Indonesian Journal of Electrical Engineering and Computer Science
Vol. 30, No. 1, April 2023, pp. 33~45
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v30.i1.pp33-45 ❒ 33

Fuzzy C-means clustering based on micro-spatial analysis for electricity load profile characterization

Adri Senen[1,2], Tri Wahyu Oktaviana Putri[2], Jasrul Jamani Jamian[1], Eko Supriyanto[1], Dwi Anggaini[2]
[1]Department of Electrical Power Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[2]Department Electrical Engineering, Institut Teknologi PLN, Jakarta, Indonesia

Article Info

Article history:

Received Oct 6, 2022
Revised Dec 20, 2022
Accepted Dec 27, 2022

Keywords:

Clustering
Descriptive statistic
Fuzzy C-means
Grid
Load profile

ABSTRACT

As the rising of electricity demand, electricity load profile characterization (ELPC) is the integral aspect in planning, operating system, and distribution network development. The approach in the existing ELPC is still relatively macro in nature and does not involve other aspects outside the electricity variable, so the results tend to be biased for areas experiencing rapid land use changes. Therefore, this paper proposes an ELPC approach based on micro-spatial. Microspatial analysis is done by dividing area in the form of the smallest grids involving various electrical, demographic, geographic and socio-economic variables, which are then grouped using adaptive clustering based on fuzzy C-means (FCM). The adaptive clustering algorithm is proven to be able to determine the degree of membership of each grid data against each cluster with the ability to determine the number of clusters automatically according to the attribute data provided. The ELPC results which consist of 5 clusters are then analyzed using descriptive statistic, plotted, and mapped to obtain more accurate and realistic load characteristics in accordance with the pattern and geographical conditions of the region, so that the results can be used as a reference in load forecasting, network development, and distributed generation (DG) integration.

Corresponding Author:

Adri Senen
Department of Electrical Power Engineering, Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Johor Bahru, Malaysia
Email: adrisenen@itpln.ac.id

1. INTRODUCTION

One key factor that determine the quality of electrical energy is the existence of the distribution system planning that meets applicable technical standards and criteria. That feasible plan will be able to provide optimum service flexibility, so it can anticipate the increased electrical energy needs and the served load density. The distribution system planning must also be in accordance with the load characteristic pattern [1], so that the location and capacity of the distribution substations that will serve the load area can be determined in accordance with the regional profile.

Commonly, electrical load forecasting is used as one of the references in distribution system planning. Usually, many loads forecasting done has not considered to use load profile characterization (LPC) as its integral reference to get more accurate prediction result. Whereas the more precise load forecast can result in the better quality of network development planning, the better design of energy conservation, the increasing efficiency of the power system due to the high quality of load management strategies, precise customer classification, as well as precise marketing policies and electricity tariff [2], [3]. However, the usual LPC result is based on macro and sectoral, making it difficult to map and determine the existence of load centers [4]. This will be even worse for areas that has rapid and dynamic land use changes which is usually in line with the increasing of electricity load as the result of both economic and population growth [5].

Therefore, this research provides a novel approach in determining load profile characterization by performing a micro-spatial-based adaptive fuzzy (FCM) clustering technique to identify the pattern of load needs, types, and characteristics of the load C-means served by a smaller area (micro-spatial) [6]. The clustering method is needed so the physical areas can be grouped into some classes that consist of another similar characteristic areas [7], [8]. Fuzzy C-means clustering is one of the algorithms of clustering besides another algorithms which are hierarchical, K-means, and dynamic [9]. The well-known clustering method is K-means which basically uses iterative method for clustering [8], [10], [11] then the result can be used to analyze the power consumption behaviour [12]. If the clustering is done by considering each data as a separate cluster and then merging the similar ones, it is called hierarchical clustering as done in some literatures [10], [13], [14]. Fuzzy clustering or fuzzy C-means algorithm is used in this research because it can eliminate the clustering inaccuracy because of overlapping clusters [15]–[17]. The cluster results then is analyzed for its load profile characterization. The statictical method based on time series data can be implemented to get the statistic descriptive [12]. Another approach of load profile characterization is by using the frequency domain analysis [18], but the accuracy depends on the data sampling frequency.

The area studied in this research is a sub-district area of 120 sub-districts (grid). The low pin count (LPC) carried out considers data that consist of 12 variables, including electrical and non-electrical variables, from each sub-district. With so many grids and variables, the clustering method based on unsupervised learning using adaptive clustering algorithm is suitable to be applied because it does not require data labeling. Clustering is a grouping method based on unsupervised learning where the algorithm tries to find a certain structure or pattern in a data set, where the objects in each cluster show a certain degree of similarity [19]. Clustering method is widely implemented to classify some areas then identify electrical load profile characterization [6], [20]. Commonly that methods have 2 types: hard clustering and soft clustering. If a number of data are divided into different clusters where each data element is only a member of 1 cluster then the method is referred as hard clustering [21]. On the other hand if each data element is assigned a degree of membership for each cluster then the method is referred as soft clustering [22]. Soft clustering is more suitable to be applied because it represents the state of the data and the relationship between the variables. However, the analysis of the regional (spatial) electrical load profile that has been carried out by [23] only in which the clustering was done by using neural network.

Therefore, in order to obtain the most accurate possible load profile grouping, the clustering method used in this study is soft clustering which is implemented in the adaptive clustering algorithm. In this case of LPC clustering, the algorithm applied for the grouping is adaptive clustering. The algorithm is proven to be able to determine the degree of membership of each grid data against each cluster with the ability to determine the number of clusters or the number of groups that have the best performance automatically according to the attribute data provided.

This is very different from the conventional clustering algorithm where the clustering is done randomly or sequentially from the lowest number of clusters and then its performance is evaluated manually using the Silhouette cluster validation method. In addition, the LPC approach with this method is able to display and identify load characteristics based on geographic, demographic and economic structures for a smaller area by analytic-descriptive methods in each cluster, so that information relating to how large and where the electrical load centre is located can be mapped more accurately to be used in obtaining an optimal service pattern.

