p-ISSN 2338 – 3216
e-ISSN 2528 – 1070

JURNAL STATISTIKA
UNIVERSITAS MUHAMMADIYAH SEMARANG
12(1) 2024: 01-09

Jurnal Statistika
Universitas Muhammadiyah Semarang

CLUSTERING DISTRICT/CITY IN WEST KALIMANTAN BASED ON FACTORS CAUSING STUNTING USING K-HARMONIC MEANS METHOD

Rahmania Andarini Hatti Imanni[1*], Evy Sulistianingsih, Hendra Perdana[3]
[1,2,3] Statistics Study Program, Department of Mathematics, Faculty Mathematic and Natural Science, University of Tanjungpura, Indonesia

***e-mail**: *rahmaniaandarini@student.untan.ac.id*

Article Info:
Received: July 23, 2023
Accepted: March 4, 2024
Available Online: May 31, 2024

Keywords:
Stunting; Cluster; Optimal

Abstract: Stunting is a chronic nutritional problem caused by inadequate dietary intake over time. The results of the Indonesian Nutrition Status Survey (SSGI) 2021 show that the percentage of stunting in West Kalimantan is 29.8%, higher than the national average. Based on the high number of stunting cases in West Kalimantan, it is necessary to group districts/cities in West Kalimantan based on the factors that cause stunting. This study aims to analyze the clustering of districts/cities in West Kalimantan based on the factors that cause stunting using the K-Harmonic Means method and analyze the number of optimal clusters using the silhouette coefficient. The percentage of households without access to clean drinking water (X_1), the rate of exclusive breastfeeding (X_2), percentage of babies with low birth weight who are born safely (X_3), the percentage of households without proper sanitation facilities (X_4) in 2021 are the variables analyzed in this study. The analysis results show that the optimal number of clusters is 4 with a silhouette coefficient value of 0.744, indicating a solid structure in the grouping. Cluster 1 is a cluster with a very high causal factor for stunting. The most influential factors in cluster 1 are households without access to clean drinking water, lack of exclusive breastfeeding, and low birth babies born safely.

1. INTRODUCTION

Stunting can be defined as a chronic nutritional condition due to insufficient intake of nutrients over a long period of time. Stunting is often associated with insufficient food intake by nutritional needs [1]. One of the provinces with high stunting rankings at the national level in 2021 is West Kalimantan. The results of the Indonesian Nutrition Status Survey (SSGI) in 2021 stated that the stunting rate in West Kalimantan was 29.8%, higher than the average stunting rate at the national level of 24.4% [2]. This figure is 1.7% lower than the results of the 2019 Indonesian Nutrition Status Survey (SSGI) which was at 31.5%. The high level of stunting in West Kalimantan makes West Kalimantan one of the priority provinces to accelerate stunting control [3].

The problem of stunting is influenced by several factors that cause stunting, including the lack of exclusive breastfeeding, lack of access to clean drinking water in the household, lack of sanitation facilities, and low birth weight of babies born safely [4]. Therefore, to find

[DOI: 10.14710/JSUNIMUS.12.1.2024.1-9]

out the factors that cause stunting in each district/city in West Kalimantan, it is necessary to group according to the characteristics of the elements that cause stunting using cluster analysis. Cluster analysis is a clustering technique used to group objects according to the similarity of features in these objects [5].

One of the cluster analysis methods that is often used to perform clustering is the non-hierarchical method. K-Means is a commonly used non-hierarchical method and is relatively easy to implement. However, the sensitivity of cluster results to initialization or initial determination value at the cluster center is a weakness of K-Means [6]. To overcome the problems that occur, the K-Harmonic Means method developed, which is a development of the K-Means method. The K-Harmonic Means method aims to minimize the harmonic mean across all existing data points to all cluster centers. Compared to K-Means, the clustering produced by K-Harmonic Means is better because K-Harmonic Means is not too sensitive to the initial cluster center value (centroid) [6]. Another thing that needs to be considered in cluster analysis is the evaluation of clustering results. According to the data, clustering results are evaluated to get the most appropriate grouping. If the review is not done, it will affect the clustering results. This research estimates clustering results with an internal approach, namely the silhouette coefficient [7].

