

# Clusterization of customer energy usage to detect power shrinkage in an effort to increase the efficiency of electric energy consumption

Yessy Asri, Dwina Kuswardani, Efy Yosrita, dan Ferdinand Hendrik Wullur

Department of Telematika Energi, Institute Technology PLN, Jakarta, Indonesia

---

## Article Info

### Article history:

Received Jun 1, 2020

Revised Aug 10, 2020

Accepted Sep 25, 2020

---

### Keywords:

Automatic meter reading

Clustering

Davies-bouldin index

Electricity

K-means

Losses

---

## ABSTRACT

Automatic meter reading (AMR) is a reading system result the measurement of electrical energy consumer, both locally and remotely. The problems faced is the high non-technical shrinkage of AMR customers due to installation, maintenance errors as well as dishonest actions some consumers, this has a major influence on electrical power losses. PT. PLN Disjaya currently faces difficulties having to choose which customers should be checked first, so the field can only find a little damage. The K-means method based on historical electric power usage and determine the most optimal number of groups the davies-bouldin index (DBI) method. Based on the results of testing with 2-6 sets of clusters, the cluster set results are the most optimal is set cluster 4 because it has the smallest DBI value 0.893. The set of 4 clusters has the best performance in data grouping of historical power usage of AMR customers the business class, each centroid of each cluster is used as an attribute and value of the AMR customer power usage business chart. The testing phase is customers who categorized as customers with un-normal usage electricity power. The test is, by determining the distance data testing each centroid in the cluster 4 set.

*This is an open access article under the [CC BY-SA](#) license.*



---

## Corresponding Author:

Yessy Asri

Department of Telematika Energi

Institute Technology PLN, Jakarta, Indonesia

---

## 1. INTRODUCTION

Losses or loss of electrical power in electric power distribution systems which is usually used at certain times, is one measure that is efficient or not an electric power system operation [1-6]. Currently it has been applied the method of measuring electrical energy using the AMR (automatic meter reading) system, namely, a system of reading or retrieving data from the measurement of electrical energy on consumers locally and remotely, where the reading schedule can be determined as needed. This AMR system can be used optimally for account issuance, customer load analysis, calculation of losses or distribution losses, electricity network development planning.

According to the results of the interview with one of the employees in the sector of distribution PT. PLN (Persero) distribution of Jakarta Raya, the AMR (automatic meter reading) system that has been applied can detect losses (electrical power losses) both technical and non-technical. Currently to detect losses itself is still checked directly by reading and checking every incoming customer data. One of the problems faced by PLN is the high non-technical shrinkage of potential AMR customers due to faulty installation and maintenance as well as one of the contributors to high losses is the fraudulent acts committed by some consumers. The research method used in this study is a qualitative approach using the case study method, which analyzes and calculates directly [1-3, 7-10]. The causes of energy losses that can be identified through AMR data analysis

include; meter reading errors, measurement abnormalities and errors during wiring, which is a mistake based on non-technical losses [3, 7, 8].

Data analysis and evaluation of energy use using the download data of AMR customers owned by PT. PLN (Persero) distribution Jakarta Raya. The data will be processed and grouped based on the usage pattern of its power. This study uses the K-means clustering data mining method. The K-means method is the oldest and most widely used clustering algorithm in a variety of small to medium applications because of its ease of implementation and fast time complexity [7, 8, 11-19]. Apart from that K-means also has a fairly high accuracy on the size of objects, so this algorithm is relatively more measurable and efficient for processing large amounts of objects [20]. The weaknesses in the K-Means algorithm are analyzing and determining the best number of K in clustering data in a dataset [21-25]. To get the optimal K value, the author uses the Davies-Bouldin index (DBI) method. Davies-Bouldin index measures the average similarity between each cluster and one of the most similar. The minimum DB Index value is the most optimum clustering scheme [16-18].

It is expected that the results of clustering on the historical data electrical power usage of AMR customers of PT. PLN (Persero) distribution of Jakarta Raya with the K-Means method produces optimal cluster results in order to find a group of customer power usage patterns, so that it can be used as a reference in analyzing each usage chart from historical data on customer electrical power usage and detect non-technical losses from customer.

