IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 3, September 2024, pp. 3533~3544
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i3.pp3533-3544

Sectoral electricity micro-spatial load forecasting based on partitional clustering technique

Adri Senen[1,2], Jasrul Jamani Jamian[1], Hasna Satya Dini[1,2], Eko Supriyanto[3], Dwi Anggaini[2]
[1]Department of Electrical Power Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor, Malaysia
[2]Department of Electrical Engineering, Faculty of Electricity and Renewable Energy, Institut Teknologi PLN, Jakarta, Indonesia
[3]Department of Biomedical Engineering & Health Sciences, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor, Malaysia

Article Info

Article history:

Received Jul 18, 2023
Revised Feb 23, 2024
Accepted Mar 13, 2024

Keywords:

Grid
Load forecasting
Microspatial
Partitional clustering
Sectoral load

ABSTRACT

Load demand forecasting is crucial in energy supply planning due to economic progress and territorial expansion, where land utilization transforms dynamically. An accurate sectoral load prediction can preclude the loss of beneficial opportunities arising from excessive load demand or excessive investment at a low-growth juncture. However, the particular area in this sectoral approach is still relatively large, rendering it incapable of precisely projecting load at minor points (micro-spatial). This study has proposed a micro-spatial load prediction strategy that categorizes identified areas into smaller grids or districts. This procedure includes clustering similar sites together for improved accuracy. K-Means is one of the partitional clustering approaches, a clustering algorithm utilizing object-based centroid-based partitioning approaches. The algorithm determines a cluster's centroid or centre as the average point for the cluster. This technique is advantageous as it can process extensive data efficiently and is appropriate for circular data. This technique can divide the data into multiple partitions, ensuring that each object belongs to precisely one cluster. Subsequently, mathematical modelling is used to predict the load of each cluster, which can then be utilized to more accurately evaluate the positions and sizes of prospective substations, transmission, and distribution facilities.

Corresponding Author:

Adri Senen
Department of Electrical Power Engineering, Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Skudai, Johor Bahru, Johor 81310, Malaysia
Email: adrisenen@itpln.ac.id

1. INTRODUCTION

Load demand forecasting is an essential task in ensuring the efficient functioning and management of electricity distribution systems in a certain period in an observed area [1]–[5]. The ability to forecast future electricity consumption enables utility companies to plan and allocate appropriate resources, such as power generation capacity and transmission infrastructure, effectively meeting the growing demands of consumers [6], [7]. Precise load demand forecasting helps optimize energy production schedules, reducing costs associated with over or under-production [8]. Moreover, it aids in maintaining a balance between supply and demand, which enhances grid stability by preventing potential blackouts or system failures caused by excessive strain on the network due to unexpected spikes in consumer usage [5], [9].

Load forecasting by sectors involved four sectors, i.e., household, commercial, social, and industry [10]–[14]. This approach offers improved accuracy by estimating load growth within each sector. As a part of the master plan distribution, information on micro-spatial load growth is required to decide the location of the distribution substation [7], [15]. However, most observed areas are macro-scale, making it difficult to pinpoint the load center in smaller areas [16], [17]. For this reason, micro-spatial load forecasting is suggested in this research. Micro-spatial load forecasting is a method used for the planning of distribution systems in a small area. This method is applied with several approaches: multivariate, time series, artificial neural network, and land usage [5], [16], [18], [19].

Multivariate analysis is a technique used to analyze an object that contains more than two variables [16], [20], [21]. This method involves examining load expansion within specific grids, which are determined based on causal factors associated with load growth, utilizing historical, and current data. In sectoral load forecasting, the generalization of the equation model may cause an error in an area with different characteristics [22]. Consequently, this study will employ a load forecasting method involving multivariate and clustering analysis. This approach requires the grouping (clustering) of areas with similar characteristics. Furthermore, the data grids will be organized into distinct groups based on their similarities within their respective sectors.

The previous research [16], [18], [19] has used the hierarchical method in grouping areas into clustering. Clustering techniques group the grid through a chart in the form of a hierarchy, where the two closest groups are merged in each iteration or division of the entire data set into clusters. However, this method has weaknesses, such as the researcher's subjectivity since it only sees the dendrogram picture, and it would be difficult for researchers to evaluate the number of groups created if the data is too heterogeneous between research objects [23]–[27]. Consequently, the number of clusters generated is often distorted due to the lack of testing of the clusters effects.

Therefore, the authors utilized partitional clustering to overcome that problem in this research. This algorithm is more efficient, accelerated in data processing, and suitable for circular data. The use of clustering algorithms in spatial load forecasting aims to reduce the number of calculated volumes from the calculated areas. For instance, if 118 regions need to be assessed, applying a clustering approach eliminates the necessity of creating 118 separate models for spatial load forecasting. Instead, implementing predefined clustering models reduces the number of required models. In the context of this research, 54 clusters have been formed to address the specific case. Hence, this method minimizes variations between data in one cluster and maximizes data variations between clusters. Besides that, partitioning clustering can cluster large data and handle outlier data, Producing the right cluster group that can represent all the calculated data (area). Then, several mathematical modelings are formed, which project the load growth of each cluster.

2. METHOD

2.1. Micro-spatial load forecasting flow using clustering technique

Sectoral load forecasting based on the multivariate analysis used in this research retrieved historical data of load density and other variables to predict sectoral load density in the later year. Clustering in each area is employed before determining the correct equation model using a micro-spatial load forecasting algorithm. The flow for the proposed methodology in this research is displayed in Figure 1.