2. METHOD

2.1. Microspatial analysis

Micro-spatial analysis is an area analysis based on land use simulation modeling, which begins by dividing the service area and electricity utility into a set of small areas [3], [6]. The size of the grid varies depending on the availability of data and the variables used. Information on the pattern of land use development is then translated into that the results will be more accurate. Based on Figure 1 we can can see details in pattern of development of load requirements, so that LPC can be determined in a small area. In this research, the division of areas in the form of grids is based on the existing condition of the area, up to the sub-district level. The variables used include the electricity variable that is the peak load per sector and also non-electricity variables which include geography, demography, and socioeconomics

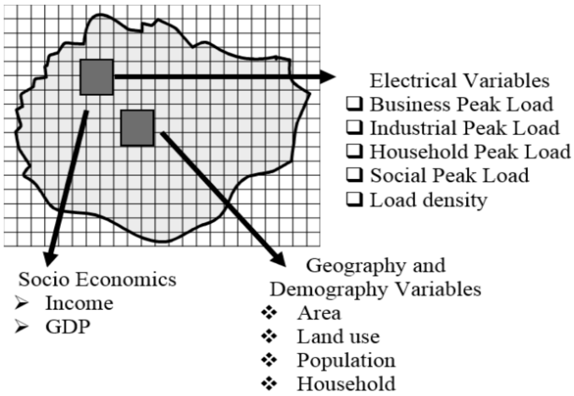


Figure 1. The region division in micro-spatial area

2.2. Soft clustering

Soft clustering is more suitable to be applied because it represents the state of the data and the relationship between the variables. One common method of soft clustering is fuzzy C-means (FCM). The FCM algorithm is [11]:

- Input the data as data X, in the form of a matrix of size $n \times p$ (n=number of data samples, p=attributes of each data).

  X_{kj} = data sample k-th ($k = 1,2, \dots, n$), data attribute j-th ($j = 1,2,3, \dots, m$).

- Determine the number of clusters, weighting power, maximum iteration, expected error, and objective function and initial iteration.
- Generate random number (μ_{ik}, $i = 1,2, \dots, c$; $k = 1,2, \dots, n$), as elements of the initial partition matrix U.

$$U_0 = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \cdots & \mu_{1c}(x_c) \\ \vdots & \vdots & & \vdots \\ \mu_{11}(x_1) & \mu_{12}(x_2) & \cdots & \mu_{1c}(x_c) \end{bmatrix} \quad (1)$$

The partition matrix in fuzzy clustering must meet the following conditions:

$$\mu_{ik} = [0,1];\ (1 \le i \le c;\ 1 \le k \le n) \quad (2)$$

$$\textstyle\sum_{i=1}^{n} \mu_{ik} = 1; 1 \le i \le c$$

$$0 < \textstyle\sum_{i=1}^{c} \mu_{ik} < c\ ; 1 \le k \le n$$

Count the number of each column (attribute):

$$Q_j = \textstyle\sum_{i=1}^{c} \mu_{ik} \quad (3)$$

with $j = 1,2,3, \dots, m$
then calculate the degree of membership of each data.

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \quad (4)$$

- Calculate the cluster centroid of the k-cluster: V_{ij}, where $i = 1,2,3,\dots,c$ and $j = 1,2,3,\dots,m$

$$V_{ij} = \frac{\sum_{k=1}^{n}((\mu_{ik})^m * X_{kj})}{\sum_{k=1}^{n}(\mu_{ik})^m} \quad (5)$$

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{c1} & \cdots & v_{cm} \end{bmatrix}$$

- Calculate the objective function, P_t

$$P_t = \sum_{k=1}^{n} \sum_{i=1}^{c} \left(\left[\sum_{j=1}^{m} \left(X_{kj} - V_{ij} \right)^2 \right] (\mu_{ik})^m \right) \tag{6}$$

- Calculate the partition matrix change:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{p} \left(X_{kj} - V_{ij} \right)^2 \right]^{\frac{-1}{p-1}}}{\sum_{i=1}^{c} \left[\sum_{j=1}^{p} \left(X_{kj} - V_{ij} \right)^2 \right]^{\frac{-1}{p-1}}} \tag{7}$$

- Check stops condition:
  a. If ($|Pt - Pt{-}1| < \xi$) or (t< max iteration) so it is done.
  b. If not, then $t=t+1$ repeat to step 4.

2.3. Cluster validation

Cluster validation or clustering validation index (CVI) is done to see the quality and strength of the cluster [24]. Optimal cluster results can be influenced by the clustering method used, the characteristics of the dataset, the structure of the data density, the size of the data, the number of clusters used. Thus, it is important to observe the quality of the cluster using CVI. There are many methods for CVI like Silhouette index [25], [26], Davies-Bouldin score, Calinski-Harabasz score, Dunn's index [27], [28], and many more. Some literatures combined two CVIs [25], [26].

In this paper, the validation method used is the silhouette coefficient or silhouette index. The silhouette coefficient method is a combination of two methods, namely the cohesion method which functions to measure the proximity of the data in one group and the separation method which functions to measure the closeness between the formed groups. When applied, the Silhouette algorithm will measure the average distance from an object to all objects in the same cluster with objects in other clusters [29]. The silhouette plot shows the silhouettes of all groups so that the quality of the groups can be compared based on the width (darkness) of the silhouettes. The equation in determining the silhouette coefficient is:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i)} \tag{8}$$

where s (i) = value of silhouette coefficient
(i) = average distance of i-data
b (i) = average distance of i-data with all members

The interpretation of silhouette value is determined with the Silhouette Width Index as shown in Table 1 [30].

Table 1. Silhouette width index

Silhouette Coefficient	Interpretation between objects and groups formed
0.7 < SC < 1	Strong structure
0.5 < SC < 0.7	Medium structure
0.25 < SC < 0.5	Weak structure
SC < 0.25	No relation

2.4. Adaptive clustering

The novelty in this research is the application of the adaptive clustering algorithm. Adaptive is an appropriate term because the algorithm can automatically generate data grouping with the most optimal number of clusters based on the evaluation of cluster results. This is very different from the conventional clustering algorithm where the clustering is done randomly or sequentially from the lowest number of clusters and then its performance is evaluated manually using the cluster validation method. It is obvious that the conventional clustering methods are ineffective and more time consuming.

The clustering algorithm used in this adaptive clustering method is fuzzy C-means with group evaluation based on silhouette validation. The concept of adaptive clustering can be seen in Figure 2. The output generated from the adaptive clustering algorithm is the most optimal number of clusters, members of each cluster, and the average silhouette value. From the flowchart in Figure 2, the clustering process and its cluster evaluation, are done simultaneously to give the most optimal cluster number based on any data input. This is more effective than processing the clustering and its evaluation as separated algorithm.

