P-ISSN: 1979-5661
E-ISSN: 2622-321X

Grouping of Electricity Regions with K-Medoid Algorithm

Dewa Ayu Indah Cahya Dewi[a1], Dewa Ayu Kadek Pramita[b2], Putri Alit Widyastuti Santiary[a3], I Ketut Swardika[a4], I Gusti Agung Made Sunaya[a5], I Ketut Bangse[c6]

[a]Departement of Electrical Engineering, Bali State Polytechnic
Badung, Bali, Indonesia
[1]ayuindahcahyadewi@pnb.ac.id (Corresponding author)
[3]putrialit@pnb.ac.id
[4]swardika@pnb.ac.id
[5]a_sunaya@pnb.ac.id

[b]Departement of Computer Systems, Institut Bisnis dan Teknologi Indonesia
Denpasar, Bali, Indonesia
[2]pramita.wayu@instiki.ac.id

[c]Departement of Mechanical Engineering, Bali State Polytechnic
Badung, Bali, Indonesia
[3]bangse@pnb.ac.id

Abstract

All human activities are inseparable from the current use of electricity. Electricity sourced from renewable energy is an alternative for areas that have difficulty accessing electricity from PLN. This research gathers data on electricity consumers in Indonesia by utilizing the K-Medoid algorithm and the Davies-Bouldin Index (DBI). The DBI is used to test the quality of clustering so that the optimal cluster in this study is obtained with the lowest DBI value of 0.386 with 3 clusters. The cluster analysis results reveal that fewer areas utilize electricity sourced from PLN compared to non-PLN or renewable energy sources. Based on this, recommendations can be proposed to enhance development, making it more accessible for villages to be electrified from PLN or to increase the electrification of villages currently not using electricity. source of electricity from non-PLN to realize the Bright Indonesia Program.

Keywords: *Electricity, Clustering, K-Medoid Algorithm, Davies Bouldin Index, Renewable Energy*

1. Introduction

The use of electricity is currently supporting almost all sectors of human life. With electricity, it can increase work ethic and productivity so that it can create community welfare. It is undeniable that there are still areas in Indonesia that have not been electrified by PLN due to the difficulty of access. The use of electricity sourced from renewable energy can be an alternative for areas that have not been electrified by PLN [1]. Renewable energy sources include solar cells, wind power, micro hydro, and various other alternative sources that can be utilized [2]. This study analyzes groups of areas that are electrified and the patterns formed from the characteristics of each group member so that they can support decision making. This research focuses on grouping data related to electricity consumers in Indonesia, utilizing the K-Medoid clustering technique, and assessing the outcomes through the Davies-Bouldin Index (DBI). The selection of the K-Medoid method for clustering is driven by its resilience to noise and data outliers [3]. The cluster validity method was used to determine the quality of the cluster results, namely the dunn index, davies-bouldin index, silhouette index, and others [4], [5]. The quality of the optimal cluster results is demonstrated by the lower DBI value [6].

The author's research draws on several previous studies for its foundation. One such study conducted in 2022 by Radiyanto Dekaprasetya, titled "Clustering to Determine the Promotion Strategy of the University of Muhammadiyah Jember Using the K-Medoids Algorithm," focused

P-ISSN: 1979-5661
E-ISSN: 2622-321X

on grouping student data to acquire student profile information as decision support for an effective promotion strategy. The study yielded an optimal cluster with a low DBI value of k = 3 [7]. Cici's 2020 research titled "Development of the K-Medoid Datamining Method in the Case of Electricity Distribution in Indonesia" delves into the clustering of electricity distribution areas using the K-Medoid method. The study resulted in two clusters: the high distribution level cluster (C1) and the low distribution level cluster (C2). The data centroid for the high distribution level cluster is 38,544.51, while the data centroid for the low distribution level cluster is 910.51 [8]. In Desi's 2019 study, "Application of the K-Medoid Method in Grouping Households in the Treatment of Sorting Waste by Province," waste sorting behavior classification was examined using the K-Medoid data mining algorithm. The research resulted in two clusters: a low-level waste sorting cluster (C1) and a high-level waste sorting cluster (C2). [9].