This study aims to analyze the clustering of districts/cities in West Kalimantan according to the factors that cause stunting using the K-Harmonic Means method and analyze the number of optimal clusters using the silhouette coefficient. The variables analyzed in this study are based on the factors that cause stunting in West Kalimantan in 2021, namely The percentage of households without access to clean drinking water (X_1), the rate of exclusive breastfeeding (X_2), percentage of babies with low birth weight who are born safely (X_3), the percentage of households without proper sanitation facilities (X_4). The clusters (K) analyzed in this study were 2, 3, and 4, using parameters (p) of 2, 3, and 4.

2. LITERATURE REVIEW

2.1. Data Standardization

Data standardization is carried out when the observed variables have significant differences in unit size. This significant difference in unit size can cause invalid cluster analysis calculations, so it is necessary to standardize the unit data to make the data units identical. Measures using data standardization can be seen in Equation 1 [8]:

$$\widehat{\boldsymbol{x}}_{\boldsymbol{i,j}} = \frac{\boldsymbol{x}_{\boldsymbol{i,j}} - \boldsymbol{\mu}_{\boldsymbol{j}}}{\boldsymbol{\sigma}_{\boldsymbol{j}}} \tag{1}$$

with N is the amount of observation data, μ_j is the average in the variabel j, σ_j is the standard deviation of variabel j and $\hat{x}_{i,j}$ is the standardization of the data i in variabel j.

2.2. Cluster Analysis

Cluster analysis is a multivariate analysis technique that aims to group objects according to their characteristics [9]. Cluster analysis has two assumptions: the sample taken must be representative, and the absence of multicollinearity [10].

1) Representative Sample

In research, it is essential to use a representative sample that can represent the population. The result obtained using a representative sample is that the research results can reflect the condition of the population to the maximum. But if the research only uses the entire population, then the representative sample is assumed to be fulfilled.

[DOI: 10.14710/JSUNIMUS.12.1.2024.1 -9]

2) No Multicollinearity

Multicollinearity is a significant correlation between two or more variables. In cluster analysis, it is necessary to pay attention to the multicollinearity problem because it has a considerable influence on determining the influence or effect of each variable and has the potential to affect the final result of clustering [11]. Multicollinearity calculation using the correlation coefficient calculation (Pearson product-moment correlation) is as follows.

$$\boldsymbol{r_{X_j,X_l}} = \frac{\boldsymbol{N}\left(\sum_{i=1}^{N} \boldsymbol{x_{i,j}x_{i,l}}\right) - \left(\sum_{i=1}^{N} \boldsymbol{x_{i,j}}\right)\left(\sum_{i=1}^{N} \boldsymbol{x_{i,l}}\right)}{\sqrt{\boldsymbol{N}\left(\sum_{i=1}^{N} \boldsymbol{x_{i,j}}^2\right)\left(\sum_{i=1}^{N} \boldsymbol{x_{i,j}}\right)^2}\sqrt{\boldsymbol{N}\left(\sum_{i=1}^{N} \boldsymbol{x_{i,l}}^2\right)\left(\sum_{i=1}^{N} \boldsymbol{x_{i,l}}\right)^2}}. \quad (2)$$

with $r_{X_jX_l}$ is the value of the correlation coefficient betwen variabel j and variable l, and N is the amount of observation data.

2.3. Distance Between Objects

Based on cluster analysis, objects with high similarity will be grouped into homogeneous groups. The distance between each object and other objects is measured [12]. In this study, the distance between objects is calculated using Euclidean distance. The following is the equation for calculating Euclidean distance.

$$\|\boldsymbol{x_i} - \boldsymbol{x_r}\| = \sqrt{\sum_{\boldsymbol{q=1}}^{\boldsymbol{M}} (\boldsymbol{x_{iq}} - \boldsymbol{x_{rq}})^2}, \quad (3)$$

where x_{iq} is the data on the object i in the variable q, x_{rq} is the data on the object r in the variable q and M is the number of variable [13].