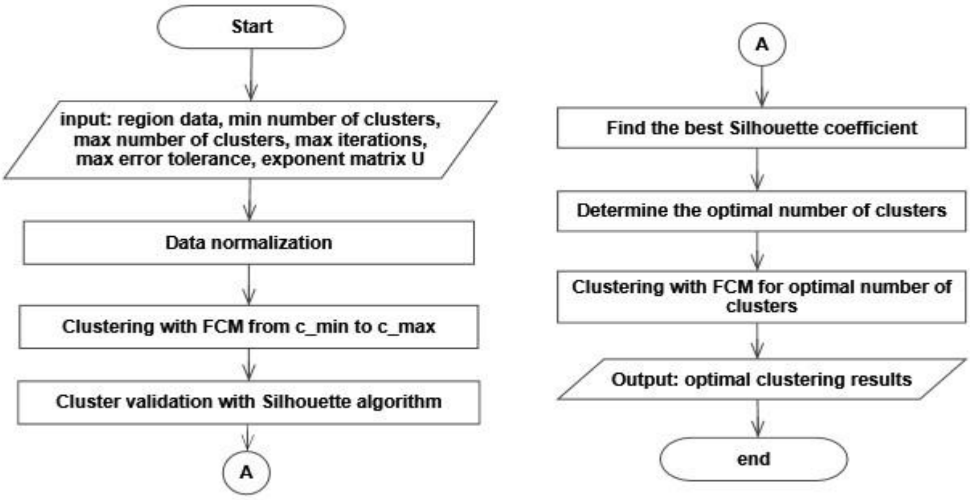


Figure 2. Adaptive clustering flowchart

2.5. Area mapping

The area mapping process is carried out to help when the available geographic data is large in number and size and consists of many interrelated themes. This process is carried out after a cluster is formed which contains grids that are analyzed to connect various data at a certain point on the earth, combine them, analyze, and finally map the results. The data to be processed is spatial data, the results obtained are geographically oriented data and are locations that have a certain coordinate system, as the basis for reference.

Spatial data and plotting carried out such as land zones and topography on maps will have data attributes that contain information that can be adapted to LPC. The spatial data can also be combined with other spatial data so that it becomes layers containing complementary data. The use of Geographic Information Systems requires technical standards such as map projection systems and types of layers in order to maintain the level of accuracy and is useful in facilitating the planning and development of distribution networks.

2.6. Load Profile Characterization

Determination of LPC using descriptive statistical method for each cluster. Each member (grid) in the cluster will have an identical LPC. It will be displayed and analyzed in accordance with the variables entered. Descriptive statistics is a statistical analysis process that focuses on the management, presentation, and classification of data. With this process, the data presented will be easier to understand, and able to provide more meaning for data users.

The graphic representation of descriptive statistics method to determine the LPC is shown in Figure 3, where Figure 3(a) is skewness coefficient criteria and Figure 3(b) is kurtosis coefficient criteria. Both coefficient criteria can give the information of data distribution and its dominant value. The size of the slope (skewness) refers to the departure of the distribution from symmetry [31]. If the slope value is known, then it is also known how the distribution model is, whether symmetrical, negative, or positive. To calculate the slope coefficient, the following Person's moment coefficient of skewness formula is used [32].

$$a_3 = 3\left[\frac{\mu - m}{\sigma}\right] \tag{9}$$

With a_3 = skewness (Slope Coefficient)
μ = mean
m = median
σ = deviation standard

Criteria for the form of skewness are positive model if the skewness value is more than zero; negative model if skewness is less than zero; symmetrical model if the skewness value is equal to zero. The measure of curvature (kurtosis) is the degree of height of the peak of a frequency distribution, usually relative to the normal distribution [33]. To find out the type of sharpness of the graph, the kurtosis value can be calculated using the following formula:

$$K = \frac{\frac{1}{n}\sum(x-\mu)^4}{\sigma^4} \tag{10}$$

with K = kurtosis coefficient
n = amount of data
x = data value

The criteria for the distribution model of kurtosis can be categorized into three, those are [34]: platykurtic distribution if the kurtosis coefficient < 0.263; mesokurtic distribution if the kurtosis coefficient = 0.263; and leptokurtic distribution if the kurtosis coefficient > 0.263.

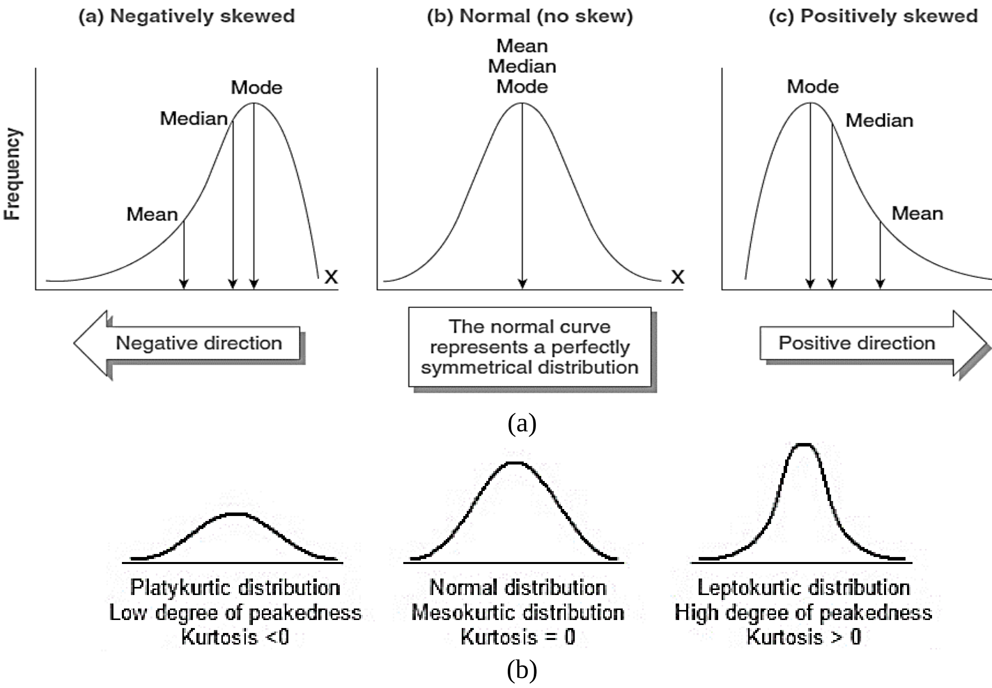


Figure 3. Graphic representation of descriptive statistics method of (a) skewness and (b) kurtosis coefficient criteria

3. RESULTS AND DISCUSSION

3.1. Data

The area studied in this research is a sub-district area of 120 sub-districts (grid). The LPC carried out considers data that consist of 12 variables, including electrical and non-electrical variables, from each sub-district. The electrical variables are load for each sector (residential, industry, commercial, and social) and load density. Non-electrical variables are total household, percentage of gross domestic product (GDP) growth, land use of each sector, and total area of sectors. The following tables (Tables 2 to 5) shows the data that contains 12 variables. Those are showed only 10 sub-districts out of 120 sub-districts data used in this research.