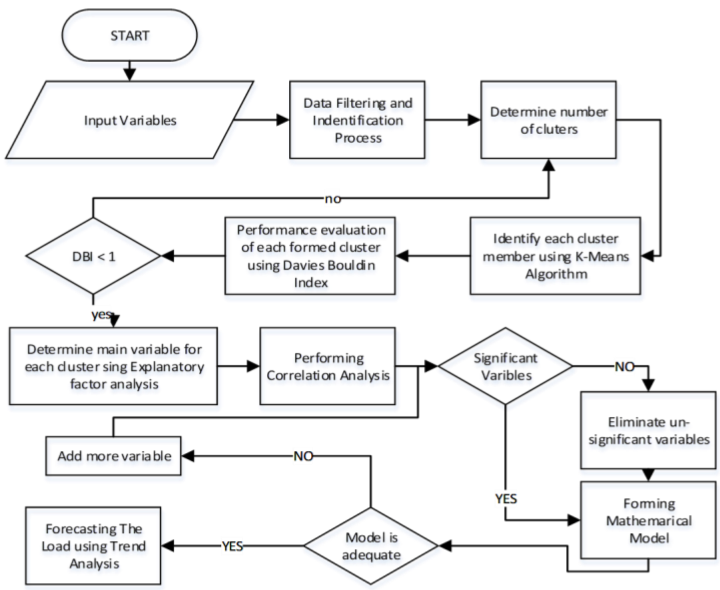

Figure 1. Flowchart for micro-spatial load forecasting based on partitional clustering technique

The proposed method begins with collecting and entering data on electrical variables (peak load of each sector) and non-electrical variables (such as number of households, area per sector and income per capita). Then K-means method is employed to define several clusters. To test and validate the result of the cluster is required the Davies Bouldin index (DBI) method, which internally evaluates to measure cluster evaluation based on separation and cohesion values. Cohesion is an amount of data closeness to the cluster centre of the cluster that is followed. Meanwhile, the distance between the cluster centre and the cluster is called separation [21], therefore, the best cluster results and maximize the quality of cluster results can be expected.

Each generated cluster will have different dominant variables, so exploratory factor analysis is required to determine the most dominant variable from several existing variables. Therefore, each cluster will reduce the number of variables from the previous number. These dominant variables will later be used to determine and form the mathematical model of each cluster. Where this mathematical model is then used as the basis for carrying out load forecasting, the results of which are then broken down according to the pattern of regional development in each forecasted region.

2.2. Partitional clustering

Clustering means the partition of a certain subset to group the data with similar characteristics into one cluster. The partition clustering algorithm creates different partitions and then classifies them based on the set of criteria. Because each instance is placed in one of k mutually exclusive clusters, they are also known as non-hierarchical. Because a typical partitional clustering method develops just one set of clusters, the user must specify the desired number of clusters. One of the most commonly used partitional clustering algorithms is the K-means clustering algorithm. K-Means clustering categorizes data into fixed groups [28], [29].

Identifying the attribute type of each variable is the first step in preparing to group. During this process, the data is divided into four data types: nominal, ordinal, interval, and ratio. Nominal data means the attribute data encompassed different symbols in a closed set, for instance, these labels (dry, wet, and humid). Ordinal data, also known as numerical data, is a data type that has continuous order, and the distances between data are not always equal. For instance, interval data has equal order and value range (1-2, 3-4, 5-6). Ratio data is a comparison between values. After determining the attribute of each variable, clustering is implemented by several steps [24], as follows:

a. Decide the number of desired clusters.
b. Initiate centroid in each group.
c. Assign the object to the closest group, the distance between the object and the centroid is defined by using Euclidean distance (d)

$$d_{euclidian}\,(x,y) = \sqrt{\sum_{i=1}^{F}(x_i - y_i)} \tag{1}$$

where, x_i : object value
y_i : centroid value
F : numbers of attributes

d. Calculate the new centroid using (2):

$$c = \frac{1}{N}\sum_{i=1}^{N}\sum_{e=1}^{F} d_{i[e]} \tag{2}$$

Where:
N : number of cluster members
F : number of attributes
d_i: member-i in a cluster
$d_{i[e]}$: Attribute e in data d_i

e. Reassign the object to the closest group based on the new centroid.

The K-Means clustering algorithm can be used to identify outliers in the data, to determine the number of clusters in the data, and to draw conclusions about the data. The cluster analysis helps to classify the data into groups. The groups can then be used to identify patterns and make predictions. Results from K-means clustering is represented in the table that shows the dataset assigned in certain clusters, which is used to make decisions or take action.

2.3. Davis bouldin index

One problem found in the clustering technique is that it needs to determine the number of clusters with the lowest variance. However, the lowest variance value comes at the expense of the greater number of clusters (variance=0 when the number of clusters – number of data sets). This means a method to justify the optimal number of designed clusters is needed, and thus, this research used DBI to solve this problem [24],

[25]. DBI calculates the sum of intra-cluster (within-cluster) variances divided by the distance between centroids for their closest neighbouring cluster as illustrated in Figure 2. The lowest value of the DBI indicates a sepa ration between clusters. The indication for good clustering is when the value of the DBI is less than 1.

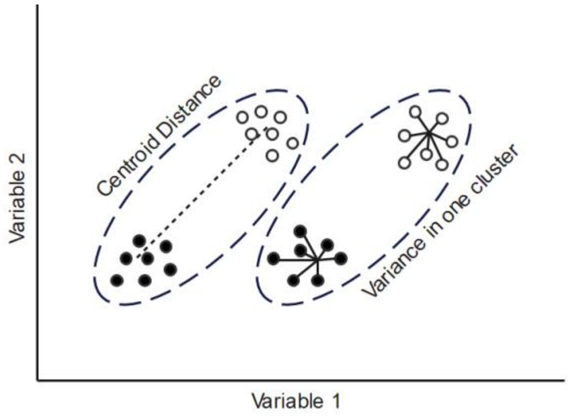

Figure 2. Illustration of variances and centroid distance between clusters

2.4. Principal component analysis

Load forecasting is a critical process in the domain of energy management and grid operations. It involves the development of a mathematical model that effectively captures the intricate relationship between dependent and independent variables, enabling grid operators and energy managers to make informed decisions. To create this mathematical model, one indispensable technique employed is principal component analysis (PCA), which is widely recognized for its capabilities in dimensionality reduction and data simplification [16], [20]. With the increasing complexity of data sets in modern energy systems, PCA comes to the forefront as a valuable tool for transforming large and correlated data sets into more manageable, smaller, and uncorrelated components. The PCA can be expressed as (3):

$$KU_1 = a_1 x = a_{11}x_1 + \cdots + a_{1p}x_p \tag{3}$$

a_1 is the eigenvector of the covariance matrix (∑), which represents the original variable with the largest eigenvalue.