2.4. K-Harmonic Means

K-Harmonic Means is a clustering technique based on the centroid as its basis. This method calculates the harmonic mean of the distance from each data point to the center point [13]. This method aims to minimize the harmonic mean of all existing data points to all cluster centroids. The value of the objective function in K-Harmonic Means is obtained by calculating the total harmonic mean at all data points to each cluster center point. Stages in the K-Harmonic Means method, namely [14]:

1. Randomly assign the initial centroid of each cluster.
2. Calculate the objective function value using Equation 4 below

$$\boldsymbol{KHM(X,C)} = \sum_{i=1}^{N} \frac{K}{\sum_{k=1}^{K} \frac{1}{\|x_i - c_k\|^p}}. \quad (4)$$

with p as the parameter. The value of p is determined by the research, which is usually $p \geq 2$.

3. Calculate the membership value $m(c_k|x_i)$ in each data (x_i) for all cluster center points (centroid) (C_k) according to Equation 5 below

$$\boldsymbol{m(c_k|x_i)} = \frac{\|\boldsymbol{x_i} - \boldsymbol{c_k}\|^{-p-2}}{\sum_{k=1}^{K} \|\boldsymbol{x_i} - \boldsymbol{c_k}\|^{-p-2}}. \quad (5)$$

4. For each data (x_i), the weight is calculated based on Equation 6 below

$$\boldsymbol{w(x_i)} = \frac{\sum_{k=1}^{K} \|\boldsymbol{x_i} - \boldsymbol{c_k}\|^{-p-2}}{\left(\sum_{k=1}^{K} \|\boldsymbol{x_i} - \boldsymbol{c_k}\|^{-p}\right)^2}. \quad (6)$$

[DOI: 10.14710/JSUNIMUS.12.1.2024.1 -9]

5. Recalculate the calculation on the cluster center point (centroid) (C_k) of all data (x_i) according to the membership value and wight owned by each data. According to Equation 7 below

$$\boldsymbol{c_k} = \frac{\sum_{i=1}^{n} \boldsymbol{m(c_k|x_i)w(x_i)x_i}}{\sum_{i=1}^{n} \boldsymbol{m(c_k|x_i)w(x_i)}}. \quad (7)$$

6. Return to steps 2 through 5 and repeat until the object function value does not changes significantly.
7. Assign the membership of data (x_i) into the cluster to the cluster center point (centroid) (C_k) at the membership value of (x_i) with (C_k).

2.5. Silhouette Coefficient

The silhouette coefficient method is one method of evaluating clustering results using internal criteria. The silhouette coefficient is used to evaluate the placement of each cluster object by measuring the average distance of objects in each cluster and estimating the average distance between objects with different clusters [15]. The steps for calculating the silhouette coefficient include [7].

1. Calculate the average distance of an i-th data to data located in the same cluster with Equation 8 below

$$\boldsymbol{a_i(k)} = \frac{\mathbf{1}}{\boldsymbol{n_k - 1}} \sum_{\boldsymbol{r=1}}^{\boldsymbol{n_k-1}} \left\|\boldsymbol{x_{i,k} - x_{r,k}}\right\|, \boldsymbol{r \neq i}. \quad (8)$$

with k = 1, 2, 3, ... , K, $a_i(k)$ is the average distance of the i-th data to the data in the same cluster, $\left\|x_{i,k} - x_{r,k}\right\|$ is the distance between the i-th data and the r-th data in the same cluster, $d_i(k)$ is the distance between the i-th data and all data in different cluster, $\left\|x_{i,k} - x_{r,l}\right\|$ is the distance between the i-th data and the r-th data in different clusters , $b(k)$ is the smallest value $d_i(k)$, $SC_{i,k}$ is the silhouette coefficient value of the i-th data, $SC(k)$ is the silhouette coefficient value of the k-th , SC is the global silhouette coefficient value, n_k is the number of the k-th cluster, n_l is the number of the l-th cluster, c_k is the k-th cluster, and x_i is the i-th observation data.