3.2. Clustering proccess

Before performing the clustering process, the data must be normalized. The normalization is critical step to group the attributes of various entities in a relation to form a good relationship structure (without redundancy or data repetition) and most of the ambiguity can be eliminated. Data plot comparison in Figure 4 shows the significant effect of normalization process to the data. It is clear from the results in Figure 4(a), which shows the plot of actual data, that the raw data are not evenly distributed to a point, so there tends to be ambiguity. However, after normalizing the data which is shown in Figure 4(b), the normalized data distribution forms a simple, non-redundant entity, so that it can be ensured that the data is in good quality for the next step. Clustering using adaptive clustering based on FCM methods and silhouette algorithm to find the best number of clusters. This process needs data input which is normalized data, minimum and maximum number of clusters, maximum iteration, and maximum error. By using adaptive clustering, the clustering processes start grouping the data into minimum number of clusters until the maximum number of clusters that determined in the initial step. Then, the result of each execution of clustering algorithm is evaluated by using silhouette methods. The best silhouette index is chosen to get the best number of clusters.

Table 2. Demographic and economic parameters

Sub-district (Grid)	∑ Household	GDP growth (%)
Kemuning	2379	11.38
Solear	1559	12.08
Rajeg	1610	8.08
Kedoya Utara	8351	99.87
Duri Kosambi	16228	208.98
Cibadak	2132	5.76
Bitung Jaya	2808	4.43
Cipete	1933	15.06
Batu ceper	4436	36.75
Kamal Muara	2083	99.96

Table 3. Geographical parameters

Sub-district (Grid)	Land Use (Ha)				Area
	Residential	Industry	Business	Social	(Ha)
Kemuning	39.8	2598.3	1732.2	73.6	4444
Solear	97.5	2737.9	1825.2	89.4	4750
Rajeg	85.1	1818.5	1212.3	86.1	3202
Kedoya Utara	266.9	2.8	4.1	50.2	324
Duri Kosambi	238.4	101.4	82.7	80.4	502.9
Cibadak	136.3	1255.8	837.2	122.8	2352
Bitung Jaya	167.3	931.2	620.8	150.7	1870
Cipete	166.4	0.0	0.0	50.8	217
Batu ceper	89.9	15.7	36.7	8.8	151
Kamal Muara	84.3	183.3	72.7	713.2	1053.4

Table 4. Electrical parameters

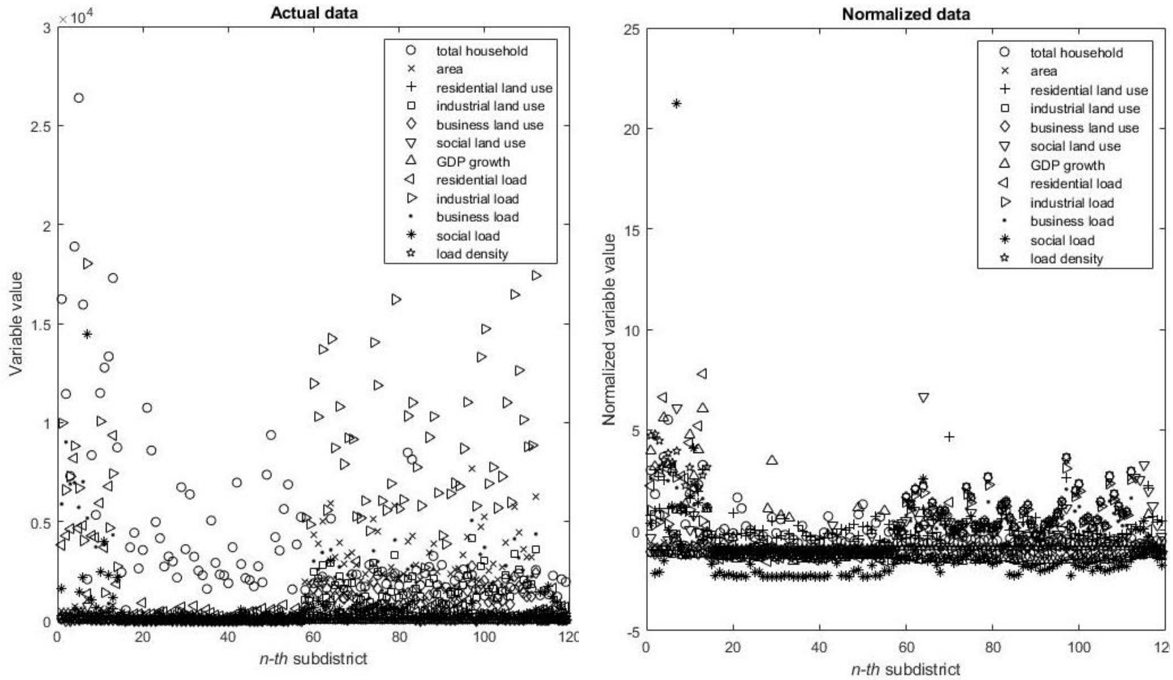
Sub-district (Grid)	Load (kW)				Load density (kW/Ha)
	Residential	Industry	Business	Social	
Kemuning	141.8	12639.8	3168.9	289.0	3.654
Solear	347.0	13318.6	3339.1	350.9	3.654
Rajeg	302.9	8846.4	2217.8	337.8	3.655
Kedoya Utara	4286.1	277.2	288.1	1019.7	18.120
Duri Kosambi	3827.4	9969.1	5884.6	1633.8	42.384
Cibadak	485.0	6108.9	1531.5	481.8	3.660
Bitung Jaya	595.4	4530.0	1135.7	591.5	3.665
Cipete	592.4	0.0	0.0	199.3	3.648
Batu ceper	320.0	76.4	67.1	34.7	3.299
Kamal Muara	1353.1	18023.0	5170.3	14489.4	37.057

This research determined to cluster the data from 2 to 10 number of clusters that was executed 20 times to ensure the consistency of the results to cluster the data. Figure 5 shows the plot of silhouette index after 4 times validity running. The graph is limited to show only 4 times validity running so the trend of the silhouette index from each clustering validity can be presented clearly. Based on the clustering that has been carried out several times, the number of clusters that often appear and form is 4 and 5 number of clusters. While the best silhouette index is 0.70414 for 5 number of clusters. The total members of each cluster, both from 4 and 5 number of clusters, can be seen in Table 5.

Cluster 1, 3, and 4 have the same total members. Although it has the same total members, the sub-districts listed are not exactly the same because of different centroid. For example, in cluster 4 that contains 51 sub-districts from both 4 and 5 number of clusters, there are 38 sub-districts listed in both, while the rest 13 sub-districts are different. The centroid for each variable for 4 and 5 number of clusters is shown in Table 6.

