Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi
Vol. 12, No. 4, Desember 2024

ISSN 0216 – 0544
e-ISSN 2301– 6914

IDENTIFYING THE CLUSTER OF FAMILIES AT RISK OF STUNTING IN YOGYAKARTA USING HIERARCHICAL AND NON-HIERARCHICAL APPROACH

[a]Ersa Riga Puspita, [b]Mujiati Dwi Kartikasari*

[a,b]Department of Statistics, Universitas Islam Indonesia, Kaliurang St., Km. 14.5 Sleman, Yogyakarta, 55584, Indonesia
E-mail: 19611178@students.uii.ac.id, mujiatikartikasari@uii.ac.id*

Abstract

Stunting, or short stature, is a growth disorder usually caused by chronic dietary deficiencies from the prenatal stage to early childhood, typically becoming evident in children after the age of 2. Stunting cases in Yogyakarta Province experienced a decline in 2020. With this development, the government aims to achieve zero stunting in Yogyakarta Province by 2024. To support this goal, a research study was conducted in 2021 to analyze family factors associated with stunting risks in Yogyakarta Province. The study aimed to assist the government in addressing the issue and achieving the target. In this research, a hierarchical clustering algorithm using the Ward technique and a non-hierarchical clustering algorithm using the Fuzzy C-Means (FCM) approach were applied. The optimal number of clusters was determined using the average distance and figure of merit approach. Stability validation, which also used the average distance and figure of merit approach, demonstrated that the best results were achieved by the non-hierarchical clustering algorithm employing FCM. As a result, six clusters were identified: cluster 1 with 5 sub-districts, cluster 2 with 18 sub-districts, cluster 3 with 21 sub-districts, cluster 4 with 17 sub-districts, cluster 5 with 14 sub-districts, and cluster 6 with 3 sub-districts.

Key words: Cluster, Stunting, Hierarchical Clustering, Non-Hierarchical Clustering, Ward, Fuzzy C-Means.

INTRODUCTION

Data is a critical component in producing People anticipate being in good health because it boosts their productivity and lowers their risk of death. Despite their ignorance about the causes of many diseases, humans are highly susceptible to them. Therefore, everyone needs to understand the importance of learning about disease causes, including stunting, as early as possible [1]. Stunting, as we know, is a condition in which a child's growth is stunted due to prolonged nutritional deficiencies. Stunting can occur from the moment the fetus is in the womb until the start of the child's life. This exemplifies how the characteristics of malnutrition can cause children to grow slower than typical for their age [2]. The prevalence of stunting is influenced by various factors, such as family livelihoods, work opportunities, and the state of the economy. Additionally, inadequate access to health services, including clean water and sanitary facilities, significantly impacts a child's growth [3].

Impaired development and growth in children as a result of persistent infections and chronic malnutrition are manifested as below-average height. In the Yogyakarta Special Region, the frequency of stunting was 21.03%

in 2019 and dropped to 19.88% in 2020. According to Presidential Regulation No. 72 of 2021, the city of Yogyakarta is expected to achieve zero stunting by 2024, with a target of reducing the risk of stunting to 14% by that same year [4]. Although the occurrences of stunting have decreased, it would be beneficial to adopt promotion and prevention strategies to take appropriate safety measures. Stunting can be prevented and controlled through the implementation of preventative and promotional strategies [1]. However, since not every part of Yogyakarta has the same family background, there's a chance that this strategy won't fully support the actual targets. It is critical to comprehend the distribution pattern of stunting. Therefore, an analysis of public health data is needed to understand the pattern of stunting distribution in each sub-district in Yogyakarta. The technique generally used for this purpose is data mining.