2. Calculate the average distance of an i-th data with all data located in a different cluster using Equation 9 below

$$\boldsymbol{d_i(k)} = \frac{\mathbf{1}}{\boldsymbol{n_k}} \sum_{\boldsymbol{r=1}}^{\boldsymbol{n_l}} \left\|\boldsymbol{x_{i,k} - x_{r,l}}\right\|. \quad (9)$$

Then the smallest value is taken using Equation 10

$$\boldsymbol{b(k)} = \boldsymbol{min}\{\boldsymbol{d_i(k)}\}. \quad (10)$$

3. Calculate the silhouette coefficient value on all i-th data. According to Equation 11

$$\boldsymbol{SC_{i,k}} = \frac{\boldsymbol{b(k) - a_i(k)}}{\boldsymbol{max}\{\boldsymbol{a_i(k), b(k)}\}}. \quad (11)$$

The average of $SC_{i,k}$ values for all data belonging to the cluster is formulated by Equation 12

[DOI: 10.14710/JSUNIMUS.12.1.2024.1 -9]

$$SC(k) = \frac{1}{n_k} \sum_{i=1}^{n_k} SC_{i,k} \,. \tag{12}$$

Next, the global SC value is calculated, obtained from the average calculation of the $SC(k)$ value in all clusters using Equation 13

$$SC = \frac{\sum_{k=1}^{K} (n_k \times SC(k))}{\sum_{k=1}^{K} n_k} \,. \tag{13}$$

The calculated silhouette coefficient value can vary between 0 and 1. If SC = 1, object x_i is already in the appropriate cluster. if SC = 0, object x_i needs to be clarifed whether it should be included in cluster A or B because its is between two clusters. The following is the global silhouette coefficient value range[10].

Table 1. Global Silhouette Cofficient Range

Global Silhouette Cofficient Range	Caption
$0{,}7 < SC \leq 1$	Strong Structure
$0{,}5 < SC \leq 0{,}7$	Medium Structure
$0{,}25 < SC \leq 0{,}5$	Weak Structure
$SC \leq 0{,}25$	No Structure

3. METHODOLOGY

The data used is secondary data on the factors that caused stunting in the district/city of West Kalimantan in 2021 obtained from the Health Office and the Central Statistics Agency of West Kalimantan Province. The data has been standardized to equalize the units of the data. The variables in this study are: The percentage of households without access to clean drinking water (X_1), the rate of exclusive breastfeeding (X_2), percentage of babies with low birth weight who are born safely (X_3), and the percentage of households without proper sanitation facilities (X_4).

4. RESULTS AND DISCUSSION

4.1. Overview of Stunting in Each District/City in West Kalimantan

The data characteristics of each causal factor of stunting in each district/city in West Kalimantan are presented in Table 2.

Table 2. Statistic Descriptive

Variable	Caption	Units	*Maximum*	*Mean*	*Minimum*
X_1	Households without access to clean drinking water	%	57,25	23,06	1,16
X_2	Exclusive breastfeeding	%	78,06	50,38	20,18
X_3	Low birth weight babies born safely	%	6,55	3,80	1,45
X_4	Households without proper sanitation facilities	%	37,15	22,89	3,52

Based on Table 2, the average percentage of households without access to clean drinking water is 23.06%, the highest percentage of households without access to clean

[DOI: 10.14710/JSUNIMUS.12.1.2024.1 -9]

drinking water is 57.25% in Ketapang District, and the lowest is 1.16% in North Kayong District. The average exclusive breastfeeding rate was 50.38%. The Sambas district had the highest percentage of 78.06%, while the Bengkayang district had the lowest percentage of 20.18%. The percentage of low birth-weight babies born safely averages 3.80%. Kota Singkawang had the highest percentage of low-birth-weight babies born safely, and Kabupaten Sanggau had the lowest at 1.45%. Then, the average percentage of households without proper sanitation facilities was 22.89%. Landak District has the highest percentage of households without proper sanitation facilities, with a value of 37.15%, and the lowest percentage of 3.52% is found in Pontianak City.