The underlying concept of PCA is to reveal the hidden structure in complex data, which employs the covariance matrix of the principal components as a new covariance matrix. Each principal component captures a portion of the overall variance in the dataset. The covariance matrix of the principal components describes how these new variables (the principal components) are correlated with each other. It does so by constructing new variables, the principal components, that are linear combinations of the original data. These components are designed to be orthogonal, meaning they are uncorrelated, which simplifies the dataset while preserving the essential information. This approach is articulated through the PCA equation, which, in its essence, signifies the transformation of a high-dimensional dataset into a more succinct, uncorrelated, and interpretable representation. Therefore, the covariance matrix of the principal component can be formulated as (4):

$$\sigma^2_{KU1} = a'_1 \textstyle\sum a_1 \tag{4}$$

Further, the principal component value is obtained by a vector (a_2), which is the second-largest eigenvalue of the matrix ∑.

Once the principal component value is obtained through the PCA process, a mathematical model can be formed in a multiple linear regression as described in the (5) [23], [25]:

$$Y = b_1 + b_2X_2 + b_3X_3 + \cdots + b_kX_k + e \tag{5}$$

In this equation, Y is the dependent variable representing the predicted load, which is the value you are trying to forecast. b_1, b_2, b_3, and b_k are coefficients that represent the relationship between the dependent variable (load) and each of the independent variables (including the principal component value, X) in the model. X_2, X_3, and X_k denote the other independent variables used in the model, such as weather data, historical consumption, economic factors, and any other relevant parameters. Then, e represents the error term, which captures the difference between the predicted load and the actual observed load values.

By fitting this multiple linear regression model, load forecasters can incorporate the information derived from the principal component analysis, along with other relevant factors, to make more accurate and reliable load predictions. This approach is particularly valuable for understanding and quantifying the influence of various independent variables on load, which is essential in optimizing energy management, grid operations, and resource allocation. It ultimately contributes to more efficient and data-driven decision-making in the realm of load forecasting and energy management.

2.5. Sectoral electricity load forecasting

Load forecasting in multivariate analysis is a multifaceted endeavor encompassing an array of intricate processes and methodologies. Beyond relying on historical load data from previous years, it also hinges on comprehensive historical data about the myriad variables that influence load density. These variables can encompass an extensive range, including meteorological data, socio-economic factors, industrial activity, and even emerging trends in energy consumption patterns. The multifaceted nature of load forecasting is underpinned by the understanding that no single variable operates in isolation. Instead, the intricate interplay and interdependencies among these variables shape the overall load density.

To achieve this complex task, load forecasters employ various statistical techniques and predictive models driven by an exhaustive dataset that incorporates historical values of all relevant variables. One of the key aspects of this process is discerning the most suitable trend outcomes for each variable, often involving rigorous analysis and computation. These trends serve as the foundation for load density forecasting within specific clusters. Importantly, the forecasters evaluate the accuracy of these trends by employing metrics such as the mean absolute percentage error (MAPE), which quantifies the deviation between the forecasted and actual load density. The trend with the lowest MAPE is ultimately selected as it represents the most accurate projection, ensuring the highest level of forecasting precision.

Once these trends are ascertained, the next critical step involves estimating load density within each sector of the respective cluster. This sector-wise estimation is essential as different sectors can exhibit varying patterns of load behavior, each influenced by its unique set of variables. These sector-specific forecasts contribute to a more granular and precise load density projection for the entire cluster.

The culmination of this intricate process is calculating the total energy requirements for the districts within the cluster. This is achieved by summing up the projected energy consumption of each sector. Consequently, the load forecasters can provide a comprehensive view of energy demand for the entire district and detailed insights into how different sectors contribute to the overall energy profile. Such detailed forecasting is indispensable for utility companies, grid operators, and policymakers to make informed decisions regarding energy supply, resource allocation, and grid management. Ultimately, it enables more efficient and reliable energy distribution, fostering sustainability and resilience in energy systems.

3. RESULTS AND DISCUSSION

3.1. Data

The data taken for this research are electrical and non-electrical data. The data will be the independent and dependent variables. Independent variables consist of the number of households, regional gross domestic product (GRDP), area size, and loads by sectors (housing, industry, business, and social), while the load density of the districts represents the dependent variable. The object of this research was a total of 118 districts. Data preview for this research is distributed into three parameters, which are demographic and economic Table 1, geographical parameters Table 2, and electrical parameters Table 3.

Table 1. Demographic and economic parameters

District (grid)	∑ Household	GRDP growth (%)
Caringin	1545	4.87
Babakan	1138	4.59
Pala Sari	1881	10.94
Serdang Wetan	1939	9.57
Rancagong	2217	12.55
Legok	1640	5.26
Bojong Nangka	5149	14.01
Pasir Muncang	1569	8.01
Cikande	1883	9.92

Table 2. Geographical parameter

District (grid)	Land use (Ha)				Area (Ha)
	Residential	Industry	Commercial	Social	
Caringin	52.15	1,067.4	711.6	142.9	1974
Babakan	56.49	998.8	665.9	154.8	1876
Pala Sari	61.07	2,461.6	1,641.0	167.3	4331
Serdang Wetan	90.18	2,113.6	1,409.1	247.1	3860
Rancagong	78.14	2,814.5	1,876.3	214.1	4983
Legok	63.40	1,145.3	763.6	173.7	2146
Bojong Nangka	280.94	2,927.6	1,951.7	769.8	5930
Pasir Muncang	75.30	1,792.8	1,195.2	127.7	3191
Cikande	91.89	2,220.8	1,480.5	155.8	3949