**2. RESEARCH METHOD**

Research that has been done before is related to PT. PLN in 2017 with the title research "Development of Multimedia-Based Community Electricity Model" and in 2019 with a research theme on processing customer historical data of PT. PLN for monitoring customer data Pasca Bayar, especially customer data in arrears in order to simplify field officers in the process of installing new electricity or in the billing process for customers who are in arrears for one month, two, or three months in arrears.

The next research in the same year 2019 was about electricity customers PLN AMR, for automation clusterization of historical power utilization customers to detect power losses in an effort to increase the efficiency of electric energy consumption. The data used in this study is the historical data of customer power usage of AMR (automatic meter reading), electricity tariffs for business purposes, B-3 tariff rates or medium voltage (power limits above 200kVA) at PT. PLN (Persero) distribution Jakarta Raya. While the attributes that will be used are Power based on Peak Load Time (WBP, hours 18:00-22:00) and Power based on outdoor peak load time (LWBP, 22:00-18:00) for 7 days.

- a) Data training is used by 103 customers belonging to customers with normal power electricity usage.
- b) Data testing is used by 3 customers who are classified as un-normal power electricity usage (Non-technical losses).

Figure 1 Describes the research methods which have 4 steps, namely entry, process, exit and the final step is validation. Figure 2 describes the steps of preprocessing data to obtain the closest cluster. This stage is the majority work in data mining. Where this stage is the process of understanding data and initial data processing which will later be used in the calculation process with predetermined methods. The data to be processed is a load profile data on the power usage of each AMR customer for business customers of the B3/TM tariff group. The author takes 5 (five) data samples that have passed the pre-processing stage to be used as an example in the calculation process.

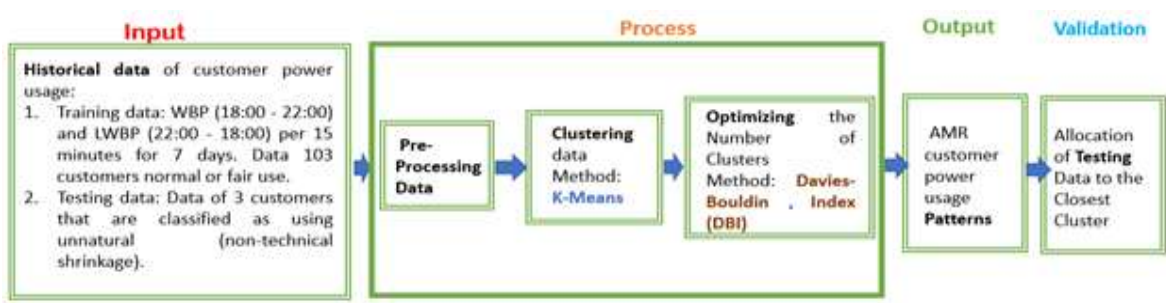


Figure 1. Research Method

Next is the calculation using the K-means method to determine the group pattern of automatic meter reading (AMR) customer power usage business class. The calculation stages will be explained based on the

flowchart process below. Figure 3 is the working process of the K-Means method process. The following is an example of the calculation process for centroid grouping, going directly to the last iteration stage, the data can be seen in Table 1. After doing K-means calculations and getting a centroid grouping, the following is an example of the data to be tested as shown in Table 2.



Figure 2. Design Process

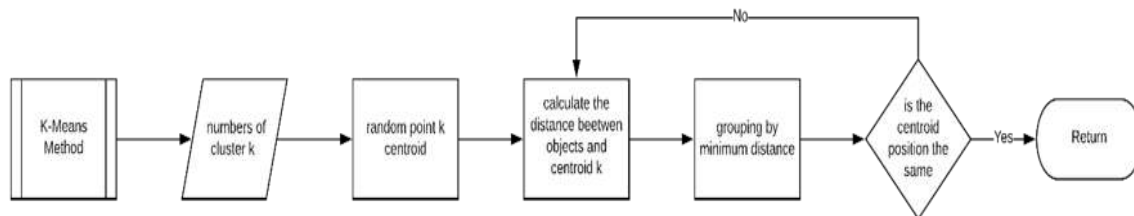


Figure 3. K-Means Process

Table 1. Centroid set 2 cluster final

	A	B	C	D	E	F	G
C1	0,266	0,250	0,284	0,252	0,279	0,247	0,273
C2	0,630	0,476	1,213	1,220	1,244	1,250	1,151
	H	I	J	K	L	M	N
C1	0,247	0,276	0,246	0,267	0,247	0,266	0,255
C2	1,249	1,227	1,231	1,197	1,304	0,635	0,808