3.3. Cluster validation

To ensure the optimum number of clusters formed that is 5 clusters, then validation using silhouette algorithm is carried out using (8), the results of cluster validation are shown in the Figure 6. It shows the silhouette index for each sub-district. The value shows the interpretation between sub-district and groups formed. The good value for silhouette coefficient is close to +1, in which it means the grid is far from other groups formed. If the value equals to 0, the gris is on or very close to the decision boundary grid If the value is close to +1. More negative value means the grid is overlapping, so the grid is more appropriate to be grouped in another cluster. From Figure 6 the average value of the silhouette coefficient is 0.704141, based on Table 1, the silhouette value which is above 0.7 the interpretation between the objects formed is very strong structure.



(a) (b)

Figure 4. The data plot comparison between: (a) actual data and (b) normalized data

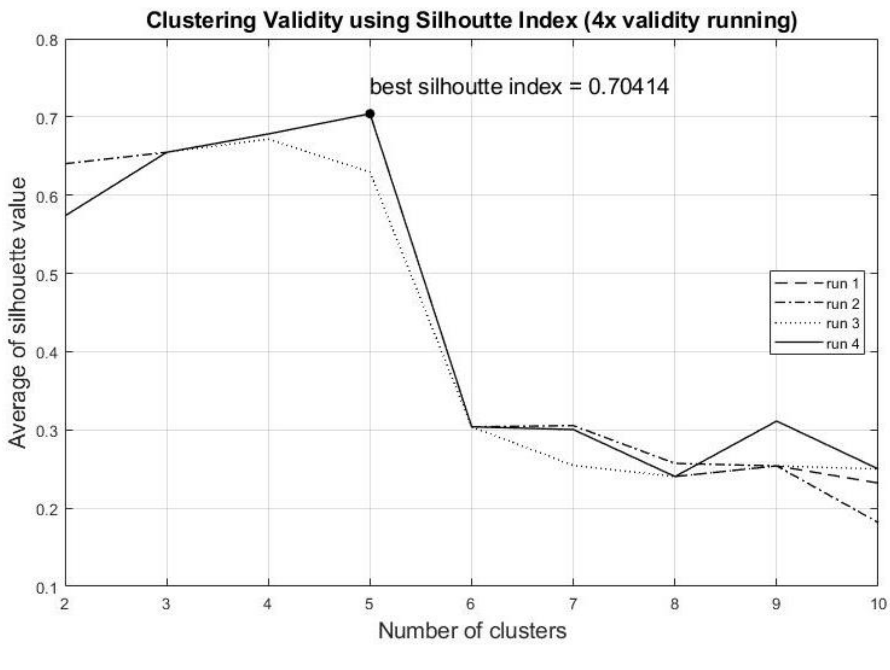


Figure 5. Best silhouette index from clustering validity processes

Table 5. Total members of each cluster

Number of Cluster	Total member of k-th cluster				
	1	2	3	4	5
5 Clusters	35	13	19	51	1
4 Clusters	35	14	19	51	0

Table 6. Centroid for 4 and 5 number of clusters

Variable	Centroid	
	5 Clusters	4 Clusters
∑ Household	-0,343	-0,339
Area	1,770	1,761
Land Use: Residential	0,263	0,265
Land Use: Industry	1,893	1,878
Land Use: Business	1,872	1,857
Land Use: Social	0,444	0,508
GDP Growth	-1,295	-1,277
Load: Residential	-1,045	-1,036
Load: Industry	1,607	1,620
Load: Business	0,908	0,921
Load: Social	-1,288	-1,112
Load Density	-0,837	-0,802

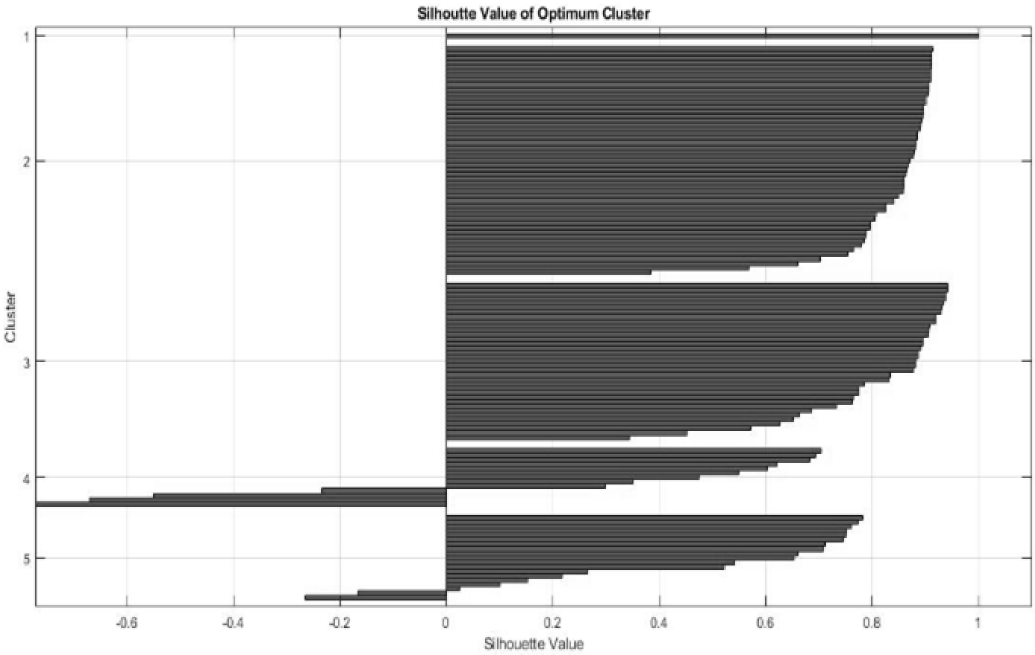


Figure 6. Cluster validation

3.4. Area mapping

The clustering results obtained are then plotted and mapped to see the distribution, location of the area of each grid in each cluster, to make it easier to perform LFC from each area. The results of mapping the area can be seen in Figure 7. From the Table 5, there is a cluster that only has 1 member that is Kamal Muara sub-districts. The location of Kamal Muara can be seen in Figure 7. Only that sub-district that clustered distinctively. From a brief observation, if a grid tends to make their own cluster with only that grid as the cluster member, it means the attributes of that grid are far different from another. From the raw data, Kamal Muara sub-district has rather unique characteristics where the social sector land use and social sector load are high, no other areas have such that characteristics. Geographically, Kamal Muara sub-district is located in the coastal area with the mangrove forest dominated the area.

3.5. Load profile characterization

LPC is obtained after all areas are clustered using descriptive statistical method for each cluster. The descriptive statistics used are skewness and kurtosis analysis using SPSS software. Table 7 shows the result of descriptive statistics analysis of cluster 4.