Table 3. Electrical parameter

District (grid)	Load (kW)				Load density (kW/Ha)
	Residential	Industry	Commercial	Social	
Caringin	185.6	5192.3	1301.8	560.9	3.67
Babakan	201.1	4858.9	1218.2	607.6	3.67
Pala Sari	217.3	11974.5	3002.1	656.8	3.66
Serdang Wetan	320.9	10282.0	2577.8	969.8	3.67
Rancagong	278.1	13691.2	3432.5	840.4	3.66
Legok	225.6	5571.5	1396.8	681.9	3.67
Bojong Nangka	1000.0	14241.5	3570.4	3021.4	3.68
Pasir Muncang	268.05	8721.28	2186.48	501.2	3.66
Cikande	327.09	10803.19	2708.43	611.6	3.66

3.2. Clustering process

The clustering process was operated by RapidMinder software as depicted in Figure 3. The process began by determining the attribute of each variable. All data used were numerical; hence, the data were categorized into nominal data.

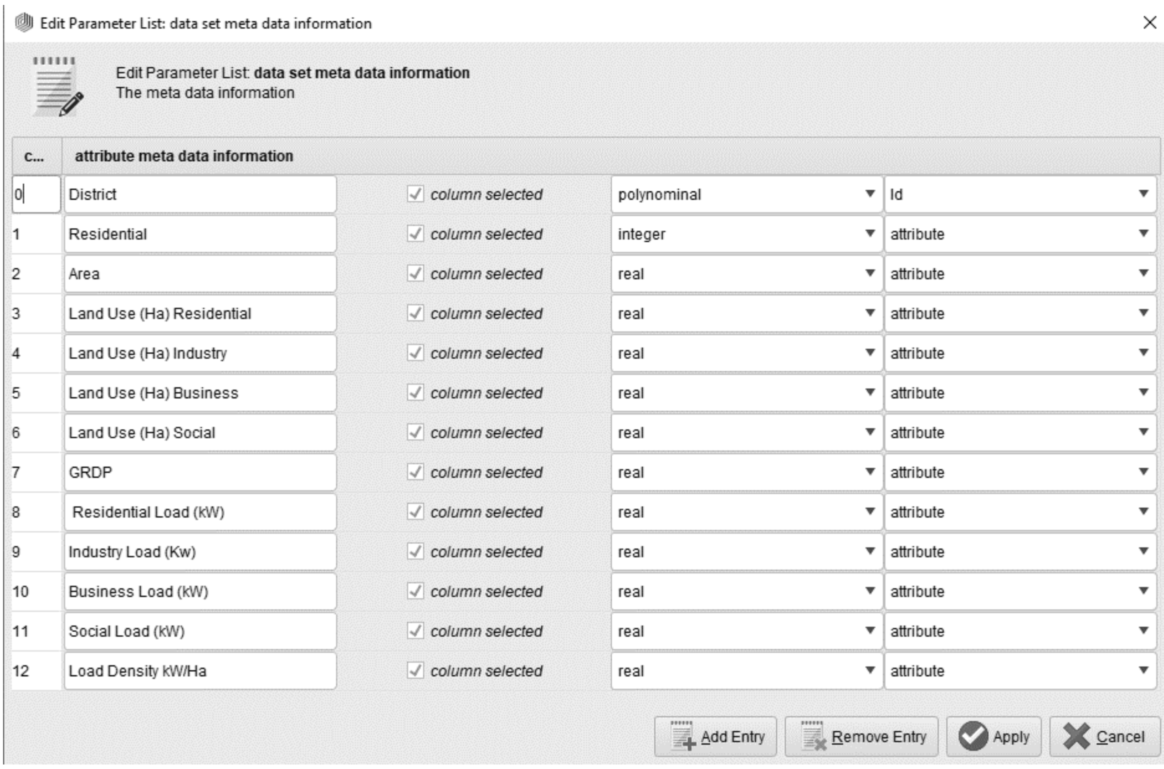

Figure 3. Determining the attribute of each variable

The block diagram was designed as depicted in Figure 4. The process began with data input in fixed attributes. The clustering method was decided, as proposed in this research, using K-means. The numbers of clusters were selected in the iteration process to ascertain proper clustering output. For this research, initiated clusters were 5 clusters.

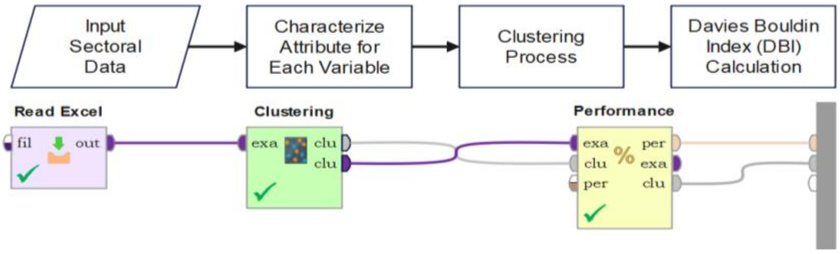

Figure 4. Determining the attribute of each variable

3.3. Forming cluster

The object of this research is the distribution networks encompassing 118 districts within Tangerang and certain regions of Jakarta. The clustering process was performed using the K-means method, and thus, 5 clusters were identified through the process. The purpose of clustering is to group districts with homogenous characteristics. Table 4 presents a preview of the data output from clustering districts in Tangerang and Jakarta, which demonstrates that the quantity of grids assigned to each cluster exhibits variation contingent upon the unique grid characteristic associated with each individual cluster. The distinctive attributes and qualities of each cluster were meticulously analyzed and subsequently interpreted into their corresponding centroids, which are thoroughly illustrated and elaborated upon in Table 5.