4.2. Clustering Using K-Harmonic Means Method

The clustering of districts/cities in West Kalimantan according to the causes of stunting in this study uses the K-Harmonic Means method. The first stage that must be done in this method is to determine the initial cluster center point randomly. Then, the objective function is calculated using Equation 3. The next step is calculating the membership and weight values to determine the new cluster center point using Equations 4 and 5. After obtaining the new cluster center point using Equation 6, the next step is to assign membership to the observation data for clusters of 2 using parameters 2, 3, and 4, as presented in Table 3.

Table 3. Results of Determining the Membership Value of Observation Data 2 Cluster

x_i	$p = 2$	$p = 3$	$p = 4$
1	2	2	1
2	1	1	2
3	1	1	2
4	2	2	1
5	1	1	2
6	1	1	2
7	1	1	2
8	1	1	2
9	1	1	2
10	1	1	2
11	2	2	1
12	2	2	1
13	2	2	1
14	2	2	1

Based on Table 3, the results obtained from the placement of observation data in cluster 2 using a parameter of 2 obtained that eight districts/cities are members of cluster 1 and 6 districts/cities are members of cluster 2. The results of placement using a parameter of 3 are eight districts/cities are members of cluster 1 and 6 districts/cities are members of cluster 2, and the results of placement using a parameter of 4 are six districts/cities are members of cluster 1 and 8 districts/cities are members of cluster 2. The steps for determining the membership of each cluster for clusters 3 and 4 can be carried out in the same way.

4.3. Silhouette Coefficient Method

The global silhouette coefficient values obtained using Equation (12) for clusters of 2,3 and 4 with parameters of 2,3 and 4 are presented in Table 4.

[DOI: 10.14710/JSUNIMUS.12.1.2024.1 -9]

Table 4. Comparison of Clustering Results Using Global SC Value

Number Of Clusters	Parameter (p)	*SC* global
2	2	0,040
	3	0,040
	4	0,052
3	2	0,591
	3	0,541
	4	0,541
4	2	0,774
	3	0,669
	4	0,681

Based on Table 4, the most significant global silhouette coefficient value is found in group 4 clusters using a parameter equal to 2, 0.744. So, the optimal clustering in grouping districts/cities in West Kalimantan using the K-Harmonic Means method is 4 clusters with a parameter equal to 2. The clustering results for 4 clusters using a parameter equal to 2 are presented in Table 5.

Table 5. Comparison of Clustering Results Using Global SC Value

Cluster	Number Of Cluster	Members In The Cluster
Very high	3	Bengkayang, Ketapang dan Melawi
High	4	Landak, Sanggau, Sintang dan Sekadau
Low	4	Mempawah, Kapuas Hulu, Kayong Utara dan Kubu Raya
Very Low	3	Sambas, Pontianak dan Singkawang

4.4. Characteristic of Each Cluster

The variable averages of each cluster presented in Table 6 can be used as a center to interpret the characteristics of each cluster.