Table 2. Examples of Data Testing

Customer	A	B	C	D	E	F	G
1	0,374	0,342	0,398	0,347	0,402	0,347	0,391
Customer	H	I	J	K	L	M	N
1	0,344	0,400	0,352	0,390	0,352	0,376	0,363

Calculate the maximum distance value (max DC) in each cluster that is formed (For example, the optimal cluster set = 2, based on the previous stage) the data can be seen in Table 3. Calculate the DC of each training data object to centroid, using the Euclidean formula, the data can be seen in Table 4.

Compare the value of DC data testing with each max DC in each cluster that has been calculated. The results of the appeal can be categorized as follows:

- a) If all DC data testing  $\leq$  max (DC) per cluster, then the data is classified as Normal in use.
- b) If one of the DC data testing  $>$  max (DC) per cluster, then the data is classified as Normal in use.
- c) If all DC data testing  $>$  max (DC) per cluster, then the data is classified as abnormal data in use.

DC1 Testing: DC1 Training = 0,409 < 0,418,

DC2 Testing: DC2 Training = 2,776 > 0,

Table 3. Maximum DC in each cluster

Customer	DC1	DC2
1	0,418	
2	0,148	
3	0,292	
4		0
5	0,273	
MAX DC	0,418	0

Table 4. DC value

DC1	DC2
0,409	2,776

Based on the results of the above comparison, it can be concluded, that data testing is classified as normal data in power usage. From the calculation process described above, the sample used is 5 AMR business class customers (B-3) by grouping using cluster set 2 and using 14 attributes in which 14 attributes are divided into 2, 7 of which are LWBP power usage data for 7 7 days and again is the data of WBP power usage for 7 days, in its calculations get results, namely:

- a) There are 2 patterns formed from the results of 5 customer data clustering. The power usage pattern is formed from the final centroid value obtained after the result of the iteration ends.
- b) In the first cluster there are 4 customers, while for the second cluster there are only 1 customer. That is, of the five customers there are 4 customers who have similarities in the pattern of power usage and 1 other customer has a different pattern from the 4 customers.
- c) For the optimal level of the set of 2 clusters for the 5 customers based on the calculation obtained a value of 0.095 based on calculations using the davies-bouldin index formula.

After the pattern is obtained, a testing test is performed to later make a comparison whether the testing data is entered or shaped like the pattern obtained or not. If the results of the testing data do not have a similarity with the pattern, in this case the distance to the centroid of each cluster that is obtained exceeds the maximum distance of the members in the cluster, then the customer testing data is unfair in the use of electric power and is set as the operating target.

### 3. RESULTS AND ANALYSIS

In this study, the power load profile data is not grouped based on certain criteria, but it is grouped generally based on the pattern of power usage. Because the number of clusters to be used is unknown, the author decides to use several clusters for the K-Means method grouping test. The cluster set used is from 2 cluster sets to 6 cluster sets. After calculating using the K-Means method, the optimization level of each cluster is then calculated using the davies-bouldin index (DBI) method. Following are the results of calculations on the last iteration, the data can be seen in Table 5 until Table 12.