Table 4. Clustering output using k-means

Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4
Cengkareng Barat	Karawaci Baru	Pala Sari	Pabuaran	Kamal Muara
Semanan	Karawaci	Serdang Wetan	Budi Mulya	-
Kalideres	Cimone	Rancagong	Bojong	-
Pegadungan	B u g e l	Bojong Nangka	Suka Mulya	-
Tegal Alur	Pabuaran	Pasir Muncang	Cikupa	-

Table 5. Clustering output using k-means

Attribute	Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4
Residential	3811.926	15110.100	1876.829	2958.263	2083
Area	245.345	567.813	2533.771	4884.053	1053.400
Land use (Ha) residential	126.841	349.883	136.067	166.084	84.272
Land use (Ha) industry	21.616	68.106	1381.698	2721.647	183.292
Land use (Ha) business	35.145	71.337	921.132	1814.432	72.685
Land use (Ha) social	61.770	78.405	94.873	181.890	713.152
GEDP	34.206	209.181	6.184	11.640	99.964
Residential load (kW)	600.522	5617.761	484.334	591.179	1353.080
Industry Load (kW)	160.719	6696.874	6721.442	13239.787	18023.033
Business Load 9kW)	172.080	5074.409	1685.106	3319.294	5170.290
Social load (kW)	299.685	1592.991	372.378	713.917	14489.448
Load density (kW/Ha)	4.694	33.976	3.656	3.658	37.057

Regarding Table 5**Error! Reference source not found.**, the characteristic of each cluster is different d epending on the variable and dominant factor. Thus, it can be inferred that each cluster needs a different mathematical model in load forecasting. Based on the clustering process carried out by hierarchical clustering and partitional clustering, the significant differences are running time, assumptions, input parameters, and the resulting clusters. Partitional grouping tends to be faster than hierarchical grouping. During the clustering process, Hierarchical clustering only requires similarity value, whereas partitional clustering requires stronger assumptions such as the number of clusters and initial centroid. The result of hierarchical clustering is much more subjective than partitional clustering. Partitional clustering also delivers a better cluster set since testing is carried out after the clusters are formed.

3.4. Evaluation of clustering performance

The output of clustering needs to be evaluated using DBI. As tabulated in Table 6, the lower value of the DBI (less than 1) indicates the distance between the closest clusters compared with the distance between dissimilar clusters. The DBI value shows the validity of each cluster. Based on the DBI test, the optimum cluster for this data is 5 clusters since the average value of the DBI obtained was 0.5447, which means the clustering output is appropriate. The DBI result for clustering the data into 4, 5, and 6 clusters is shown in Table 6.

Table 6. DBI test

Number of cluster	Clustering	Number of grids	DBI test
2	Cluster_0	58	3.2 x 107
	Clsuter_1	61	
3	cluster_0	57	1.8 x 107
	cluster_1	10	
	cluster_2	52	
4	cluster_0	43	0.655
	cluster_1	13	
	cluster_2	9	
	cluster_3	54	
5	cluster_0	9	0.5447
	cluster_1	54	
	cluster_2	19	
	cluster_3	36	
	cluster_4	1	
6	cluster_0	53	0.756
	cluster_1	35	
	cluster_2	5	
	cluster_3	19	
	cluster_4	1	
	cluster_5	6	

Each cluster will have different regional characteristics, but each region within the cluster will have similar traits. The area is grouped into 5 clusters that show the area's distribution, as mapped in Figure 5. The map of clustering output depicts the region's overall cluster distribution, demonstrating that nearby locations tend to be in the same cluster. Table 7 shows the regional characteristics of a cluster – 1.

