no energy waste and is operating efficiently. Based on the results of clustering on the times of electrical energy use from industrial and business customers, the optimal times for 7 days are shown in Figures 4(a) and (b).

The trend of electrical energy consumption in Figures 4(a) and (b) occurs evenly in the industrial electricity customer group, and in the business customer group the trend of electrical energy consumption for 7 days looks optimal at 07:00 to 18:00. This time group can be used by power supply companies to plan and manage future electricity infrastructure. Another finding obtained from this research is that in the Business group electricity customers found an anomaly on the centroid generated with the k-means algorithm, it shows in Table 2 the results of the 2nd cluster the PF value on the centroid exceeds the normal threshold limit value, this can happen because of several possibilities such as imbalance disorders when measuring, the use of improper measuring instruments or calibration errors, and the presence of high harmonic conditions.



(a) (b)

Figure 4. Optimal times in industrial; (a) business and (b) customers

6. CONCLUSION

The following are some of the findings that emerged from this research: the results of clustering on seven-day, twenty-four-hour electrical energy consumption data using apparent (kVa), PF, and active power (kW) show that FSC produces better clusters; as indicated by DBI evaluation values of 0.25 for business customer groups and 0.31 for industrial customer groups. Additionally, the obtained cluster results indicate that optimal electrical energy consumption is typically more consistent and uniformly always distributed for industrial customers. This contrasts with business customers, who typically maximize their electrical energy consumption between 07:00 and 18:00 daily. Further research can be considered to investigate the application of the clustering method on a larger scale, such as in smart grid networks and smart technology that can assist in real-time load management.

APPENDIX

Table 2. Cluster result in business electricity customer group using k-means

Number of clusters	DayTime	Watt	Va	*PF*	Member of cluster
127	Tuesday 21:00	0.14	0.17	0.85	Monday 00:00, Monday 01:00, Monday 02:00, Monday 03:00, Monday 04:00, Monday 05:00, Monday 06:00, Monday 07:00, Monday 16:00, Monday 17:00, Monday 18:00, Monday 19:00, Monday 20:00, Monday 21:00, Monday 22:00, Monday 23:00, Tuesday 00:00, Tuesday 01:00, Tuesday 02:00, Tuesday 03:00, Tuesday 04:00, Tuesday 05:00, Tuesday 06:00, Tuesday 07:00, Tuesday 15:00, Tuesday 16:00, Tuesday 17:00, Tuesday 18:00, Tuesday 19:00, Tuesday 20:00, Tuesday 21:00, Tuesday 22:00, Tuesday 23:00, Wednesday 00:00, Wednesday 01:00, Wednesday 02:00, Wednesday 03:00, Wednesday 04:00, Wednesday 05:00, Wednesday 06:00, Wednesday 07:00, Wednesday 16:00, Wednesday 17:00, Wednesday 18:00, Wednesday 19:00, Wednesday 20:00, Wednesday 21:00, Wednesday 22:00, Wednesday 23:00, Thursday 00:00, Thursday 01:00, Thursday 02:00, Thursday 03:00, Thursday 04:00, Thursday 05:00, Thursday 06:00, Thursday 06:00, Thursday 07:00, Thursday 09:00, Thursday 10:00, Thursday 11:00, Thursday 12:00, Thursday 13:00, Thursday 14:00, Thursday 15:00, Thursday 17:00, Thursday 18:00, Thursday 19:00, Thursday 20:00, Thursday 21:00, Thursday 22:00, Thursday 23:00, Friday 00:00, Friday 01:00, Friday 02:00, Friday 03:00, Friday 04:00, Friday 05:00, Friday 06:00, Friday 12:00, Friday 18:00; Friday 19:00, Friday 20:00, Friday 21:00, Friday 22:00, Friday 23:00, Saturday 00:00, Saturday 01:00, Saturday 02:00, Saturday 03:00, Saturday 04:00, Saturday 05:00, Saturday 06:00, Saturday 07:00, Saturday 14:00, Saturday 15:00, Saturday 16:00, Saturday 17:00, Saturday 18:00, Saturday 19:00, Saturday 20:00, Saturday 21:00, Saturday 22:00, Saturday 23:00, Sunday 00:00, Sunday 01:00, Sunday 02:00, Sunday 03:00, Sunday 04:00, Sunday 05:00,

Table 2. Cluster result in business electricity customer group using k-means (*continued*)