Table 6. Average Value of Each Variable for Each Cluster

Variabel	Caption	Cluster of-K			
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
X_1	Households without access to clean drinking water	38,902	37,377	9,639	6,046
X_2	Exclusive breastfeeding	26,448	45,610	63,557	63,120
X_3	Low birth weight babies born safely	4,737	2,903	3,470	4,500
X_4	households without proper sanitation facilities	23,582	32,962	24,818	6,215

Cluster 1 is Bengkayang Regency. Bengkayang Regency is one of the districts/cities in West Kalimantan with the highest stunting rate in West Kalimantan due to the difficulty of getting clean water from polluted rivers. The contamination of river water makes it difficult for residents to utilize it for food and drinking needs. The lack of clean water makes it difficult for residents to get clean drinking water. Another factor that causes high stunting cases is the importance of exclusive breastfeeding for newborns until the age of 2 years. Lack of understanding of nutrition among pregnant women leads to premature birth and underweight babies. There needs to be cooperation between the central government, local governments, and

[DOI: 10.14710/JSUNIMUS.12.1.2024.1 -9]

the community in prioritizing districts/cities with very high stunting causal factors to accelerate the decline in stunting incidence.

Cluster 2 is categorized as a cluster with high stunting-causing factors. Cluster 2 has the highest percentage of households without proper sanitation compared to other clusters. Meanwhile, the percentage of households without access to clean drinking water and exclusive breastfeeding has a higher causal factor for stunting than clusters 3 and 4. Meanwhile, the percentage of low birth weight babies born safely is the lowest causal factor for stunting compared to other clusters. Landak District is a member of cluster 2 because there are still many people defecating in open places, causing contamination of the water used for daily life. Then, the lack of proper sanitation facilities causes the risk of stunting in children due to contaminated water and hurts community welfare.

Cluster 3 is a cluster with low contributing factors to stunting. Cluster 3 has fewer households without access to clean drinking water than Cluster 1 and Cluster 2. The percentage of safe birth weight is also lower compared to cluster 1. Mempawah District is a district that is a member of cluster 3. The low safe birth weight in Mempawah District is one of the government's hard work to reduce stunting cases in West Kalimantan. The number of programs conducted to reduce the incidence of stunting is one of the government's efforts that can be proud.

Cluster 4 is the cluster with the lowest causal factor for stunting. One of the members of cluster 4 is Pontianak City. The low number of households needing access to clean drinking water is due to easy access to clean drinking water. Proper sanitation facilities are also widely provided, leading to a lack of causal factors for stunting in this area. However, the safe birth weight of babies in Pontianak City is high compared to other clusters due to the suboptimal understanding of nutrition among pregnant women, which causes babies to weigh less than 2,500 grams. Exclusive breastfeeding is low, as most mothers in urban areas also work and choose to give formula milk rather than exclusive breastfeeding.

5. CONCLUSION

The clustering analysis results using the K-Harmonic Means method obtained cluster 1, a cluster with very high stunting causal factors. The most influential factors in cluster 1 are households without access to clean drinking water, exclusive breastfeeding, and low birth weight. The members of Cluster 1 are Bengkayang, Ketapang, and Melawi. Cluster 2 is a cluster with high stunting-causing factors. The most influential factor in cluster 2 is households without proper sanitation. Members of cluster 2 are Landak, Sanggau, Sintang, and Sekadau. Cluster 3 is a cluster with low causal factors for stunting; the most influential factor is exclusive breastfeeding. Members of cluster 3 are Mempawah, Kapuas Hulu, North Kayong, and Kubu Raya. Cluster 4 is a cluster with deficient causal factors for stunting. The most influential factors in this cluster are exclusive breastfeeding and households without proper sanitation facilities. Members of this cluster are Sambas, Pontianak, and Singkawang.

The optimal cluster formed based on the silhouette coefficient value is 4 clusters using $p = 2$. The silhouette coefficient value obtained is 0.744, the most significant value of the number of other clusters, and the cluster has entered into a strong cluster structure.

[DOI: 10.14710/JSUNIMUS.12.1.2024.1 -9]

REFERENCES

[1] Promkes.kemkes.go.id, "Mengenal Stunting dan Gizi Buruk. Penyebab, Gejala, dan Mencegah," *Promkes.kemkes.go.id*. https://promkes.kemkes.go.id/?p=8486 (accessed Nov. 20, 2022).