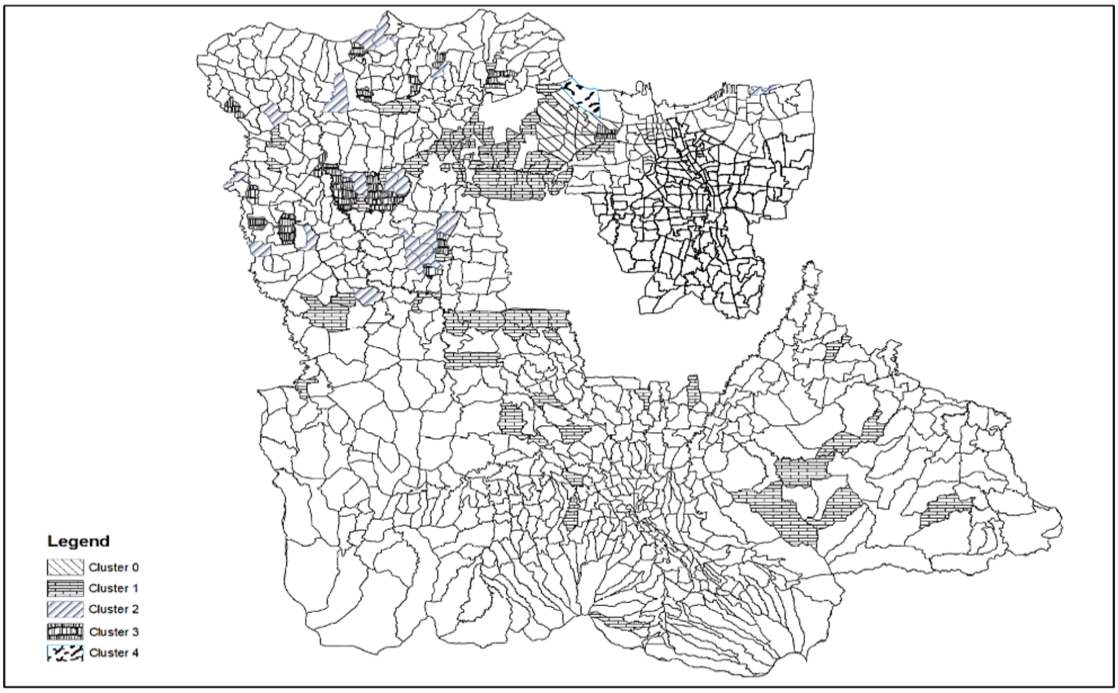

Figure 5. Clustering output of district

Table 7. Characteristics of cluster - 1

Variables	N	Minimum	Maximum	Mean	Std. deviation
		Descriptive statistics			
Residential (unit)	54	529	10744	3811.93	2374.365
Area (Ha)	54	58	1070	245.34	207.840
Residential land use (Ha)	54	20.56	460.40	1.2684	75.51688
Industry land use (Ha)	54	0.00	549.75	21.6158	75.72367
Business land use (Ha)	54	0.00	366.50	35.1452	67.99717
Social land use (Ha)	54	0.00	437.13	61.7703	101.41784
GRDP (Million/year)	54	2.55	190.97	34.2060	38.46717
Residential load (KW)	54	73.18	4286.07	6.0052	787.92362
Industry load (KW)	54	0.00	2721.99	1.6072	511.86931
Business load (KW)	54	0.00	3716.89	1.7208	569.81060
Social load (KW)	54	0.00	2151.51	2.9968	487.55458

3.5. Sectoral load forecasting

Based on the findings of the variable test using principal component analysis, modelling of all parameters that contribute to the first rating value (load density) is carried out at this final stage. Multiple regression modelling was used to obtain the result depicted in Figure 6. A regression equation, which represents each cluster based on historical data, was formed using the output of clustering. The regression equation was later set as the foundation of sectoral load forecasting for each sector in the districts. Finally, the result of sectoral load forecasting for the year 1-5 and the year 6-10 is expressed in Figures 7 and 8, respectively. These load growth forecasts can establish load points for future years based on geographical location, demographic characteristics, and load characteristics. The load growth of every grid (district) can be applied as the basis of a greater-scale load growth (region). Thus, the micro-spatial sectoral load forecast can be used as the basic ground of the distribution master plan.

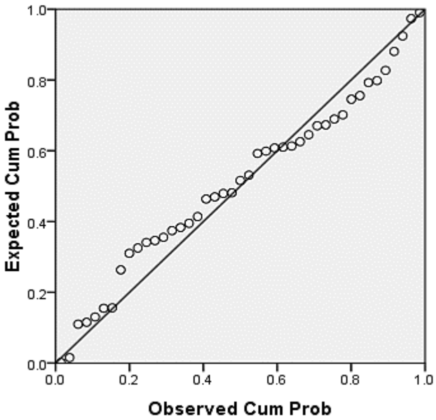

Figure 6. Block diagram of clustering process

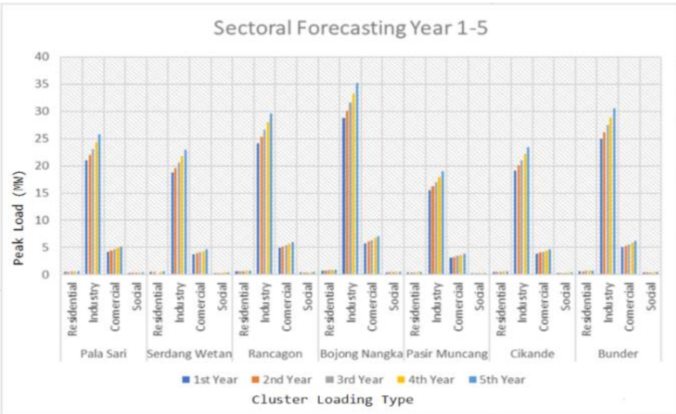

Figure 7. Sectoral load forecasting year 1-5

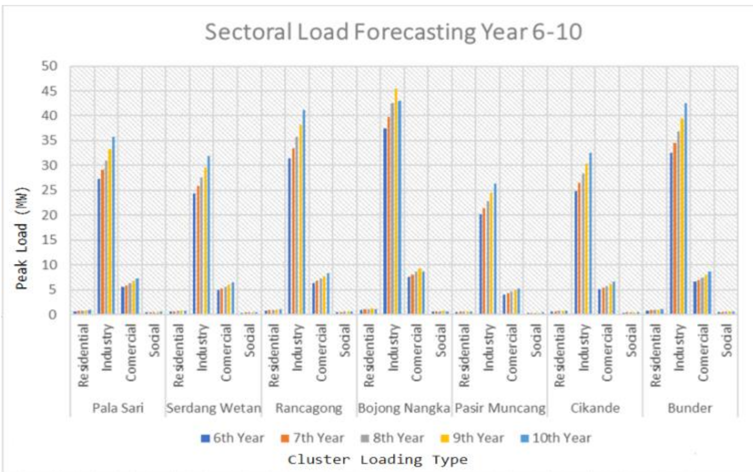

Figure 8. Sectoral load forecasting year 6-10

These graphics illustrate higher demand and load growth in the industrial sector. This is due to the naturally high annual growth in the industrial sector. For the residential sector, the growth trend pattern is nearly identical to the industrial sector. However, for the business and social sectors, there is a tendency for the growth pattern to be relatively small, primarily because the overall expansion of the business and social areas is not very significant. The average growth percentage in the industrial sector is highly dynamic when compared to the loads in other sectors. Nonetheless, the system's growth follows a linear pattern with an average growth rate of 6.4%.

3.6. Comparison study of the spatial load forecast strategies

Table 8 presents a brief overview of prior research conducted by various authors in the field of sectoral electricity micro-spatial load forecasting. These studies have employed diverse strategies aimed at identifying the optimal approach for electric load forecasting. The proposed methodology introduces a novel perspective by enabling implementation at the smallest spatial scale, such as individual grids or cells. This involves intricate multivariate calculations and classification using a soft-clustering technique. The outcomes of the clustering process are subsequently visualized and mapped using geographical information system (GIS) tools. Concurrently, linear regression is utilized to predict regional load requirements while maintaining adaptability to changes in land use patterns.