2. Reseach Methods

2.1. K-Medoid Algorithm

The K-Medoids algorithm is used to calculate the closeness between medoids and non-medoids objects and minimizes the dissimilarity of objects in the cluster. The following is a description of the k-medoids algorithm:

1. Selecting the number of clusters and setting the initial medoid value
2. Calculating the proximity distance between medoid objects and non-medoid objects using Euclidean calculations with equation 1.
   $$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_n - y_n)^2} \quad (1)$$
   Where :
   d = object distance
   x = data
   y = centroid/medoid
3. Randomly select non-medoid objects to become new medoids
4. Calculating the proximity between a medoid object and a non-medoid object using Euclidean calculations
5. The total deviation (S) is computed by subtracting the value of the new total distance from the old total distance. If the total deviation is less than zero (S < 0), the objects' data clusters are swapped to create a new set of k objects as medoids; otherwise, the medoids remain unchanged [10].
6. Steps 3 to 6 are iteratively repeated until no further changes occur in the medoids [11].

2.2. Davies-Bouldin Index (DBI)

The Davies-Bouldin Index (DBI) is a method used to assess the quality of clustering results. A clustering outcome is considered better when the DBI value approaches 0, indicating a more optimal clustering scheme with smaller DBI values [6], [12], [13], [14].

The Sum of Squares Within Cluster (SSW) is an equation used to calculate the cohesion matrix in the i-th cluster, as shown in the following equation.

$$SSW_i = \frac{1}{m_i}\sum_{j=i}^{m_i} d(x_j, c_i) \quad (2)$$

Where:
mi = the data number in the i cluster
ci = centroid of cluster i
d (xj, ci) = the euclidean distance of each data point to the centroid is calculated.

The Sum of Squares Between Clusters (SSB) is utilized to evaluate the degree of separation between clusters, as demonstrated in the following equation.

$$SSB_{i,j} = d(c_i, c_j) \quad (3)$$

Where :

P-ISSN: 1979-5661
E-ISSN: 2622-321X

d (ci, cj) = distance between centroids

Once the cohesion and separation values are acquired, the next step is to calculate the ratio measurement (Rij) for determining the comparative value between the i-cluster and j-cluster, as shown in the following equation.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (4)$$

The formula for computing the Davies-Bouldin Index (DBI) value is as follows.

$$DBI = \frac{1}{K}\sum_{i=1}^{k} max_{i\neq j}\left(R_{i,j}\right) \quad (5)$$

Where :
k = clusters number
The proximity of the DBI value to 0 indicates a superior cluster quality obtained from the clustering of the data. [15].

2.3. Preprocessing data

The data preprocessing stage in this study includes the process of transforming data in number format and reducing missing value data so that the data obtained is clean and free from noise [16], [17]. The secondary data variables used by the Badan Pusat Statistik in the excel form which consist several villages in each province, in Indonesia that uses electricity sourced from PLN and non-PLN in 2021 [18]. The data used in the study uses data from the Badan Pusat Statistik in 2021, totaling 34 units. province in Indonesia. The following are data variables from the Badan Pusat Statistik used for clustering, namely PLN variables and non-PLN variables:

	A	B	C
1	**Province**	**PLN**	**Non PLN**
2	Aceh	6503	170
3	North Sumatera	5931	692
4	West Sumatera	1276	166
5	Riau	1828	646
6	Jambi	1535	324
7	South Sumatera	3137	626
8	Bengkulu	1511	107
9	Lampung	2626	307
10	Bangka Belitung	393	38
11	Riau	392	141
12	DKI Jakarta	267	0
13	West Java	5957	41
14	Central Java	8560	23
15	DI Yogyakarta	438	3
16	East Java	8473	140
17	Banten	1551	14
18	Bali	716	6
19	West Nusa Tenggara	1145	45
20	East Nusa Tenggara	3024	1468
21	West Kalimantan	1612	1010
22	Central Kalimantan	1088	775
23	South Kalimantan	1980	229
24	East Kalimantan	834	519
25	North Kalimantan	295	250
26	North Sulawesi	1821	71
27	Central Sulawesi	1861	516
28	South Sulawesi	2947	493
29	Southeast Sulawesi	2188	416
30	Gorontalo	725	157
31	West Sulawesi	595	271
32	Maluku	918	419
33	North Maluku	1005	287
34	West Papua	968	961
35	Papua	1258	3453
36	**Deviation Standard**	**2237.912**	**632.991**
37	**Mean**	**2216.412**	**434.8235**

Figure 1. Data and Variables

The characteristics of village data, encompassing both electricity users from PLN and non-PLN sources, demonstrate substantial dispersion, as evidenced by a standard deviation value greater than the mean value [19]. In this research, the data utilized for clustering exhibits heterogeneity, primarily attributed to the variance distribution of the data [20].

2.4. Research Procedure

The research flow to conduct this research as follows.

1. Conducting literature studies and data collection related to data analysis using K-Medoids for data clustering and cluster validity testing with DBI
2. Preprocessing data before data is clustered.
3. The clean data after going through the preprocessing stage is then clustered using the K-Medoid method.
4. Perform cluster validity testing so that optimal clustering results are obtained
5. The research flowchart can be observed in Figure 2 provided below.

P-ISSN: 1979-5661
E-ISSN: 2622-321X

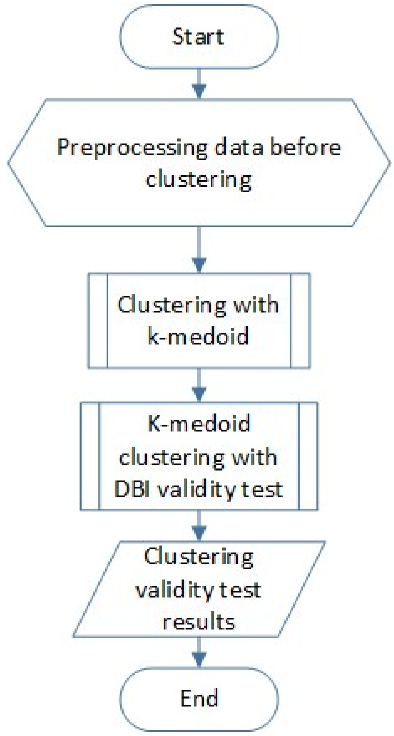

Figure 2. Flowchart of the Study

P-ISSN: 1979-5661
E-ISSN: 2622-321X

2.5. Research Overview

In this research, the focus lies in comparing the clustering outcomes for various numbers of clusters to identify the optimal number of clusters when using the k-medoid algorithm for clustering electricity user data.