From Table 7, electrical variables can be analyzed based on the value of skewness and kurtosis for household (residential) sector load, industrial sector load, business sector load, and social sector loads. The load analysis can be represented as a graph of skewness and kurtosis as shown in Figure 8. Thus, the load profile characterization of each sector is:

- Residential sector: the skewness value is more than zero and the kurtosis value is more than 0.263, which means that the residential load of each customer tends to be low but the number of consumers for residential sector tend to be high. In addition, it can be seen based on the graph of kurtosis and Skewness in Figure 8(a).
- Industrial sector: the skewness value is less than zero and the kurtosis value is less than 0.263, this shows that the industrial load of each customer tends to be high with a relatively small number of customers. The skewness and kurtosis graphs for industrial loads are described as Figure 8(b).
- Business sector: the skewness value is less than zero and the kurtosis is less than 0.263, this is the same as the industrial sector, where the business sector load of each customer tends to be high with a relatively small number of customers. The frequency distribution of business expenses according to the following Figure 8(c).
- Social sector: as seen in Figure 8(d), the skewness value is more than zero and the kurtosis value is more than 0.263, which means that the social load of each customer but the number of consumers for social sector tend to be high.

Therefore, the characteristics of the electrical load in cluster 4 are dominated by residential loads because the kurtosis value is the highest among the other three variables, which is 0.76, so that the residential load kurtosis curve is sharper than the industrial load, business load, and social load. The same method can be implemented in another clusters. By knowing the load profile characterization of every cluster, the distribution system planning will be more specific in the level of sub-district, so that the location and capacity of the distribution substations that will serve the load area can be determined in accordance with the regional profile.

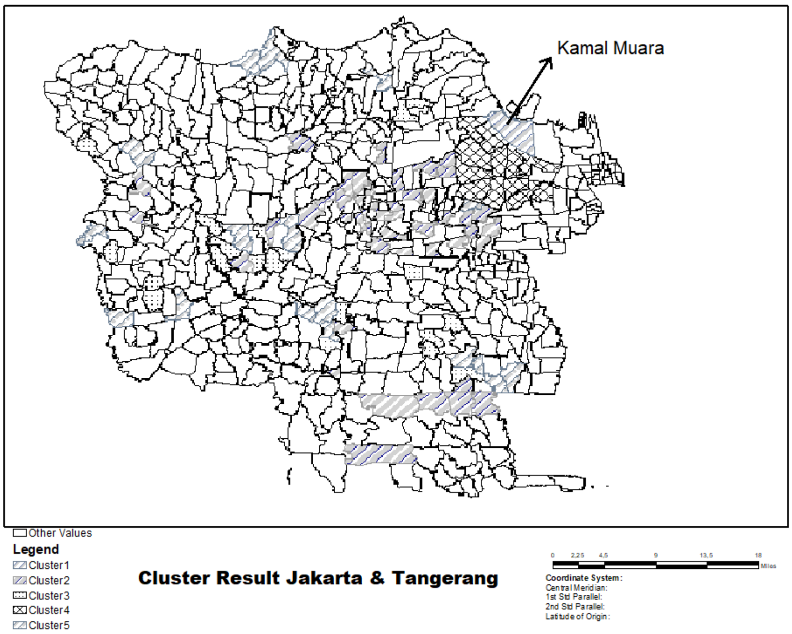
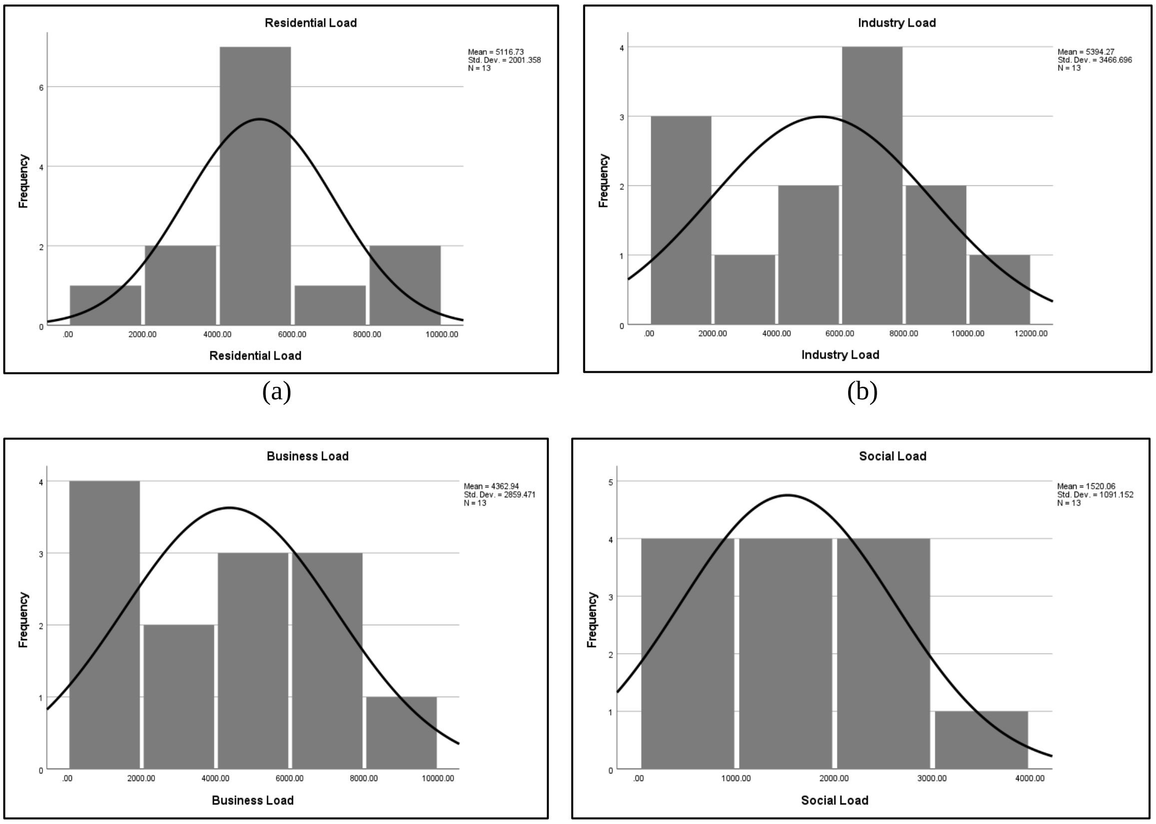


Figure 7. Mapping result of 5 clusters and the position of Kamal Muara sub-district