Data mining is the process of extracting knowledge and valuable information from massive databases by applying mathematical, statistical, artificial intelligence, and machine learning techniques. Clustering is one of the many tasks involved in data mining. The clustering method groups observations or objects based on their shared characteristics. Cluster analysis can identify high-risk areas for disease, making it useful in disease prevention by examining the number and distribution of diseases. With this knowledge, the government may find it simpler to instruct paramedics on how to eradicate or combat potential illnesses in the region. Data clustering is one of the unsupervised data mining techniques. Hierarchical and non-hierarchical clustering are the two forms of data grouping frequently employed in the data grouping process [5]. The outcomes of this case grouping are anticipated to inform health services about diseases that may be present in the area. Hierarchical clustering and non-hierarchical clustering are the two types of clustering algorithms used in data mining. Cluster analysis using the hierarchical clustering method is conducted by calculating the proximity to each object, which creates a dendrogram. There are several methods in hierarchical clustering, including Ward, Average Linkage, Complete Linkage, and Single Linkage. Additionally, there are several techniques for non-hierarchical clustering, such as fuzzy clustering, K-Means, and K-Medoids [6].

The hierarchical clustering approach has advantages in cluster analysis, such as speeding up processing and saving time because the incoming data will generate its own dendrogram or level, making it easier to analyze. The non-hierarchical clustering technique also has advantages, including the ability to analyze large samples more efficiently. To select the optimum algorithm, we will compare two methods: the Ward technique and Fuzzy C-Means (FCM). Ward's approach offers advantages in terms of efficiency and tends to build clusters until they reach the smallest size [6]. Additionally, the advantage of the FCM approach is that by repeatedly improving the cluster center, the cluster center moves towards the proper place. FCM also has a high level of accuracy and rapid computation time [7]. The FCM approach benefits from placing cluster centers more precisely than previous cluster methods [8]. Given the scenario of stunting cases, it is necessary to group families at risk of stunting based on sub-districts in Yogyakarta Province, which remains a focal point. Therefore, researchers conducted a study on clustering areas with family variables at risk of stunting using the Ward and FCM methods.

MATERIAL AND METHODS

The National Family Planning Coordination Board, also known as the Badan Koordinasi Keluarga Berencana Nasional (BKKBN), provided the secondary data used in this study. The data includes the number of families in Yogyakarta Special Region at risk of stunting in 2021.

There were multiple phases to this study. First, secondary data were gathered by compiling information on families in Yogyakarta Special Region at risk of stunting. After that, a descriptive analysis was performed to obtain a summary of the data. Next, outliers and multicollinearity were examined to test the clustering process's assumptions. The best cluster and method were identified by utilizing validation indices, namely the average distance and figure of merit, following a comparison of the Ward and FCM methods to determine which is better. Lastly, the clusters of the best method results were profiled. Figure 1 displays the research flowchart.

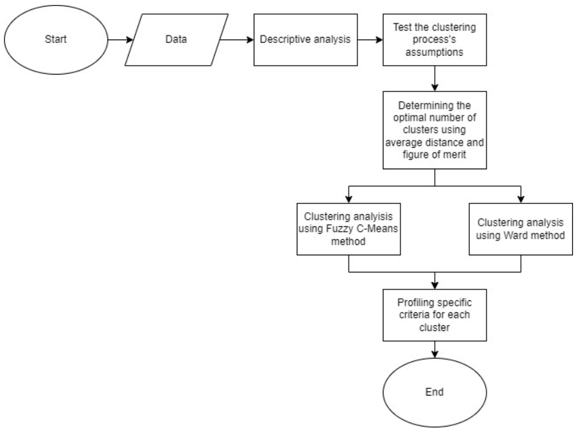

Fig 1. The research flowchart

Fuzzy C-Means

Dunn made the original discovery of the FCM method in 1973, and Bezdek further developed it in 1981. The basic concept of the method is similar to the C-Means method. Each data point is grouped according to its membership value in the group in FCM, which is based on fuzzy logic [9].

One clustering technique is fuzzy clustering, which utilizes a set of fuzzy measurements as well as a classification process based on membership degree. There is a chance that every piece of data will belong to a particular cluster. This indicates that not all of the data will fit perfectly into a single cluster [10].

A type of soft clustering, in which each data point can belong to more than one cluster, is called FCM, an extension of the Hard C-Means method [11]. dditionally, the FCM algorithm can be utilized to cluster data based on the membership degree of each point in the data [12]. Below is the FCM algorithm [9].