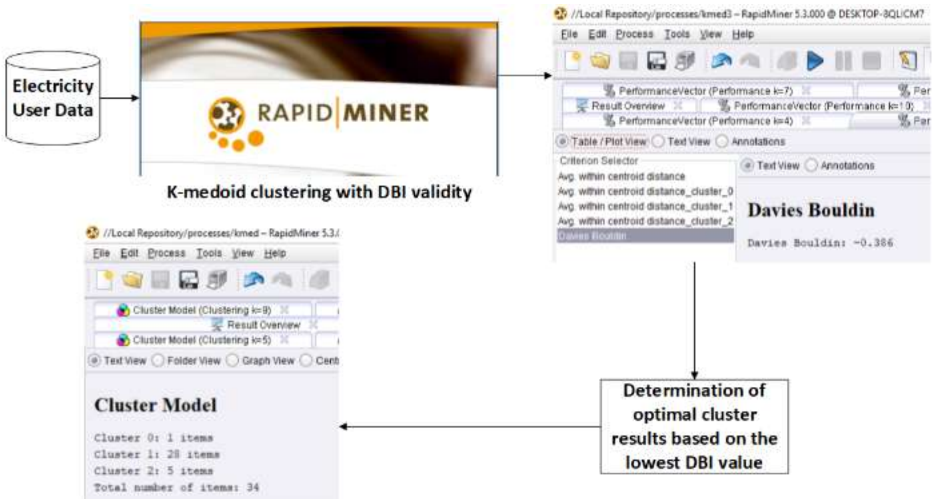

Figure 3. Research Overview

Figure 3 provides an overview of the K-Medoid clustering process with cluster validity testing using the Davies-Bouldin Index in Rapid Miner. The initial step involves data collection from the Badan Pusat Statistik, followed by data preprocessing in the second step. The third step is to test the clustering quality based on the DBI value on Rapid Miner, then analyze the cluster results so that the optimal clusters number is obtained in the k-medoid clustering with DBI method.

3. Result and Discussion

3.1. Grouping data with K-Medoid

The data was grouped using the k-medoid clustering method through the Rapid Miner application, and the cluster validity was assessed using DBI values. The grouping process started with 2 clusters and continued up to 10 clusters.

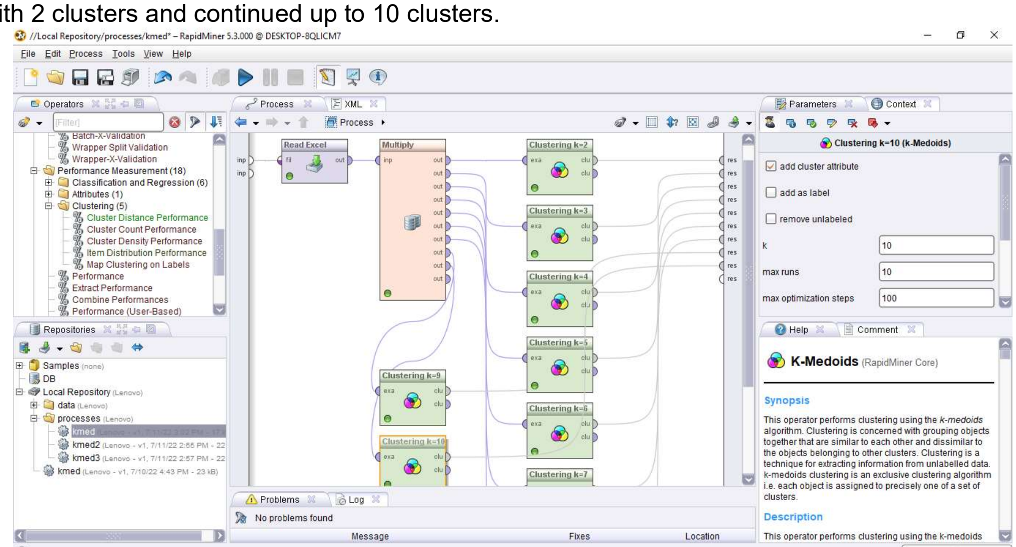

Figure 4. Clustering using Rapid Miner

P-ISSN: 1979-5661
E-ISSN: 2622-321X

After implementing the K-Medoid algorithm with two clusters, the analysis unveiled the first cluster, which primarily consisted of electricity users from PLN, encompassing 1258 villages. In contrast, the second cluster comprised non-PLN electricity users, comprising a total of 3453 villages located in Papua. Members of this cluster center amounted to 1 data. The second cluster center is electricity users from PLN as many as 968 villages and electricity users from non-PLN as many as 961 villages located in West Papua. The second cluster consists of 33 data with electricity users from PLN in the range of 267 to 8560 villages and electricity users from non-PLN in the range 0 to 1468 villages. The first cluster is classified as a village that uses more electricity than the second cluster.