Table 5. The number of members of each iteration set cluster 6

Iteration	Number of Members
1	Cluster 1: 5 Member Cluster 2: 55 Member Cluster 3: 31 Member Cluster 4: 2 Member Cluster 5: 9 Member Cluster 6: 1 Member
2	Cluster 1: 4 Member Cluster 2: 51 Member Cluster 3: 35 Member Cluster 4: 1 Member Cluster 5: 10 Member Cluster 6: 2 Member

Table 6. The end centroid value of cluster 1 in cluster set 6

A	B	C	D	E	F	G
0,4	0,4	0,5	0,4	0,5	0,4	0,4
07	93	48	39	13	91	79
H	I	J	K	L	M	N
0,4	0,5	0,4	0,3	0,3	0,2	0,2
18	38	33	99	73	67	88

Table 7. The end centroid value of cluster 2 in cluster set 6

A	B	C	D	E	F	G
0,062	0,064	0,076	0,104	0,075	0,111	0,074
H	I	J	K	L	M	N
0,108	0,074	0,109	0,081	0,112	0,072	0,068

Table 8. The end centroid value of cluster 3 in cluster set 6

A	B	C	D	E	F	G
0,155	0,148	0,188	0,197	0,19	0,203	0,188
H	I	J	K	L	M	N
0,202	0,191	0,202	0,191	0,206	0,167	0,156

Table 9. The end centroid value of cluster 4 in cluster set 6

A	B	C	D	E	F	G
0,641	0,788	1,608	1,329	1,635	1,558	1,690
H	I	J	K	L	M	N
1,633	1,653	1,671	1,729	1,640	0,719	1,351

Table 10. The end centroid value of cluster 5 in cluster set 6

A	B	C	D	E	F	G
0,333	0,274	0,348	0,277	0,35	0,279	0,337
H	I	J	K	L	M	N
0,277	0,348	0,276	0,332	0,283	0,328	0,293

Table 11. The end centroid value of cluster 6 in cluster set 6

A	B	C	D	E	F	G
0,359	0,480	1,043	1,125	1,045	1,219	1,020
H	I	J	K	L	M	N
1,200	1,050	1,200	1,034	1,269	0,434	0,646

Table 12. The results of the DBI value calculation for each set of clusters

Set Cluster	DBI	Number of Members
2	1.234	Cluster 1: 15 Cluster 2: 88
3	0.931	Cluster 1: 4 Cluster 2: 58 Cluster 3: 41
4	0.893	Cluster 1: 12 Cluster 2: 54 Cluster 3: 34 Cluster 4: 3
5	1.174	Cluster 1: 4 Cluster 2: 51 Cluster 3: 35 Cluster 4: 3 Cluster 5: 40
6	0.990	Cluster 1: 4 Cluster 2: 51 Cluster 3: 35 Cluster 4: 1 Cluster 5: 10 Cluster 6: 2

Based on the calculation results, the cluster set is the most optimal is set cluster 4 because it has the smallest DBI value, that is,  $0.893$ , that means set cluster 4 has the density of each object with the best centroid and the distance between the clusters is also well separated. Because the set of 4 clusters has the best performance in data grouping of historical power usage of AMR (automatic meter reading) customers in the business class, each centroid or center point of each cluster is used as an attribute and value of the AMR customer power usage business chart in PT. PLN (Persero) distribution of Jakarta Raya.

After getting the optimal set of clusters, then the testing phase. At this step the data being tested is data of 3 customers categorized as customers with un-normal usage electricity power. The test is, by determining the distance of each data testing object to each centroid in the cluster 4 set then the 3 data are tested with the output that is, the 3rd data is not normal in electricity power usage. The test is, by determining the distance of each data testing object to each centroid in the cluster 4 set then the 3 data are tested with the output that is, all 3 data are classified into customers with unnatural usage because based on the data allocation process to the centroid set the closest cluster 4, the distance of the testing data exceeds the maximum distance of each object in each cluster in the cluster 3 set. It can be seen the comparison of normal and un-normal electrical energy usage based on comparison of chart shapes between Figure 4 and Figure 5. The energy usage pattern in Figure 5 does not have the same pattern or according to each pattern in Figure 4. Besides being seen from the DC value of each client, it can also be compared between the WBP and LWBP values in each cluster as shown in Figures 4, 5, and 7, and the WBP and LWBP values of each client as shown in Figures 6 and 8.

Customers who are said not to fit the pattern or do not enter into any cluster, can be seen in the DC value of each customer, which has also determined the maximum DC value for each pattern, if the customer's DC value exceeds the maximum DC value, then it can be said that the customer is not appropriate predetermined pattern. Besides looking at the DC value of each customer, it can also be compared between the WBP and LWBP values in each cluster and the WBP and LWBP values for each customer. The highest value

of week WBP in each cluster, has the highest value of 0.453 in cluster 4, while for customers, the week's WBP value is 2.770, which has already passed the highest value of each cluster.