Number of clusters	DayTime	Watt	Va	*PF*	Member of cluster
127	Tuesday 21:00	0.14	0.17	0.85	Sunday 06:00, Sunday 07:00, Sunday 08:00, Sunday 09:00, Sunday 10:00, Sunday 11:00, Sunday 12:00, Sunday 13:00, Sunday 14:00, Sunday 15:00, Sunday 16:00, Sunday 17:00, Sunday 18:00, Sunday 19:00, Sunday 20:00, Sunday 21:00, Sunday 22:00, Sunday 23:00.
1	Thursday 16:00	1	0.52	1.9	Friday 13:00.
40	Friday 13:00	0.66	0.72	0.90	Monday 08:00, Monday 09:00, Monday 10:00, Monday 11:00, Monday 12:00, Monday 13:00, Monday 14:00, Monday 15:00, Tuesday 08:00, Tuesday 09:00, Tuesday 10:00, Tuesday 11:00, Tuesday 12:00, Tuesday 13:00, Tuesday 14:00, Wednesday 08:00, Wednesday 09:00, Wednesday 10:00, Wednesday 11:00, Wednesday 12:00, Wednesday 13:00, Wednesday 14:00, Wednesday 15:00, Thursday 08:00, Thursday 16:00, Friday 07:00, Friday 08:00, Friday 09:00, Friday 10:00, Friday 11:00, Friday 14:00, Friday 15:00, Friday 16:00, Friday 17:00, Saturday 08:00, Saturday 09:00, Saturday 10:00, Saturday 11:00, Saturday 12:00, Saturday 13:00.

Table 4. Cluster result in business electricity customer group using FSC

Number of clusters	DayTime	kW	kVa	*PF*	Member of cluster
123	Saturday 22:00	0.14	0.16	0.84	Monday 00:00, Monday 01:00, Monday 02:00, Monday 03:00, Monday 04:00, Monday 05:00, Monday 06:00, Monday 07:00, Monday 17:00, Monday 19:00, Monday 20:00, Monday 21:00, Monday 22:00, Monday 23:00, Tuesday 00:00, Tuesday 01:00, Tuesday 02:00, Tuesday 03:00, Tuesday 04:00, Tuesday 05:00, Tuesday 06:00, Tuesday 07:00, Tuesday 15:00, Tuesday 16:00, Tuesday 17:00, Tuesday 18:00, Tuesday 19:00, Tuesday 20:00, Tuesday 21:00, Tuesday 22:00, Tuesday 23:00, Wednesday 00:00, Wednesday 01:00, Wednesday 02:00, Wednesday 03:00, Wednesday 04:00, Wednesday 05:00, Wednesday 06:00, Wednesday 07:00, Wednesday 18:00, Wednesday 19:00, Wednesday 20:00, Wednesday 21:00, Wednesday 22:00, Wednesday 23:00, Thursday 00:00, Thursday 01:00, Thursday 02:00, Thursday 03:00, Thursday 04:00, Thursday 05:00, Thursday 06:00, Thursday 07:00, Thursday 09:00, Thursday 10:00, Thursday 11:00, Thursday Thursday 12:00, Thursday 13:00, Thursday 14:00, Thursday 15:00, Thursday 17:00, Thursday 18:00, Thursday 19:00, Thursday 20:00, Thursday 21:00, Thursday 22:00, Thursday 23:00, Friday 00:00, Friday 01:00, Friday 02:00, Friday 03:00, Friday 04:00, Friday 05:00, Friday 06:00, Friday 12:00, Friday 18:00, Friday 19:00, Friday 20:00, Friday 21:00, Friday 22:00, Friday 23:00, Saturday 00:00, Saturday 01:00, Saturday 02:00, Saturday 03:00, Saturday 04:00, Saturday 05:00, Saturday 06:00, Saturday 07:00, Saturday 14:00, Saturday 15:00, Saturday 16:00, Saturday 17:00, Saturday 18:00, Saturday 19:00, Saturday 20:00, Saturday 21:00,Saturday 22:00, Saturday 23:00, Sunday 00:00, Sunday 01:00, Sunday 02:00, Sunday 03:00, Sunday 04:00, Sunday 05:00, Sunday 06:00, Sunday 07:00, Sunday 08:00, Sunday 09:00, Sunday 10:00, Sunday 11:00, Sunday 12:00, Sunday 12:00, Sunday 13:00, Sunday 14:00, Sunday 15:00, Sunday 16:00, Sunday 17:00, Sunday 18:00, Sunday 19:00, Sunday 20:00, Sunday 21:00, Sunday 22:00, Sunday 23:00.
45	Wednesday 14:00	0.56	0.62	0.91	Monday 08:00, Monday 09:00, Monday 10:00, Monday 11:00 Monday 12:00, Monday 13:00, Monday 14:00, Monday 15:00 Monday 16:00, Monday 18:00, Tuesday 08:00, Tuesday 09:00 Tuesday 10:00, Tuesday 11:00, Tuesday 12:00, Tuesday 13:00, Tuesday 14:00, Wednesday 08:00, Wednesday 09:00, Wednesday 10:00, Wednesday 11:00, Wednesday 12:00, Wednesday 13:00, Wednesday 14:00, Wednesday 15:00, Wednesday 16:00, Wednesday 17:00, Thursday 08:00, Thursday 16:00, Friday 07:00, Friday 08:00, Friday 09:00, Friday 10:00, Friday 11:00, Friday 13:00, Friday 14:00, Friday 15:00, Friday 16:00, Friday 17:00, Saturday 08:00, Saturday 09:00, Saturday 10:00, Saturday 11:00, Saturday 12:00, Saturday 13:00.