Table 1. The results of grouping data with 2 clusters

Result	Centroid Position	Centroid		Amount of Cluster
		PLN	Non PLN	
Cluster 1	Papua	1258	3453	1
Cluster 2	West Papua	968	961	33

Grouping the data with 3 clusters produces 3 cluster centers located in Papua, West Papua and East Java. The first cluster is 1 data, the second cluster is 28 data. The third Cluster Center in East Java has electricity users from PLN as many as 8473 villages and non-PLN as many as 140 villages. The third cluster has 5 data members, namely Aceh, West Java, North Sumatra, DI Yogyakarta, DKI Jakarta with characteristics of PLN in the range of 5931 to 8560 villages and non-PLN in the range of 23 to 692 villages.

Table 2. The results of grouping data with 3 clusters

Result	Centroid Position	Centroid		Amount of Cluster
		PLN	Non PLN	
Cluster 1	Papua	1258	3453	1
Cluster 2	West Papua	968	961	28
Cluster 3	East Java	8473	140	5

3.2. Cluster Validity Test

The testing method for clustering results is classified as good or not good, it can be assessed from one of the cluster validity methods [5], [21], [22]. In this research, the Davies-Bouldin Index (DBI) method was employed to evaluate the clustered results. The optimality of a cluster is determined by its minimum DBI value [23].

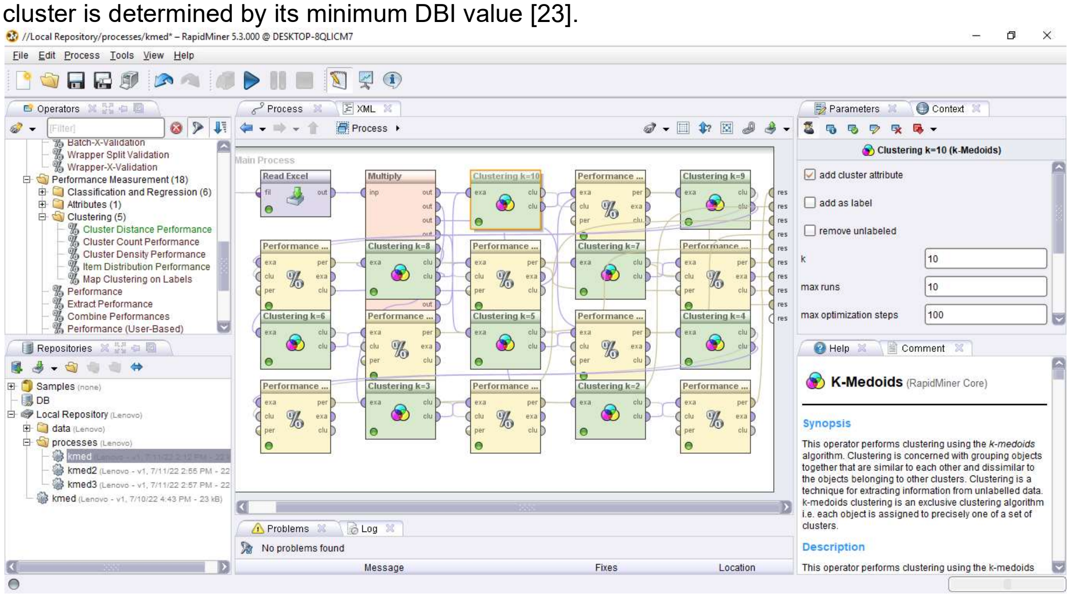

Figure 5. Validity Test in Rapid Miner

After conducting k-medoid clustering and evaluating the cluster validity using the Davies Bouldin index for each number of clusters, we obtained the following table.

Table 3. Validity Test Result

Amount of Cluster	DBI Value
2 clusters	0.724
3 clusters	0.386
4 clusters	0.671
5 clusters	1.025
6 clusters	0.761
7 clusters	0.543
8 clusters	0.66
9 clusters	0.707
10 clusters	0.651

Table 3 displays the DBI values for different cluster numbers, and it is evident that the cluster with the smallest DBI value is when there are 3 clusters with a value of 0.386. This suggests that the most optimal cluster quality is achieved with 3 clusters.