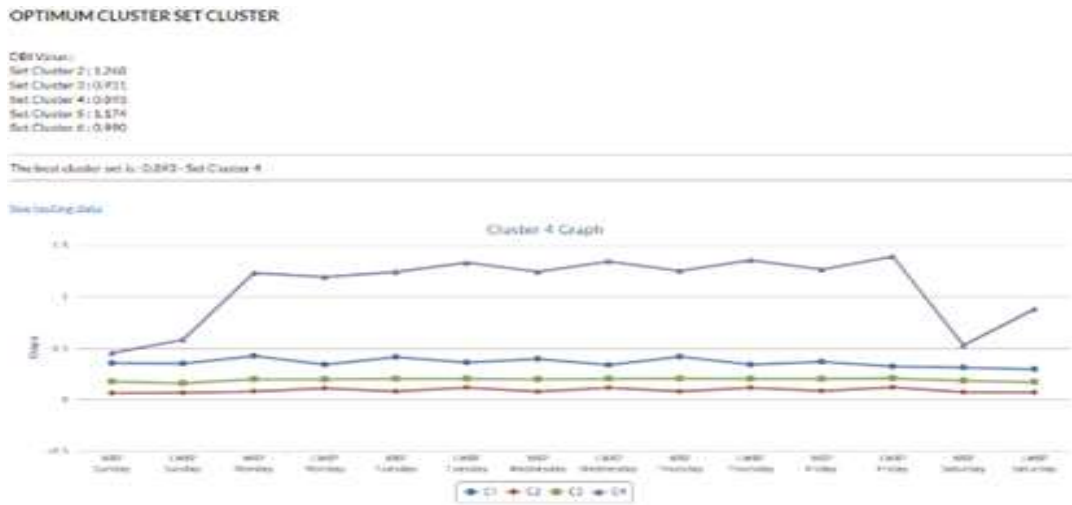


Figure 4. Chart of AMR Customer Business Class User Power Pattern Group



Figure 5. Chart of Energy Consumption Customers are

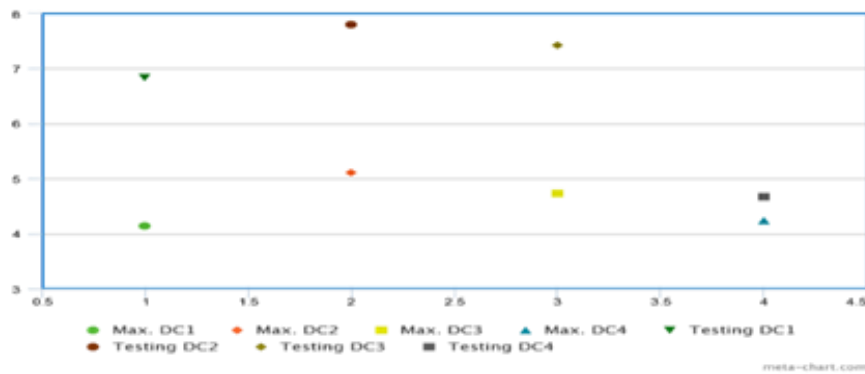


Figure 6. Scatter Plot Diagram of Consumer Testing Data



Figure 7. Normal Customer Power Usage Patterns

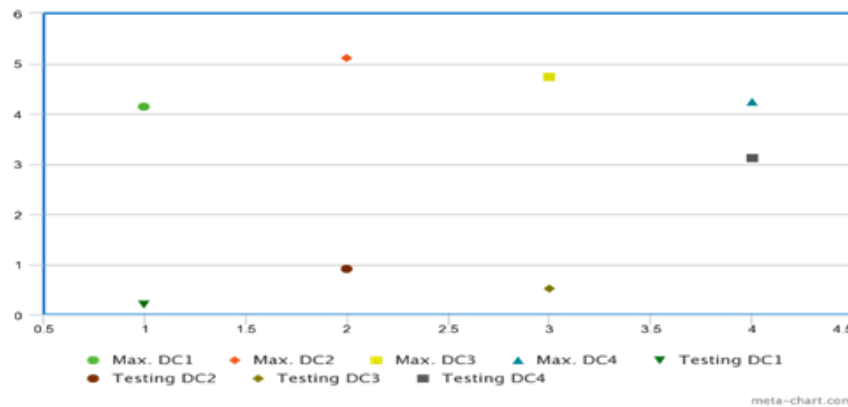


Figure 8. Scatter Plot Diagram of Consumer Testing Data

#### 4. CONCLUSION

Based on the description of the discussion above, clustering of historical data on customer power usage of AMR (automatic meter reading) business class of PT. PLN (Persero) distribution of Jakarta Raya has been successfully built and can be used to classify and determine the pattern of electric power usage based on the number of cluster sets that have been calculated for their optimization. Based on the results of testing with 2-6 sets of clusters, the cluster set results are the most optimal that is in the set of 4 clusters because it has the smallest DBI value (minimum), 0,893. Where in set 4 clusters, cluster 1 has 12 members, cluster 2: 54 members, cluster 3: 34 members and cluster 4: 3 members. There are 4 patterns of electric power usage of AMR (automatic meter reading) customers of business class (B-3 or medium voltage class, power limit above 200kVA). Customers who are said not to fit the pattern or do not enter into any cluster, can be seen in the DC value of each customer, which has also determined the maximum DC value for each pattern, if the customer's DC value exceeds the maximum DC value, then it can be said that the customer is not appropriate predetermined pattern. Besides looking at the DC value of each customer, it can also be compared between the WBP and LWBP values in each cluster and the WBP and LWBP values for each customer. The highest value of week WBP in each cluster, has the highest value of 0.453 in cluster 4, while for customers, the week's WBP value is 2.770, which has already passed the highest value of each cluster.

#### ACKNOWLEDGEMENTS

We would like to thank the LPPM Institute of Technology PLN, Jakarta, Indonesia, which helped fund this research.