ACKNOWLEDGEMENTS

This research was supported by Institut Teknologi PLN, based on Institutional flagship research projects No. 0209.SK/1/A0/07/2023 and Research Agreement No. 001/PU/A04/VIII/2023.

REFERENCES

[1] E. Proedrou, "A Comprehensive Review of Residential Electricity Load Profile Models," *IEEE Access*, vol. 9, pp. 12114–12133, 2021, doi: 10.1109/ACCESS.2021.3050074.
[2] Y. Zhao *et al.*, "An Evaluation Strategy of Energy Storage Construction for Industrial Users Based on K-Means Clustering

Algorithm," in *2019 Computing, Communications and IoT Applications, ComComAp 2019*, IEEE, Oct. 2019, pp. 396–401, doi: 10.1109/ComComAp46287.2019.9018661.
[3] G. Grigoras, M. Istrate, and F. Scarlatache, "Electrical energy consumption estimation in water distribution systems using a clustering based method," in *2013 International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2013*, IEEE, Jun. 2013, pp. 1–6, doi: 10.1109/ECAI.2013.6636174.
[4] H. B. Tambunan, D. H. Barus, J. Hartono, A. S. Alam, D. A. Nugraha, and H. H. H. Usman, "Electrical peak load clustering analysis using K-means algorithm and silhouette coefficient," in *Proceeding - 2nd International Conference on Technology and Policy in Electric Power and Energy, ICT-PEP 2020*, IEEE, Sep. 2020, pp. 258–262, doi: 10.1109/ICT-PEP50916.2020.9249773.
[5] D. Xu and Y. Tian, "A Comprehensive Survey of Clustering Algorithms," *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, Jun. 2015, doi: 10.1007/s40745-015-0040-1.
[6] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013, doi: 10.1109/TPAMI.2013.57.
[7] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2437–2447, Sep. 2016, doi: 10.1109/TSG.2016.2548565.
[8] W. Labeeuw and G. Deconinck, "Residential electrical load model based on mixture model clustering and markov models," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1561–1569, Aug. 2013, doi: 10.1109/TII.2013.2240309.
[9] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68–80, Jun. 2012, doi: 10.1016/j.energy.2011.12.031.
[10] S. M. M. Alam and M. H. Ali, "A new subtractive clustering based ANFIS system for residential load forecasting," in *2020 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2020*, IEEE, Feb. 2020, pp. 1–5, doi: 10.1109/ISGT45199.2020.9087653.
[11] X. Wang, H. Li, X. Yi, J. Kong, and X. Wang, "Analysis of User's Power Consumption Behavior Based on K-means," in *Proceedings - 2022 4th International Conference on Machine Learning, Big Data and Business Intelligence, MLBDBI 2022*, IEEE, Oct. 2022, pp. 39–42, doi: 10.1109/MLBDBI58171.2022.00013.
[12] C. Zhang and J. Mo, "A Clustering Analysis of Electricity Consumption Behavior," in *2022 5th International Conference on Energy, Electrical and Power Engineering, CEEPE 2022*, IEEE, Apr. 2022, pp. 962–969, doi: 10.1109/CEEPE55110.2022.9783311.
[13] D. D. Sharma and S. N. Singh, "Electrical load profile analysis and peak load assessment using clustering technique," in *IEEE Power and Energy Society General Meeting*, IEEE, Jul. 2014, pp. 1–5, doi: 10.1109/PESGM.2014.6938869.
[14] D. T. Kusuma, N. Ahmad, and S. S. S. Ahmad, "Clustering Algorithm for Electrical Load Profiling Analysis: a Systematic Review of Machine Learning Approaches for Improved Clustering Algorithms," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 10, pp. 5453–5468, 2024.
[15] Á. L. Oriona, C. Weiss, and J. A. Vilar, "Fuzzy clustering of ordinal time series based on two novel distances with economic applications," *arXiv*, 2023, doi: 10.48550/arXiv.2304.12249.