Table 7. Result of descriptive statistics analysis of cluster 4

Variable	Descriptive statistics		
	Mean	Skewness	Kurtosis
∑ Household	13348.46	0.82	0.98
Area	509.75	0.22	-0.14
Land Use: Residential	318.68	0.84	0.76
Land Use: Industry	54.86	-0.25	-1.23
Land Use: Business	61.33	-0.15	-1.13
Land Use: Social	74.82	0.70	0.57
GDP Growth	184.88	0.15	-0.59
Load: Residential	5116.73	0.84	0.76
Load: Industry	5394.27	-0.25	-1.23
Load: Business	4362.94	-0.15	-1.13
Load: Social	1520.06	0.70	0.57
Load Density	31.88	-0.07	-0.56



(a) (b)

(c) (d)

Figure 8. Load kurtosis and skewness graph of (a) residential sector, (b) industrial sector, (c) business sector, and (d) social sector

4. CONCLUSION

The area studied in this research is a sub-district area of 120 sub-districts (grid). The LPC carried out considers data that consist of 12 variables, including electrical and non-electrical variables, from each sub-district. The grids must be grouped into some clusters that contains the cluster members which have the similar profile, using adaptive clustering. The proposed clustering method used in this adaptive clustering algorithm is based on fuzzy C-means with group evaluation based on Silhouette validation. Based on the clustering that has been carried out several times, the number of clusters that often appear and form is 4 and 5 number of clusters. While the best silhouette index is 0.70414 for 5 number of clusters. Thus, the output of clustering process was analyzed for its electricity load profile characterization (LPC) using descriptive statistical method for each cluster. LPC has done for cluster 4 that contains 13 grids (sub-districts) which gives the result that the electrical load in cluster 4 are dominated by residential loads where the number of customers is high but the load for each residential customer tends to be low. The same method can be implemented in another clusters.

REFERENCES

[1] A. Parker, K. James, D. Peng, and M. A. Alahmad, “Framework for extracting and characterizing load profile variability based on a comparative study of different wavelet functions,” *IEEE Access*, vol. 8, pp. 217483–217498, 2020, doi: 10.1109/ACCESS.2020.3042125.

[2] K. A. Choksi, S. Jain, and N. M. Pindoriya, “Feature based clustering technique for investigation of domestic load profiles and probabilistic variation assessment: Smart meter dataset,” *Sustain. Energy, Grids Networks*, vol. 22, p. 100346, 2020, doi: 10.1016/j.segan.2020.100346.

[3] N. Anuar, N. K. K. Baharin, N. H. M. Nizam, A. N. Fadzilah, S. E. M. Nazri, and N. M. Lip, “Determination of typical electricity load profile by using double clustering of fuzzy C-means and hierarchical method,” *2021 IEEE 12th Control Syst. Grad. Res. Colloquium, ICSGRC 2021 - Proc.*, no. August, pp. 277–280, 2021, doi: 10.1109/ICSGRC53186.2021.9515295.

[4] G. Trotta, “An empirical analysis of domestic electricity load profiles: Who consumes how much and when?,” *Appl. Energy*, vol. 275, no. January, p. 115399, 2020, doi: 10.1016/j.apenergy.2020.115399.

[5] A. Senen, C. Widyastuti, O. Handayani, and P. Putera, “Development of micro-spatial electricity load forecasting methodology using multivariate analysis for dynamic area in Tangerang , Indonesia,” vol. 29, no. 4, pp. 2565–2578, 2021.

[6] C. Widyastuti, A. Senen, and O. Handayani, “Micro-spatial electricity load forecasting using clustering technique,” *Proceeding - 2020 2nd Int. Conf. Ind. Electr. Electron. ICIEE 2020*, vol. 11005, pp. 17–21, 2020, doi: 10.1109/ICIEE49813.2020.9277274.