1. Input the data to be clustered
2. Determine the values needed in the FCM calculation such as:
   Expected number of clusters = c
   Rank of weights = w
   Maximum iterations = $MaxIter$
   Least expected error = ζ
   Initial objective function = $P_0 = 0$
   Initial iteration = $t = 1$
3. Generate random numbers u_{ik} with i being the number of data and k being the number of groups as the initial elements of the initial membership matrix u_{ik}. u_{ik} is commonly referred to as the degree of membership of the first data sample in the kth cluster. The position and value of the matrix are constructed randomly, and the membership value lies in the interval 0 to 1. The initial position of the U partition matrix lacks accuracy, as does the cluster center. Consequently, the accuracy of the data's tendency to enter a cluster is compromised. Therefore, calculate the number of each attribute column using the following equation:

$$Q_i = \sum_{i=1}^{c} (u_{ik}) \tag{1}$$

   Q_i is the sum of the membership degree values per column, which is equal to 1, where $i = 1,2,3,\cdots,n$.
4. Calculate the kth cluster center (V_{kj}) using the equation below, where $k = 1,2,3,\ldots,c$; and $j = 1,2,3,\ldots,m$

$$V_{kj} = \frac{\sum_{k=1}^{n} (u_{ik})^{w} * X_{ij}}{\sum_{k=1}^{n} (u_{ik})^{w}} \tag{2}$$

5. Calculate the objective function at iteration t.

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j}^{m} (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^{w} \right) \tag{3}$$

6. Calculate the change in the membership partition matrix with $i = 1,2,3\ldots,n$; and $k = 1,2,3,\ldots,c$; using the equation below:

$$u_{ik} = \frac{\left[\sum_{j}^{m} (X_{kj} - V_{ij})^2 \right]^{\frac{-1}{p-1}}}{\sum_{k=1}^{c} \left[\sum_{j=1}^{m} (X_{kj} - V_{ij})^2 \right]^{\frac{1}{p-1}}} \tag{4}$$

7. Check the condition to determine if the process should be stopped.
   If $|J_t - J_{t-1}| < \zeta \text{ or } t > \text{MaxIter}$, then stop. If not, then $t = t + 1$, repeat step 3 with equation 1.

Ward

One popular technique for clustering that aims to produce clusters with the least amount of internal variance is Ward's approach. Calculating the average of each cluster is a common step in Ward's approach, used to measure this variation. This approach not only computes the average but also measures the Euclidean distance between each object value and its average, which is then continually added up. This method employs the least "sum of squares," which will be combined at each stage of the procedure [13].

Ward's approach maximizes homogeneity within a group and involves comprehensive computation. In a hierarchical clustering process, the agglomerative nature is the stage where every object begins in a different cluster. The sum of squared errors (SSE) is the metric used for Ward's clustering approach. The formulation of SSE is

$$SSE = \sum_{i=1}^{n}(X_i - \bar{X})(X_j - \bar{X}) \quad (5)$$

where X_i represent the value of variable X in the ith data, X_j represent the value of variable X in the jth data, $\bar{X}$ represent the mean of all object in the cluster, and n represent the total number of object in the cluster.

The processes involved in performing a hierarchical cluster analysis using the Ward technique are as follows [14]:

1. To begin, consider N clusters, where each cluster contains objects; all objects are regarded as clusters.
2. To merge two clusters into a single cluster, find the least variance between them..
3. Continue stage 2 until the goal is achieved.

Average Distance

The average distance metric calculates the average distance between observations clustered in the same cluster using both clustering based on complete data and clustering based on data with one column removed. The average distance can be computed using the formula below [15].

$$AD(K) = \frac{1}{MN}\sum_{i=1}^{N}\sum_{l=1}^{M}\frac{1}{n(C^{i,0})n(C^{i,l})}\left[\sum_{i\in C^{i,0}, j\in C^{i,l}} dist(i,j)\right] \quad (6)$$

The notation N represents the total number of observations (rows) in a dataset, M represents the total number of columns, C is cluster, $C^{i,0}$ represents the cluster containing observation i using the original clustering (based on all available data), and $C^{i,l}$ represents the cluster containing observation i where the clustering is where the clustering is based on the dataset with column l removed. Smaller values of the average distance, which range from zero to ∞, are favoured.