[16] D. Phamtoan and T. Vovan, "Building fuzzy time series model from unsupervised learning technique and genetic algorithm," *Neural Computing and Applications*, vol. 35, no. 10, pp. 7235–7252, Apr. 2023, doi: 10.1007/s00521-021-06485-7.
[17] K. Y. Khumchoo and W. Kongprawechnon, "Cluster analysis for primary feeder identification using metering data," in *2015 6th International Conference on Information and Communication Technology for Embedded Systems, IC-ICTES 2015*, IEEE, Mar. 2015, pp. 1–6, doi: 10.1109/ICTEmSys.2015.7110831.
[18] O. Austin, "Electrical Energy Losses Reduction using Power Factor Correction Technique," *Advances in Electrical and Telecommunication Engineering (AETE)*, vol. 1, no. 1, pp. 9–16, 2018.
[19] P. Bagheripour and M. Asoodeh, "Fuzzy ruling between core porosity and petrophysical logs: Subtractive clustering vs. genetic algorithm-pattern search," *Journal of Applied Geophysics*, vol. 99, pp. 35–41, Dec. 2013, doi: 10.1016/j.jappgeo.2013.09.014.
[20] I. Sangadji, Y. Arvio, and Indrianto, "Dynamic Segmentation of Behavior Patterns Based On Quantity Value Movement Using Fuzzy Subtractive Clustering Method," *Journal of Physics: Conference Series*, vol. 974, no. 1, pp. 1–8, Mar. 2018, doi: 10.1088/1742-6596/974/1/012009.
[21] Haviluddin *et al.*, "A Performance Comparison of Euclidean, Manhattan and Minkowski Distances in K-Means Clustering," in *2020 6th International Conference on Science in Information Technology: Embracing Industry 4.0: Towards Innovation in Disaster Management, ICSITech 2020*, IEEE, Oct. 2020, pp. 184–188, doi: 10.1109/ICSITech49800.2020.9392053.
[22] S. K. Jha and Z. Ahmad, "Soil microbial dynamics modeling in fluctuating ecological situations by using subtractive clustering and fuzzy rule-based inference systems," *CMES - Computer Modeling in Engineering and Sciences*, vol. 113, no. 4, pp. 443–459, 2017, doi: 10.3970/cmes.2017.113.443.
[23] F. Nuraeni, N. N. Febriani SM, L. Listiani, and E. Rahmawati, "Implementation of K-Means Algorithm with Distance of Euclidean Proximity in Clustering Cases of Violence Against Women and Children," in *2019 1st International Conference on Cybernetics and Intelligent System, ICORIS 2019*, IEEE, Aug. 2019, pp. 162–167, doi: 10.1109/ICORIS.2019.8874883.
[24] Y. Shi, T. Yu, Q. Liu, H. Zhu, F. Li, and Y. Wu, "An Approach of Electrical Load Profile Analysis Based on Time Series Data Mining," *IEEE Access*, vol. 8, pp. 209915–209925, 2020, doi: 10.1109/ACCESS.2020.3019698.
[25] K. Chowdhury, D. Chaudhuri, A. K. Pal, and A. Samal, "Seed selection algorithm through K-means on optimal number of clusters," *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 18617–18651, Jul. 2019, doi: 10.1007/s11042-018-7100-4.
[26] R. Dhanalakshmi and R. Vinodha, "Performance evaluation of fuzzy subtractive clustering scheduler in real-time spherical tank process," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 4, pp. 1640–1645, Apr. 2019, doi: 10.1166/jctn.2019.8090.
[27] S. Barchinezhad, M. Eftekhari, and H. Sanatnama, "A new feature ranking criterion based on density function of subtractive clustering," in *13th Iranian Conference on Fuzzy Systems, IFSC 2013*, IEEE, Aug. 2013, pp. 1–4, doi: 10.1109/IFSC.2013.6675624.
[28] H. Hosseini, B. Tousi, and N. Razmjooy, "Application of fuzzy subtractive clustering for optimal transient performance of automatic generation control in restructured power system," *Journal of Intelligent and Fuzzy Systems*, vol. 26, no. 3, pp. 1155–1166, 2014, doi: 10.3233/IFS-130802.
[29] L. T. Ngo and B. H. Pham, "A type-2 fuzzy subtractive clustering algorithm," in *Advances in Intelligent and Soft Computing*, vol. 125 AISC, 2012, pp. 395–402, doi: 10.1007/978-3-642-27329-2_54.