[7] H. C. Jeong, M. Jang, T. Kim, and S. K. Joo, "Clustering of load profiles of residential customers using extreme points and demographic characteristics," *Electron.*, vol. 10, no. 3, pp. 1–10, 2021, doi: 10.3390/electronics10030290.
[8] S. Cen, J. H. Yoo, and C. G. Lim, "Electricity pattern analysis by clustering domestic load profiles using discrete wavelet transform," *Energies*, vol. 15, no. 4, 2022, doi: 10.3390/en15041350.
[9] S. Deng, "Clustering with fuzzy C-means and common challenges," *J. Phys. Conf. Ser.*, vol. 1453, no. 1, 2020, doi: 10.1088/1742-6596/1453/1/012137.
[10] A. Satre-Meloy, M. Diakonova, and P. Grünewald, "Cluster analysis and prediction of residential peak demand profiles using occupant activity data," *Appl. Energy*, vol. 260, no. September 2019, p. 114246, 2020, doi: 10.1016/j.apenergy.2019.114246.
[11] H. Bian, Y. Zhong, J. Sun, and F. Shi, "Study on power consumption load forecast based on K-means clustering and FCM–BP model," *Energy Reports*, vol. 6, pp. 693–700, 2020, doi: 10.1016/j.egyr.2020.11.148.
[12] 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.
[13] L. Xu, Y. Zhang, and Z. Shao, "An approach to cluster electrical load profiles based on piecewise symbolic aggregation," *Proc. - 2021 6th Asia Conf. Power Electr. Eng. ACPEE 2021*, pp. 1000–1004, 2021, doi: 10.1109/ACPEE51499.2021.9437105.
[14] G. L. Ray and P. Pinson, "Online adaptive clustering algorithm for load profiling," *Sustain. Energy, Grids Networks*, vol. 17, p. 100181, 2019, doi: 10.1016/j.segan.2018.100181.
[15] C. Liu *et al.*, "A moving shape-based robust fuzzy K-modes clustering algorithm for electricity profiles," *Electr. Power Syst. Res.*, vol. 187, no. May, p. 106425, 2020, doi: 10.1016/j.epsr.2020.106425.
[16] M. Jain, T. Alskaif, and S. Dev, "Validating clustering frameworks for electric load demand profiles," *IEEE Trans. Ind. Informatics*, vol. XX, no. XX, 2021, doi: 10.1109/TII.2021.3061470.
[17] K. Zhou, S. Yang, and Z. Shao, "Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study," *J. Clean. Prod.*, vol. 141, pp. 900–908, 2017, doi: 10.1016/j.jclepro.2016.09.165.
[18] M. Sanabria-Villamizar, M. Bueno-López, J. C. Hernández, and D. Vera, "Characterization of household-consumption load profiles in the time and frequency domain," *Int. J. Electr. Power Energy Syst.*, vol. 137, no. September 2021, 2022, doi: 10.1016/j.ijepes.2021.107756.
[19] L. Sun, K. Zhou, and S. Yang, "An ensemble clustering based framework for household load profiling and driven factors identification," *Sustain. Cities Soc.*, vol. 53, p. 101958, 2020, doi: 10.1016/j.scs.2019.101958.
[20] A. Bossio *et al.*, "Performance assessment of load profiles clustering methods based on silhouette analysis," *IEEE*, 2021, [Online]. Available: https://ieeexplore.ieee.org/document/9584629/citations#citations.
[21] G. Franco, C. P. E. de Souza, and N. L. Garcia, "Aggregated functional data model applied on clustering and disaggregation of UK electrical load profiles," 2021, [Online]. Available: http://arxiv.org/abs/2106.11448.
[22] M. Pietrzykowski, "Mini-models based on soft clustering methods," *Procedia Comput. Sci.*, vol. 159, pp. 2512–2521, 2019, doi: 10.1016/j.procs.2019.09.426.
[23] M. Flor, S. Herraiz, and I. Contreras, "Definition of residential power load profiles clusters using machine learning and spatial analysis," *Energies*, vol. 14, no. 20, 2021, doi: 10.3390/en14206565.
[24] J. Hämäläinen, S. Jauhiainen, and T. Kärkkäinen, "Comparison of internal clustering validation indices for prototype-based clustering," *Algorithms*, vol. 10, no. 3, 2017, doi: 10.3390/a10030105.
[25] X. Wang and Y. Xu, "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 569, no. 5, 2019, doi: 10.1088/1757-899X/569/5/052024.
[26] T. Thinsungnoen, N. Kaoungku, P. Durongdumronchai, K. Kerdprasop, and N. Kerdprasop, "The clustering validity with silhouette and sum of squared errors," no. January 2015, pp. 44–51, 2015, doi: 10.12792/iciae2015.012.
[27] R. Al-Otaibi, N. Jin, T. Wilcox, and P. Flach, "Feature construction and calibration for clustering daily load curves from smart-meter data," *IEEE Trans. Ind. Informatics*, vol. 12, no. 2, pp. 645–654, 2016, doi: 10.1109/TII.2016.2528819.
[28] H. Bulut, Y. Öner, and Ç. Sözen, "Clustering of member and candidate countries of the European Union," *Int. J. Sci. Basic Appl. Res.*, vol. 36, no. 7, pp. 18–25, 2017, [Online]. Available: http://gssrr.org/index.php?journal=JournalOfBasicAndApplied.
[29] N. Thamrin and A. W. Wijayanto, "Comparison of soft and hard clustering: a case study on welfare level in cities on Java Island," *Indones. J. Stat. Its Appl.*, vol. 5, no. 1, pp. 141–160, 2021, doi: 10.29244/ijsa.v5i1p141-160.
[30] L. Kaufman and P. J. Rousseeuw, *Finding Groups in data: An Introduction to Cluster Analysis*. New York: J. Wiley & Son, 1990.
[31] A. K. Singh, L. P. Gewali, and J. Khatiwada, "New Measures of Skewness of a Probability Distribution," *Open J. Stat.*, vol. 09, no. 05, pp. 601–621, 2019, doi: 10.4236/ojs.2019.95039.
[32] M. A. Eltehiwy and A.-B. A. AbdulMotaal, "A new coefficient of Skewness for grouped data," *Biometrics Biostat. Int. J.*, vol. 9, no. 2, pp. 54–59, 2020, doi: 10.15406/bbij.2020.09.00300.
[33] K. Chumpong, K. Mekchay, and N. Thamrongrat, "Analytical formulas for pricing discretely - sampled skewness and," no. March, 2022.
[34] J. Einasto, A. Klypin, G. Hütsi, L. J. Liivamägi, and M. Einasto, "Evolution of skewness and kurtosis of cosmic density fields," *Astron. Astrophys.*, vol. 652, 2021, doi: 10.1051/0004-6361/202039999.

BIOGRAPHIES OF AUTHORS

Adri Senen received the Bachelor Degree in Electrical Engineering from Andalas University, Indonesia in 2004, and Master Degree in electrical power engineering from Bandung Insitute of Technology (ITB), Indonesia in 2008. Currently he is a PhD student in School of Electrical Engineering in Universiti Teknologi Malaysia. His research interests concern load forecasting, management energy, electrical planning, renewable energy and power system. He can be contacted at email: adrisenen@itpln.ac.id.

Tri Wahyu Oktaviana Putri is an lecturer at the Department of Electrical Engineering, Institut Teknologi PLN, Indonesia, where she has been a faculty member since 2017. She received her bachelor in control engineering from Department of Electrical Engineering, Universitas Brawijaya, Malang, Indonesia, in 2014, and an M.Sc. in Electrical Engineering with specialization in control engineering and intelligent system Institut Teknologi Bandung (ITB), Indonesia, in 2017. Her research interests are primarily in the area of control engineering, system modelling and identification, intelligent system, robotic, and artificial intelligent. She can be contacted at email: triwahyu@itpln.ac.id.

Jasrul Jamani Jamian received the Bachelor of Engineering (B. Eng. (Hons)) degree, Master of Engineering (M. Eng.) and Ph.D degree in electrical (power) engineering from Universiti Teknologi Malaysia in 2008, 2010 and 2013 respectively. He is currently director for Power Engineering Division, School of Electrical Engineering, Universiti Teknologi Malaysia. Dr Jasrul is actively involved in research as a principal investigator as well as leader in consultancy projects with several companies such as Petronas and Tenaga Nasional Berhad, which focuses on relay coordination projects and off grid solar PV design. He is the author and co-author of more than 80 publications in international journals and proceedings in the area of Power Systems and Energy. His research interest includes Network Reconfiguration, Optimization technique, and Renewable Energy. He can be contacted at email: jasrul@fke.utm.my.

Eko Supriyanto obtained his Doctor of Engineering from University of Federal Armed Forces Hamburg, Germany. He obtained his professorship from Universiti Teknologi Malaysia at early 30s age, which is one of the world youngest Professor. He published more than 300 international journal and proceeding papers during last 12 years, registered more than 50 patents and copyrights, and received more than 30 international awards, including from National Research Council of Thailand and Korea Invention Promotion Association. Currently he is a full Professor at Universiti Teknologi Malaysia, Adjunct Professor at Ilmenau University of Technology Germany, senior research fellow at University of Indonesia, Senior Consultant for PETRONAS Malaysia, Management Consultant for PLN (State Electricity Company) Indonesia. He can be contacted at email: eko@utm.my.

Dwi Anggaini received the Bachelor Degree in Mathematics Education (S.Pd) from Muhammadiyah University Prof. Dr. HAMKA, Indonesia in 2012, and Master Degree in Reseach and Evaluation of Education (M.Pd) from Graduate school of Muhammadiyah University Prof. Dr. HAMKA, Indonesia in 2015. Currently she is a lecturer at the Institut Teknologi PLN since 2017 and has developed several studies including education, statistics, data analysis and applied mathematics. She can be contacted at email: dwi_anggaini@itpln.ac.id.