Figure of Merit

Since the clustering is based on the remaining or non-deleted samples, the validity in the figure of merit displays the average intra-cluster variance of the data in the deleted column. Predictions based on the cluster average, leaving column l should be used to estimate the average error; the FOM is [15]

$$FOM(l,K) = \sqrt{\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in C_k(l)} dist(x_i, l, \bar{x}_{C_k}(l))} \quad (7)$$

where x_i, l is the value of the ith observation in the lth column in cluster $C_k(l)$, and $\bar{x}_{C_k}(l)$ is the average of cluster $C_k(l)$. Currently, the only distance available for FOM is Euclidean. The last value is averaged over all deleted columns, and has a value between 0 and ∞, with smaller values giving better results.

RESULT AND DISCUSSION

Descriptive of Data

The data used in this study are secondary data on the family risk of stunting from The National Family Planning Coordination Board, also known as the Badan Koordinasi Keluarga Berencana Nasional (BKKBN). The data consist of 4 (four) variables: the variable of not having proper drinking water (X_1), variable of not having a proper toilet (X_2), variable of not having a proper house (X_3), and variable of not having an income (X_4). The descriptive statistics of data is shown in Table 1.

Table 1. Descriptive Statistics of Data

Variable	Mean	Maximum	Minimum	Median
X_1	480.4	3011	17	345
X_2	411.8	1601	59	377
X_3	1512	4395	457	1386
X_4	87.55	321	5	73.5

Table 1 shows that the variable of not having proper drinking water has a mean of 480.4, a maximum value of 3011, a minimum value of 17, and a median of 345. The variable of not having a proper toilet has a mean of 411.8, a maximum value of 1601, a minimum value of 59, and a median of 377. The variable of not having a proper house has a mean of 1512, a maximum value of 4395, a minimum value of 457, and a median of 1386. The last variable, not having an income, has a mean of 87.55, a maximum value of 321, a minimum value of 5, and a median of 73.5.

Outliers Detection

Outlier data can be identified using the central tendency value. If a value falls outside the lower and upper limit range, then the data is considered an outlier. In outliers, there are also lower limits and upper limits [16].

In the boxplot visualization in Figure 2, it is clear that there are no outliers for each variable related to the lack of proper drinking water, lack of proper latrines, lack of proper housing, and lack of income, indicating that the assumptions are met for further analysis.

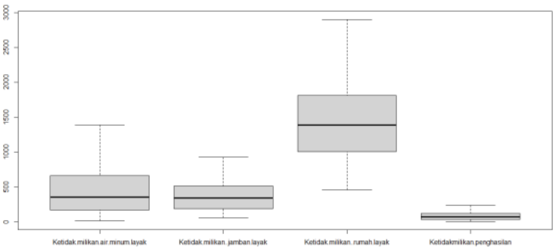

Fig 2. Boxplot visualization

Multicolinearity Test

In Table 2, it is shown that a correlation value of more than 0.85 indicates multicollinearity [17]. Therefore, since the correlation matrix in Table 2 does not contain such values, it can be concluded that there is no multicollinearity, allowing for further analysis of the results.

Table 2. Correlation Coefficient

	X_1	X_2	X_3	X_4
X_1	1.00	0.24	0.28	0.08
X_2	0.24	1.00	0.14	0.05
X_3	0.28	0.14	1.00	0.25
X_4	0.08	0.05	0.25	1.00

Determination of the Optimum Number of Clusters

To determine the number of clusters to use, we need to find the optimal number by using the average distance index and the figure of merit, as shown in Table 3.

Table 3. Optimum Number of Cluster

Cluster Method	Index	Number of Cluster				
		2	3	4	5	6
Ward	average distance	1032.33	900.26	837.83	820.66	802.72
	figure of merit	396.94	382.04	381.49	377.62	371.15
FCM	average distance	973.35	842.5	685.24	676.52	602.72
	figure of merit	395.8	377.7	358.73	368.03	357.63

Based on the results obtained from the average distance and figure of merit approaches in Table 3, the smallest value for the Ward method and Fuzzy C-Means is found in cluster 6. Therefore, the optimum number of clusters determined is 6.