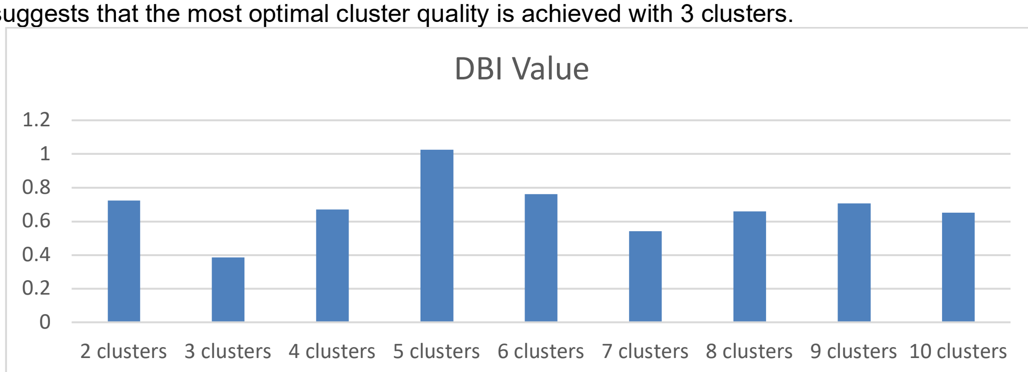

Figure 6. Davies-Bouldin Index Value

The results of clustering with 3 clusters produce 3 cluster centers located in Papua, West Papua, and East Java. The first cluster is 1 data, the second cluster is 28 data. The third Cluster Center in East Java has electricity users from PLN in as many as 8473 villages and non-PLN in as many as 140 villages. The third cluster has 5 data members, namely Aceh, DKI Jakarta, North Sumatra, DI Yogyakarta,West Java with characteristics of PLN in the range of 5931 to 8560 villages and non-PLN in the range of 23 to 692 villages. From the results of clustering village groups that use non-PLN electricity which is quite large, around 3000 villages form their clusters. Village groups that use PLN electricity are large enough to form clusters with an average of 6000-8000 villages/cluster members using electricity from PLN and 200 villages for non-PLN electricity. The third group has the characteristics of cluster members with an average number of PLN electricity users of around 1000 villages and the number of villages using non-PLN electricity from 0 -1000 villages.

4. Conclusion

After thoroughly examining the grouping of electricity user data with the K-Medoid Algorithm, the following conclusions have been drawn.

The clustering of the data into 3 clusters yields improved cluster quality, evident from the smallest DBI value achieved. The clustering results are influenced by the data characteristics within the cluster, particularly clusters with a larger dispersion size, as indicated by standard deviation values greater than the mean value. The data used for clustering exhibit significant variability. The area that is classified as the largest group of non-PLN electricity users is Papua with an average of around 3000 villages. From this group, it can be analyzed that there are fewer users of electricity from PLN than electricity sourced from non-PLN or renewable energy so recommendations can be made to improve development that can make it easier for villages in Papua to be electrified from PLN as well as to increase the number of villages that use electricity. not yet electrified in Papua using electricity from non-PLN sources. Regional groups that use PLN electricity the most are around 6000-8000 villages, namely Aceh, DKI Jakarta, North Sumatra, DI Yogyakarta,West Java. The third group that uses PLN and non-PLN electricity are around 1000 villages.

References

[1] PT. PLN (Persero), “Rencana Usaha Penyediaan Tenaga Listrik PT. PLN (Persero).” [Online]. Available: https://web.pln.co.id/statics/uploads/2018/04/RUPTL-PLN-2018-2027.pdf.

[2] R. Syahputra and I. Soesanti, “Renewable energy systems based on micro-hydro and solar photovoltaic for rural areas: A case study in Yogyakarta, Indonesia,” *Energy Reports*, vol. 7, 2021.