Table 8. A brief review of the previous sectoral electricity micro-spatial load forecasting

Ref.	Year	Grid / cell	Multivariate analysis	Mathematic model			Gis	Treading	Land usage simulation	Direct vision	Clustering	
				Time series	Multiple regresion	Gwr					Hard clustering	Soft clustering
[16]	2020	✓	×	✓	×	×	✓	×	×	×	×	✓
[19]	2020	✓	×	×	✓	×	×	×	✓	×	✓	×
[30]	2020	✓	✓	✓	×	×	×	×	✓	×	×	×
[5]	2021	✓	✓	×	✓	×	×	×	✓	×	✓	×
[31]	2021	✓	×	×	✓	×	×	×	✓	×	✓	×
[32]	2021	✓	✓	×	×	×	×	×	×	×	×	✓
[33]	2021	✓	×	✓	×	×	✓	×	✓	×	×	×
[34]	2022	✓	✓	×	×	✓	✓	×	×	✓	×	×
[35]	2021	✓	✓	×	×	×	✓	×	×	×	✓	×
[36]	2022	✓	×	×	×	×	✓	✓	×	×	×	×
Proposed Method		✓	✓	×	✓	×	✓	×	✓	×	×	✓

4. CONCLUSION

Micro-spatial load forecasting using the clustering technique proposed in this research identified grids with homogenous characteristics. A total of 119 grids were grouped into 5 clusters, followed by good performance. The output of clustering was taken as the basis of the load forecast for each cluster. Moreover, the output provided information on raising load in high accuracy, and hence appropriate for the basic ground of distribution of the master plan. The development of this methodology is possible in terms of pattern identification through artificial intelligence systems.

ACKNOWLEDGEMENTS

This research is funded by Research Funding for Lecturers of Perusahaan Listrik Negara (PLN) Institute of Technology, Jakarta.

REFERENCES

[1] Z. Yu, Z. Niu, W. Tang, and Q. Wu, "Deep learning for daily peak load forecasting-a novel gated recurrent neural network combining dynamic time warping," *IEEE Access*, vol. 7, pp. 17184–17194, 2019, doi: 10.1109/ACCESS.2019.2895604.

[2] R. D. Arias, F. P. -Yauli, and F. Q. Gudiño, "A methodological aproach to the determination of the future electrical spatial demand in electric distribution companies in Ecuador," in *2019 IEEE PES Conference on Innovative Smart Grid Technologies, ISGT Latin America 2019*, IEEE, Sep. 2019, doi: 10.1109/ISGT-LA.2019.8895274.

[3] D. Zhang, T. Ahmad, and W. A. Shah, "Efficient energy planning with decomposition-based evolutionary neural networks," *IEEE Access*, vol. 8, pp. 134880–134897, 2020, doi: 10.1109/ACCESS.2020.3010782.

[4] K. B. Lindberg, P. Seljom, H. Madsen, D. Fischer, and M. Korpås, "Long-term electricity load forecasting: Current and future trends," *Utilities Policy*, vol. 58, pp. 102–119, 2019, doi: 10.1016/j.jup.2019.04.001.

[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," *Pertanika Journal of Science and Technology*, vol. 29, no. 4, pp. 2565–2578, 2021, doi: 10.47836/PJST.29.4.18.

[6] O. E. Olabode, I. K. Okakwu, O. O. A. -Ikuesan, and I. D. Fajuke, "Performance evaluation of medium-term load forecasting approaches: A case study of Ogun State, Nigeria," *Journal of Advances in Science and Engineering*, vol. 1, no. 2, pp. 9–16, Sep. 2018, doi: 10.37121/jase.v1i2.24.

[7] S. Sasmono, N. I. Sinisuka, and M. W. Atmopawiro, "Alternative spatial approach on spatial demand forecasting for transmission expansion planning," in *Proceedings of 2012 IEEE International Conference on Condition Monitoring and Diagnosis, CMD 2012*, IEEE, Sep. 2012, pp. 577–580, doi: 10.1109/CMD.2012.6416209.

[8] S. Katruksa and S. Jiriwibhakorn, "Electricity load forecasting based on a geographic information system," in *Proceeding - 5th International Conference on Engineering, Applied Sciences and Technology, ICEAST 2019*, IEEE, 2019, doi: 10.1109/ICEAST.2019.8802591.

[9] D. E. Babatunde, A. Anozie, and J. Omoleye, "Artificial neural network and its applications in the energy sector – An overview," *International Journal of Energy Economics and Policy*, vol. 10, no. 2, pp. 250–264, 2020, doi: 10.32479/ijeep.8691.

[10] Sujito, R. R. Hadi, L. Gumilar, A. I. Syah, M. Z. Falah, and T. H. Duy, "Long-term forecasting for growth of electricity load based on customer sectors," *Journal of Mechatronics, Electrical Power, and Vehicular Technology*, vol. 13, no. 2, pp. 214–221, 2022, doi: 10.14203/j.mev.2022.v13.214-221.

[11] A. A. Mir, M. Alghassab, K. Ullah, Z. A. Khan, Y. Lu, and M. Imran, "A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons," *Sustainability*, vol. 12, no. 15, 2020, doi: 10.3390/SU12155931.

[12] B. Schminke, "Overview of the current state of research on characteristics and algorithms of energy management systems in households and buildings," *International Journal of Energy Research*, vol. 45, no. 10, pp. 14194–14206, 2021, doi: 10.1002/er.6738.

[13] T. Tumiran *et al.*, "Power system planning assessment for optimizing renewable energy integration in the maluku electricity system," *Sustainability*, vol. 14, no. 14, 2022, doi: 10.3390/su14148436.

[14] B. Yildiz, J. I. Bilbao, J. Dore, and A. B. Sproul, "Short-term forecasting of individual household electricity loads with investigating impact of data resolution and forecast horizon," *Renewable Energy and Environmental Sustainability*, vol. 3, 2018, doi: 10.1051/rees/2018003.

[15] W. Lin, D. Wu, and B. Boulet, "Spatialoral residential short-term load forecasting via graph neural networks," *IEEE Trans Smart Grid*, vol. 12, no. 6, pp. 5373–5384, Nov. 2021, doi: 10.1109/TSG.2021.3093515.