Hierarchical Clustering Using Ward Method

The Ward method is a hierarchical method that can be used when the desired number of groups is unknown. This method is mostly used for objects with a relatively small number of observations [18]. The Ward method is a specific type of variance method, which aims to obtain clusters with the smallest possible cluster interval variances and where the average for each cluster is calculated. Using the Ward method, the results obtained consist of six clusters. Figure 3 is a map visualization of families at risk of stunting based on the results from Table 4.

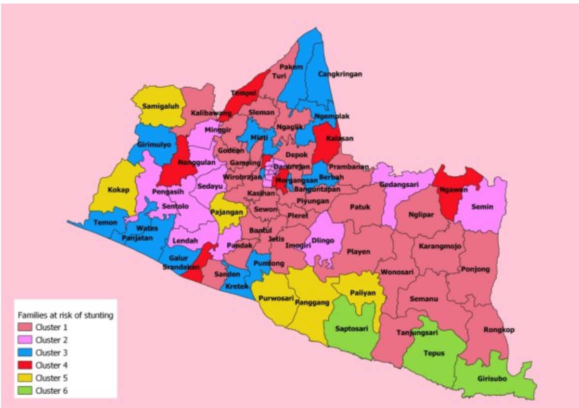

Fig 3. Cluster visualization of families at risk of stunting in Yogyakarta using the Ward method

Table 4. Clustering Results using Ward Method

Cluster	The Number of Cluster Members	Cluster Members
1	29	Gamping, Godean, Seyegan, Depok, Prambanan, Ngaglik, Sleman, Turi, Sanden, Bambanglipuro, Bantul, Jetis, Imogiri, Banguntapan, Pleret, Piyungan, Sewon, Kasihan, Wonosari, Nglipar, Playen, Patuk, Semanu, Karangmojo, Ponjong, Rongkop, Ngawen, Tanjungsari, Kalibawang
2	22	Moyudan, Minggir, Tempel, Pandak, Dlingo, Sedayu, Semin, Gedangsari, Tegalrejo, Jetis, Gondokusuman, Danurejan, Gedongtengen, Ngampilan, Mantrijeron, Kraton, Gondomanan, Pakualaman, Mergangsan, Lendah, Sentolo, Pengasih.
3	15	Mlati, Berbah, Ngemplak, Pakem, Cangkringan, Kretek, Pundong, Wirobrajan, Kotagede, Temon, Wates, Panjatan, Galur, Girimulyo.
4	3	Srandakan, Umbulharjo, Nanggulan
5	6	Pajangan, Paliyan, Panggang, Purwosari, Kokap, Samigaluh
6	3	Tepus, Saptosari, Girisubo.

Non-Hierarchical Clustering Using Fuzzy C-Means Method

The FCM algorithm can be used to cluster data based on the membership degree of each point. From the clusters obtained using the FCM method, six clusters are identified. Figure 4 is a map visualization of families at risk of stunting based on results from Table 5.

Fig 4. Cluster visualization of families at risk of stunting in Yogyakarta using the Fuzzy C-Means method

Table 5. Clustering Results using FCM Method

Cluster	The Number of Cluster Members	Cluster Members
1	5	Kalasan, Pajangan, Paliyan, Purwosari, Samigaluh.
2	18	Godean, Seyegan, Ngaglik, Sleman, Turi Bambanglipuro, Pandak, Bantul, Jetis, Piyungan, Patuk, Panggang, Karangmojo, Ponjong, Lendah, Kokap, Nanggulan, Kalibawang.
3	21	Berbah, Ngemplak, Cangkringan, Srandakan, Sanden, Kretek, Pundong, Nglipar, Danurejan, Gedongtengen, Ngampilan, Wirobrajan, Kraton, Gondomanan, Pakualaman, Kotagede, Temon, Wates, Panjatan, Galur, Girimulyo.
4	17	Moyudan, Minggir, Mlati, Rambanan, Tempel, Pakem, Dlingo, Pleret, Sedayu, Ngawen, Tegalrejo, Jetis, Gondokusuman, Mantrijeron, Mergangsan, Sentolo, Pengasih.
5	14	Gamping, Depok, Imogiri, Banguntapan, Sewon, Kasihan Wonosari, Playen, Semanu, Rongkop, Semin, Gedangsari, Tanjungsari, Umbulharjo.
6	3	Tepus, Saptosari, Girisubo.