[3] W. S. Sarle, L. Kaufman, and P. J. Rousseeuw, “Finding Groups in Data: An Introduction to Cluster Analysis.,” *Journal of the American Statistical Association*, 1991.

[4] M. A. Nahdliyah, T. Widiharih, and A. Prahutama, “Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index,” *Jurnal Gaussian*, vol. 8, no. 2, 2019.

[5] D. S. Rini, P. Novianti, and H. Fransiska, “Internal Cluster Validation on Earthquake Data in the Province of Bengkulu,” in *IOP Conference Series: Materials Science and Engineering*, 2018.

[6] D. L. Davies and D. W. Bouldin, “A Cluster Separation Measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, 1979.

[7] R. Dekaprasetya, U. Anisatur, and L. A. Muharom, “Clustering Untuk Menentukan Strategi Promosi Universitas Muhammadiyah Jember Dengan Algoritma K-Medoids Clustering To Determine The University Of Muhammadiyah Jember Promotion Strategy With K-Medoids Algorithm,” *Jurnal Smart Teknologi*, vol. 3, no. 3, pp. 312–320, 2022.

[8] C. Astria, D. Hartama, A. P. Windarto, and I. Sudahri, “Pengembangan Metode Datamining K-Medoid Pada Kasus Distribusi Listrik di Indonesia,” *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2020.

[9] D. A. Silitonga, A. P. Windarto, D. Hartama, and Sumarno, “Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi,” *Seminar Nasional Sains & Teknologi Informasi (SENSASI) SENSASI 2019 ISBN:*, 2019.

[10] N. L. Ratniasih and R. A. N. Diaz, “Comparison Of Clustering Algorithm In Employee Training Management Recommendations,” in *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, 2021, pp. 1–4.

[11] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, and D. Hartama, “ANALISIS ALGORITMA K-MEDOIDS CLUSTERING DALAM PENGELOMPOKAN PENYEBARAN COVID-19 DI INDONESIA,” *Jurnal Teknologi Informasi*, vol. 4, no. 1, 2020.

[12] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo, “PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS DALAM KLASTERISASI PRODUK ASURANSI PERUSAHAAN NASIONAL,” *Jurnal Ilmiah FIFO*, vol. 11, no. 2, 2019.

[13] B. Jumadi Dehotman Sitompul, O. Salim Sitompul, and P. Sihombing, “Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm,” in *Journal of Physics: Conference Series*, 2019, vol. 1235, no. 1.

[14] A. M. Sikana and A. W. Wijayanto, “Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering,” *Jurnal Ilmu Komputer*, vol. 14, no. 2, 2021.

[15] H. Prananditya, M. Ary, and A. Z. Fuadi, “Clustering of Electricity Usage Using K-Means Method,” in *2021 International Conference on Computer Science and Engineering (IC2SE)*, 2021, vol. 1, pp. 1–7.

[16] J. Han, M. Kamber, and J. Pei, “Data Preprocessing,” in *Data Mining*, 2012.

[17] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, “A Review on Data Preprocessing

Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," *Frontiers in Energy Research*, vol. 9. 2021.
[18] Badan Pusat Statistik, "Banyaknya Desa Menurut Keberadaan Keluarga Pengguna Listrik," 2021. [Online]. Available: https://www.bps.go.id/indicator/168/950/1/banyaknya-desa-kelurahan-menurut-keberadaan-keluarga-pengguna-listrik.html.
[19] A. Ferragamo, J. A. Rubiño-Martín, J. Betancort-Rijo, E. Munari, B. Sartoris, and R. Barrena, "Biases in galaxy cluster velocity dispersion and mass estimates in the small N galregime," *Astronomy and Astrophysics*, vol. 641, 2020.
[20] C. N. P. G. Arachchige, L. A. Prendergast, and R. G. Staudte, "Robust analogs to the coefficient of variation," *Journal of Applied Statistics*, vol. 49, no. 2, 2022.
[21] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1981.
[22] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, 1987.
[23] I. Kamila, U. Khairunnisa, and Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019.