[16] X. Bai *et al.*, "A spatial load forecasting method based on the theory of clustering analysis," *Phys Procedia*, vol. 24, pp. 176–183, 2012, doi: 10.1016/j.phpro.2012.02.027.

[17] M. Q. Raza, N. Mithulananthan, J. Li, and K. Y. Lee, "Multivariate ensemble forecast framework for demand prediction of anomalous days," *IEEE Trans Sustain Energy*, vol. 11, no. 1, pp. 27–36, 2020, doi: 10.1109/TSTE.2018.2883393.

[18] J. Jiménez, A. Pertuz, C. G. Quintero, and J. Montaña, "Multivariate statistical analysis based methodology for long-term demand forecasting," *IEEE Latin America Transactions*, vol. 17, no. 1, pp. 93–101, 2019, doi: 10.1109/TLA.2019.8826700.

[19] C. Widyastuti, A. Senen, and O. Handayani, "Micro-spatial electricity load forecasting using clustering technique," *Proceeding - 2020 2nd International Conference on Industrial Electrical and Electronics, ICIEE 2020*, no. 2, pp. 17–21, 2020, doi: 10.1109/ICIEE49813.2020.9277274.

[20] B. J. D. Sitompul, O. S. Sitompul, and P. Sihombing, "Enhancement clustering evaluation result of davies-bouldin index with determining initial centroid of k-means algorithm," *Journal of Physics: Conference Series*, vol. 1235, no. 1, 2019, doi: 10.1088/1742-6596/1235/1/012015.

[21] D. Xishuang, Q. Lijun, and H. Lei, "Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach," *2017 IEEE International Conference on Big Data and Smart Computing, BigComp 2017*, pp. 119–125, 2017, doi: 10.1109/BIGCOMP.2017.7881726.

[22] A. Azeem, I. Ismail, S. M. Jameel, F. Romlie, K. U. Danyaro, and S. Shukla, "Deterioration of electrical load forecasting models in a smart grid environment," *Sensors*, vol. 22, no. 12, Jun. 2022, doi: 10.3390/s22124363.

[23] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 911–916, 2010, doi: 10.1109/ICDM.2010.35.

[24] F. Fahiman, S. M. Erfani, S. Rajasegarar, M. Palaniswami, and C. Leckie, "Improving load forecasting based on deep learning and K-shape clustering," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017, pp. 4134–4141, 2017, doi: 10.1109/IJCNN.2017.7966378.

[25] K. Gajowniczek and T. Zabkowski, "Simulation study on clustering approaches for short-term electricity forecasting," *Complexity*, vol. 2018, no. 3, 2018, doi: 10.1155/2018/3683969.

[26] X. Dong, S. Deng, and D. Wang, "A short-term power load forecasting method based on k-means and SVM," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 11, pp. 5253–5267, 2022, doi: 10.1007/s12652-021-03444-x.

[27] A. Almalaq and G. Edwards, "A review of deep learning methods applied on load forecasting," *16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, vol. 2017, pp. 511–516, 2017, doi: 10.1109/ICMLA.2017.0-110.

[28] W. Yanbo, L. Li, P. Xinfu, and F. Enpeng, "Load forecasting based on improved K-means clustering algorithm," *China International Conference on Electricity Distribution, CICED*, pp. 2751–2755, 2018, doi: 10.1109/CICED.2018.8592023.

[29] C. Ye, Y. Ding, P. Wang, and Z. Lin, "A data-driven bottom-up approach for spatial and temporal electric load forecasting," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1966–1979, 2019, doi: 10.1109/TPWRS.2018.2889995.

[30] D. A. G. Vieira, B. E. Silva, T. V. Menezes, and A. C. Lisboa, "Large scale spatial electric load forecasting framework based on spatial convolution," *International Journal of Electrical Power and Energy Systems*, vol. 117, 2020, doi: 10.1016/j.ijepes.2019.105582.

[31] C. Zhang and R. Li, "A novel closed-loop clustering algorithm for hierarchical load forecasting," *IEEE Trans Smart Grid*, vol. 12, no. 1, pp. 432–441, 2021, doi: 10.1109/TSG.2020.3015000.

[32] 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 and System Graduate Research Colloquium, ICSGRC 2021*, pp. 277–280, 2021, doi: 10.1109/ICSGRC53186.2021.9515295.

[33] R. Zhao, L. Zhan, M. Yao, and L. Yang, "A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2.5," *Sustainable Cities and Society*, vol. 56, pp. 1–9, 2020, doi: 10.1016/j.scs.2020.102106.

[34] Y. Gao, J. Zhao, and L. Han, "Exploring the spatial heterogeneity of urban heat island effect and its relationship to block morphology with the geographically weighted regression model," *Sustainable Cities and Society*, vol. 76, 2022, doi: 10.1016/j.scs.2021.103431.

[35] C. Zhu, Z. S. Hu, X. R. Wang, and L. Wang, "A hierarchical data driven method for spatial electric load forecasting," *Proceedings - 2021 International Conference on Power System Technology: Carbon Neutrality and New Type of Power System, POWERCON 2021*, pp. 360–364, 2021, doi: 10.1109/POWERCON53785.2021.9697499.

[36] V. A. Evangelopoulos and P. S. Georgilakis, "Probabilistic spatial load forecasting for assessing the impact of electric load growth in power distribution networks," *Electric Power Systems Research*, vol. 207, 2022, doi: 10.1016/j.epsr.2022.107847.

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 Ph.D. student in Faculty 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.

Jasrul Jamani Bin 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. He 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.

Hasna Satya Dini received the Bachelor and Master Degree in Electrical Engineering from Institut Teknologi Bandung, Indonesia in 2016 and 2017. She is currently pursuing a Ph.D. student in School of Electrical Engineering in Universiti Teknologi Malaysia. Her research interests mainly focus on resiliency, optimization, and power system stability. She can be contacted at email: hasna@itpln.ac.id.

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.