Selection of the Best Method

Following the cluster analysis procedure, the Ward and FCM methods were used to determine the optimal approach. Two indices were utilized in the validation process of this study. The purpose of this validation was to determine which approach is more feasible than the others. The average distance and figure of merit are the indices utilized in this validation. The optimal approach was determined based on the findings of these two indices, which are displayed in Table 6.

Table 6. Determining the Best Method

Index	Ward	Fuzzy C Means
average distance	802.7195	602.7186
figure of merit	371.1546	357.6328

In comparison to the hierarchical approach using the Ward method, the non-hierarchical algorithm using the FCM method yields the best cluster analysis, as demonstrated by the values of each index in Table 6 with up to six clusters.

Interpretation of The Best Method

For data on families at risk of stunting in the Special Region of Yogyakarta in 2021, the FCM method is the best approach, according to the comparison of cluster methods. The results of the profiling produced using the cluster technique are shown in Table 7.

Table 7. Profilling the Best Method

Cluster	X_1	X_2	X_3	X_4
1	**1941.2**	**571**	1545.2	53.6
2	654.17	384.78	1567.33	94.55
3	162.57	309.95	807.42	58.38
4	206.29	421.41	1276.82	78
5	530.93	525.07	2218.57	**154.07**
6	544.33	438	**4101.33**	50

Based on Table 7, the profiling results focus on the highest average values for each variable: not having proper drinking water (X_1), not having a proper toilet (X_2), not having a proper house (X_3), and not having an income (X_4). This highlights the clusters with the highest profiles, which should be the main concern.

There are three clusters that need attention: cluster 1, cluster 5, and cluster 6. Cluster 1 requires the most attention because it ranks highest in terms of not having proper drinking water and toilets. It is followed by cluster 5, which has a significant issue with not having an income, and cluster 6, which faces a major problem with not having a proper house.

CONCLUSION

The results of the research indicate that the optimum number of clusters, determined using the average distance and figure of merit, is six. The best method, based on these indices, is hierarchical clustering using the FCM method. In the FCM method, there are 5 sub-districts in cluster 1, 18 sub-districts in cluster 2, 21 sub-districts in cluster 3, 17 sub-districts in cluster 4, 14 sub-districts in cluster 5, and 3 sub-districts in cluster 6.

REFERENCES

[1] Kemenkes, "Pencegahan Stunting Pada Anak," *Direktorat Promosi Kesehatan dan Pemberdayaan Masyarakat*, 2019. https://promkes.kemkes.go.id/pencegahan-stunting (accessed Jul. 02, 2023).

[2] D. Satriawan and D. A. Styawan, "Pengelompokkan Provinsi di Indonesia Berdasarkan Faktor Penyebab Balita Stunting Menggunakan Analisis Cluster Hierarki," *J. Stat. dan Apl.*, vol. 5, no. 1, pp. 61–70, 2021, doi: https://doi.org/10.21009/JSA.05106.

[3] K. Komalasari, E. Supriati, R. Sanjaya, and H. Ifayanti, "Faktor-Faktor Penyebab Kejadian Stunting Pada Balita," *Maj. Kesehat. Indones.*, vol. 1, no. 2, pp. 51–56, 2020, doi: 10.47679/makein.202010.

[4] Pemerintah Kota Yogyakarta, "Tahun 2024 Targetkan Kota Yogya Zero Stunting," *Portal Berita Pemerintah Kota Yogyakarta*, 2024. https://warta.jogjakota.go.id/detail/index/23501 (accessed Jul. 02, 2023).

[5] D. Siahaan and E. Prasetyo, "Perbaikan Metode Pemeringkatan Spesifikasi kebutuhan Berdasarkan Perkiraan Keuntungan dan Nilai Proyek dengan Mengurangi Perbandingan Berpasangan," *J. Ilm. Kursor*, vol. 6, no. 2, pp. 93–102, 2011.