## REFERENCES

- [1] S. S. Bhatti, E. M. Umair, U. Lodhi, and S. Haq, "Electric Power transmission and distribution losses overview and minimization in Pakistan," *Int. J. Sci. Eng. Res.*, vol. 6, no. 4, pp. 1108-1112, 2015.
- [2] B. Pradana, P. Purba, and E. Warman, "Analysis calculation of technical loss with load curve approach on the distribution network of PT. PLN (persero) Rayon Medan City," *Singuda Ensikom*, vol. 6, no. 2, pp. 60-64, 2014.
- [3] P. A. Yunyansyah, U. Wibawa, and T. Utomo, "Study of Estimated technical loss and alternative repairs in Kayoman Feeders at Sukorejo Substation", *College Student Journal TUEB*, vol. 3, no. 1, pp. 1-8, 2015.
- [4] A. Heriyanto, "A case study of AMR (Automatic Meter Reading) Performance at potential customer power 41.5 KVA-200 KVA in Situbondo," *Journal of Electrical Engineering, Muhammadiyah Jember University*, pp. 1-13, 2016.
- [5] E. Agustina and A. F. Amalia, "Reduction of non-technical losses in the distribution network through the automatic meter reading system in PT. PLN (PERSERO)," *J. Mechanical Engineering*, vol. 5, pp. 164-166, 2017, doi: 10.22441/jtm.v5i4.1223.
- [6] N. Putu, E. Merliana, and A. J. Santoso, "Analysis of the determination of the best number of clusters with the K-Means method," *Proceedings of the national multidisciplinary seminar and call for papers by UNISBANK (SENDI\_U)*, pp. 978-979.
- [7] A. Muhammad, H. Tumaliang, and S. Silimang, "Electrical energy loss analysis on distribution network (JTM) at PT. PLN (Persero) Gorontalo Area," *Journal of Electrical and Computer Engineering*, vol. 7, no. 3, pp. 295-302, 2018.
- [8] J. Mangundap, Silimang, Sartje, and H. Tumaliang, "Power Loss analysis distribution network at PT. PLN (Persero) Manado area 2017", *Journal of Electrical and Computer Engineering*, vol. 7, no. 3, pp. 219-226, 2018.
- [9] Rahman, F., Basuki, A., Aknuranda, "Moving Data retrieval on automatic meter reading with mesh topology", *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 6, no. 1, pp. 1-8, 2019, doi: 10.25126/jtiik.2019616959.
- [10] M. Samarth Pandit, M. Snehamandhre, and M. Meghananichal, "Smart energy meter using Internet of Things (IoT)," *VJER-Vishwakarma Journal of Engineering Research*, vol. 1, no. 2, pp. 222-229, 2017.
- [11] J. Suryanto, "Comparative analysis of 15 daily rainfall grouping in DIY Province using fuzzy clustering and K-Means clustering", *J. AGRIFOR*, vol. 16, no. 2, pp. 229-242, 2017.
- [12] F. E. Nasution and J. M. T., "The influence of industrial customer power factor on losses on the secondary side network of distribution transformers PT. PLN (PERSERO) Serpong area," *National Seminar of Technology, Faculty of Engineering, Krisnadwipayana Univ.*, pp. 148-163, 2019.