[2] V. B. Kusnandar, "10 Provinsi dengan Angka Stunting Tertinggi Nasional Tahun 2021," *databoks*, Jul. 12, 2022. https://databoks.katadata.co.id/datapublish/2022/07/12/10-provinsi-dengan-angka-stunting-tertinggi-nasional-tahun-2021 (accessed Oct. 20, 2022).

[3] N. Muharrami, "Angka Stunting KALBAR Ditargetkan 14% di Tahun 2024," *kalbarprov.go.id*, Jan. 24, 2022. https://kalbarprov.go.id/berita/angka-stunting-kalbar-ditargetkan-14-di-tahun-2024.html (accessed Oct. 20, 2022).

[4] Kementerian Kesehatan, *Buku Saku Pemantauan Status Gizi Tahun 2017*. Jakarta: Kementerian Kesehatan, 2018.

[5] A. Salsabila, T. Widiharih, D. Statistika, and F. Sains dan Matematika, "Metode K-Harmonic Means Clustering dengan Validasi Silhouette Coefficient (Studi Kasus : Empat Faktor Utama Penyebab Stunting 34 Provinsi di Indonesia Tahun 2018)," *JURNAL GAUSSIAN*, vol. 11, no. 1, pp. 11–20, 2022.

[6] I. gede A. Gunadi, "Analisis Cluster Pada Pengelompokan Siswa Diktuk Bintara Polri TA. 2018/2019, SPN Singaraja - Polda Bali Menggunakan K-Means dan K-Harmonic Means," *Jurnal Ilmiah SINUS*, vol. 17, no. 2, p. 13, Jul. 2019, doi: 10.30646/sinus.v17i2.421.

[7] D. I. Yunistya, R. Goejantoro, F. Deny, and T. Amijaya, "The Application Of K-Harmonic Means Method In District/City Grouping (Case Study: Poverty in Kalimantan Island in 2020) Penerapan Metode K-Harmonic Means Dalam Pengelompokan Kabupaten/Kota (Studi Kasus: Kemiskinan di Pulau Kalimantan Tahun 2020)," *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 1, pp. 51–64, 2022, doi: 10.20956/j.v19i1.21116.

[8] A. A. Dzikrullah, "Pengelompokan Provinsi Berdasarkan Kualitas Jaringan Internet Dengan Metode Centroid Linkage," 2022. [Online]. Available: http://www.ojs.unm.ac.id/jmathcos

[9] H. Usman and N. Sobari, *Aplikasi Teknik Multivariate untuk riset pemasaran*, 1st ed., vol. 1. Jakarta: Rajawali Pers, 2013.

[10] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 7th ed. New York: Pearson, 2010.

[11] Gujarati, N. Damodar, and C. P. Dawn, *Dasar-dasar Ekonometrika*, 5th ed. Jakarta: Salemba Empat, 2012.

[12] B. Zhang, M. Hsu, and U. Dayal, "K-Harmonic Means -A Spatial Clustering Algorithm with Boosting," 2001, pp. 31–45. doi: 10.1007/3-540-45244-3_4.

[13] Nicolaus, E. Sulistianingsih, and H. Perdana, "Penentuan Jumlah Cluster Optimal Pada Median Linkage Dengan Indeks Validitas Silhouette," 2016.

[14] S. Wijaya, M. A. Mukid, D. Ispriyanti, and S. Pengajar, "Pengelompokan Kabupaten/Kota di Jawa Tengah Menurut Kualitas Udara Ambien Menggunakan Analisis K-Harmonic Mean Cluster (Studi Kasus: Kualitas Udara Ambien pada Kawasan Pemukiman di Jawa Tengah Tahun 2015)," *Jurnal Ilmiah SINUS*, vol. 1, no. 1, pp. 978–602, 2018.

[15] U. Sa'adah, M. Y. Rochayani, D. W. Lestari, and Lusia. D.A, *Kupas Tuntas Algoritma Data Mining dan Implementasinya Menggunakan R*. Malang: UB Press, 2021.