[6] Imasdiani, I. Purnamasari, and F. D. T. Amijaya, "Perbandingan Hasil Analisis Cluster Dengan Menggunakan Metode Average Linkage Dan Metode Ward," *Eksponensial*, vol. 13, no. 1, pp. 9–18, 2022.

[7] D. L. Pardosi and I. D. Siagian, "Klasterisasi Data Lowongan Pekerjaan Berdasarkan Fuzzy C-Means," *J. Ilmu Komput. dan Sist. Inf.*, vol. 3, no. 1, pp. 27–31, 2020, doi: https://doi.org/10.9767/jikomsi.v3i1.79.

[8] I. Suciati, N. Herawati, S. Subian, and Widiarti, "Analisis Klaster Menggunakan Metode Fuzzy C-Means pada Data COVID-19 di Provinsi Lampung," 2022.

[9] L. Affandi, R. Arianto, and H. H. Firdausy, "Implementasi Algoritma Fuzzy C-Means pada Kasus Stunting Balita Berbasis Website," 2021.

[10] A. E. Haryati, Sugiyarto, and R. D. A.

Putri, "Comparison of Fuzzy Subtractive Clustering and Fuzzy C-Means," *J. Ilm. Kursor*, vol. 11, no. 1, 2021, doi: https://doi.org/10.21107/kursor.v11i1.254.

[11] J. Inayah, D. A. S. N. Maghfiroh, and D. C. R. Novitasari, "Clustering Daerah Rawan Kriminalitas Menggunakan Algoritma Fuzzy C-Means," *J. Ilm. Inform. Komput.*, vol. 27, no. 2, pp. 95–106, 2022, doi: https://doi.org/10.35760/ik.2022.v27i2.6019.

[12] W. Sanusi, A. Zaky, and B. N. Afni, "Analisis Fuzzy C-Means dan Penerapannya Dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Faktor-faktor Penyebab Gizi Buruk," *J. Math. Comput. Stat.*, vol. 2, no. 1, 2019, doi: https://doi.org/10.35580/jmathcos.v2i1.12458.

[13] M. Paramadina, S. Sudarmin, and M. K. Aidid, "Perbandingan Analisis Cluster Metode Average Linkage dan Metode Ward (Kasus: IPM Provinsi Sulawesi Selatan)," *Variansi J. Stat. Its Appl. Teach. Res.*, vol. 1, no. 2, 2019, doi: https://doi.org/10.35580/variansiunm9357.

[14] Y. I. Hartanto, A. Rusgiyono, and T. Wuryandari, "Penerapan Analisis Klaster Metode Ward Terhadap Kabupaten/Kota di Jawa Tengah Berdasarkan Pengguna Alat Kontrasepsi," *J. Gaussian*, vol. 6, no. 4, 2017, doi: https://doi.org/10.14710/j.gauss.6.4.528-537.

[15] G. Brock, V. Pihur, S. Datta, and S. Datta, "clValid : An R Package for Cluster Validation," *J. Stat. Softw.*, vol. 25, no. 4, 2008, doi: 10.18637/jss.v025.i04.

[16] I. M. S. Bimantara and I. M. Widiartha, "Optimization of K-Means Clustering Using Particle Swarm Optimization Algorithm For Grouping Traveler Reviews Data on Tripadvisor Sites," *J. Ilm. Kursor*, vol. 12, no. 1, 2023, doi: https://doi.org/10.21107/kursor.v12i01.269.

[17] Mahmudi, R. Goejantoro, and F. D. T. Amijaya, "Perbandingan Metode C-Means dan Fuzzy C-Means Pada Pengelompokan Kabupaten/Kota Di Kalimantan Berdasarkan Indikator IPM Tahun 2019," *Eksponensial*, vol. 12, no. 2, 2021.

[18] Y. Muflihan, H. Retnawati, and A. Kristian, "Analisis Cluster dengan Metode Hierarki untuk Pengelompokan Sekolah Menengah Atas Berdasarkan Rapor Mutu Sekolah di Kabupaten Nagan Raya," *Meas. Educ. Res.*, vol. 2, no. 1, pp. 22–33, 2022.

.