- [13] E. Kuswanto and Y. K. Suprpto, "Application of the K-Means algorithm with the optimization of the number of clusters for the grouping of the workforce of the province of East Java", *JAVA J. Electr. Electron. Eng.*, 2015.
- [14] Al qoyyimi, T. A., Penangsang, O., & Aryani, N. K.. "Determination of the location of short-circuit interference in the 20 kV distribution network of Tegalsari Surabaya feeders with the GIS-based impedance method (Geographic Information System)", *ITS Engineering Journal*, 6(1), doi: 10.12962/j23373539.v6i1.21297
- [15] A F Khairati, A A Adlina, G F Hertono, B D Handari, "Validity index study on K-Means enhanced algorithm and K-Mean MMCA in Bahasa," *PRISMA*, vol. 2, pp. 161-170, 2019.
- [16] Junwei Xiao, Jianfeng Lu, and Xiangyu Li, "Davies Bouldin index based hierarchical initialization K-Means," *Intelligent Data Analysis*, vol. 21, no 6, pp. 1327-1338, 2017, doi: 10.3233/IDA-163129.
- [17] Sitompul, Bernad Jumadi Dehotman, "Improvement of the evaluation results of the Davies-Bouldin index clustering by determining the initial cluster center point of the K-Means algorithm in Bahasa", *Master dissertation, Universitas Sumatera Utara, Medan, Indonesia*, 2018.
- [18] Fitri Irhamni, "Optimization of subdistrict grouping based on educational indicators using clustering methods and Davies Bouldin Index in Bahasa," *National Science and Technology Seminar*, pp. 1-5, 2014.
- [19] Sadia Nawrin, Md Rahatur Rahman, Shamim Akhter, "Exploreing K-Means with internal validity indexes for data clustering in traffic management system," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 8, no. 3, 2017, doi: 10.14569/IJACSA.2017.080337.
- [20] S. Ghosh and S. Kumar, "Comparative analysis of K-Means and fuzzy C-Means algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 4, pp. 35-38, 2013, doi: 10.14569/IJACSA.2013.040406.
- [21] T. Rahman, Wiranto, and A. Rini, "Coal trade data clustering using K-Means (case study PT. Global Bangkit Utama)," *ITSMART Journal Technology and Informatics.*, vol. 6, no. 1, pp. 24-31, 2017, doi: 10.20961/itsmart.v6i1.11296.
- [22] Ediyanto, M. N. Mara, and N. Satyahadewi, "Characteristics classification using the K-Means cluster analysis method in Language," *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, vol. 2, no. 2, pp. 133-136, 2013, doi: 10.26418/bbimst.v2i02.3033.
- [23] Suyanto, "Data mining for data classification and clustering", *Bandung: Informatics Publisher*, 2017.
- [24] J. Wu, *Advances in K-means clustering*, Berlin Heidelberg: Springer-Verlag, 2012, doi: 10.1007/978-3-642-29807-3.
- [25] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224-227, 1979, doi: 10.1109/TPAMI.1979.4766909.