BIOGRAPHIES OF AUTHORS

Dine Tiara Kusuma obtained her degree in informatics engineering from the PLN Institute of Technology, Jakarta, Indonesia in 2011, and has completed her master's studies in computer science at Budi Luhur University, Jakarta, Indonesia in 2015. Currently, she is a lecturer at the PLN Institute of Technology in the Informatics Engineering Study Program teaches artificial neural systems and databases and now she is a Ph.D. student at Universiti Teknikal Malaysia Melaka (UTeM). She is also an editor at SINTA accredited journal since 2019. Her current research fields are clustering and data science. She can be contacted at email: dinetiara@itpln.ac.id.

Norashikin Ahmad is a Senior Lecturer in Department of Software Engineering, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM). She received the degree of bachelor of information technology (computer science) from Universiti Kebangsaan Malaysia, master's in computer science from Universiti Teknologi Malaysia and Ph.D. in information technology from Monash University, Australia. Her Ph.D. thesis addresses the problem of incorporating incremental learning in artificial learning models and highlights its practical value and utility. Her research interests include database technologies and intelligence techniques, especially artificial neural networks, pattern recognition, and data mining in areas such as bioinformatics and tourism. She can be contacted at email: norashikin@utem.edu.my.

Sharifah Sakinah Syed Ahmad received B. applied science (Hons) computer modelling and M.Sc. mathematics from Universiti Sains Malaysia (USM). Her Ph.D. is in software engineering and intelligent system from University of Alberta, Canada. She is expert in computer aided geometric design (CAGD) and neural networks. He is currently an Associate Professor in the Department of Intelligent Computing and Analytics (ICA), Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM). She received her bachelor's and masters degree of applied mathematics in School of Mathematics from the University Science Malaysia. Following this, she received her Ph.D. from the University of Alberta, Canada in 2012 in intelligent system. Her research in graduate school focused on granular computing and fuzzy modeling. Her current research work is on the granular fuzzy rule-based system, evolutionary method, modeling, and data science. She can be contacted at email: sakinah@utem.edu.my.

Iriansyah BM Sangadji is currently a teaching staff at the Informatics Engineering Study Program, Faculty of Energy Telematics at the PLN Jakarta Institute of Technology and is still pursuing the Computer Science Doctoral program at IPB. Field of teaching expertise in the field of: artificial intelligence with artificial nervous system models, fuzzy systems, and multimedia. Apart from being a teacher, he is also an Assessor for the Competency Certification Institute in the field of electrical power distribution and has competence as a book and journal editor recognized by the National Professional Certification Agency (BNSP). He also contributed as an editor in several SINTA-accredited National Journals as well as popular magazines or e-Techno news portals in the energy and electricity sector. The field of research involved is the application and implementation of artificial intelligence in the fields of energy and electricity, construction management and agriculture. He can be contacted at email: iriansyah@itpln.ac.id.

Yozika Arvio completed elementary to high school education in Padang City, bachelor's degree at the PLN Jakarta Engineering College which is currently transforming into the PLN Institute of Technology (ITPLN) and completed Postgraduate at Budi Luhur University Jakarta Computer Science Study Program in 2020 as a Magna Cumlaude graduate. Currently she is a teaching staff at the Informatics Engineering Study Program, Faculty of Energy Telematics. Apart from actively participating in campus activities, she also contributes as an editor to the SINTA-accredited National Journal as well as popular magazines or e-Techno news portals in the field of opinion and entertainment. The research fields currently involved are related to machine learning, data mining and intelligent computing. She can be contacted at email: yozika@itpln.ac.id.