PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol .04 No. 02, August 2025
Pages 329-338

Implementation of Cluster Analysis on Districts/Cities in Banten Province Based on Factors Causing Stunting in Toddlers

Isnaini Mahuda[1]*, Rofiroh[2], Meida Fitriani[3]

[1]Department of Statistics, Faculty of Engineering, Universitas Sultan Ageng Tirtayasa
Jenderal Sudirman St KM 3, Cilegon, 42435, Banten, Indonesia
[2]Department of Mathematics Education, Faculty of Teacher Training and Education,
Universitas Sultan Ageng Tirtayasa
Ciwaru Raya St, No.25, Serang, 42117, Banten, Indonesia
[3]Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Yogyakarta
Komojoyo St, No.21A, Sleman, 55281, Yogyakarta, Indonesia

E-mail Correspondence Author: *isnaini.mahuda@untirta.ac.id*

Abstract

Indonesia is a country in Southeast Asia that has the fifth highest prevalence of stunting in world and as a country that has a relatively high stunting rate. Banten Province, as one of provinces in Indonesia, is in the bottom five provinces with the worst stunting cases according to results of Indonesian Toddler Nutrition Status Survey (SSGBI) 2021. Many efforts have been made to resolve this stunting problem. The condition of stunting is usually characterized by the toddler's length or height being less than the normal toddlers of the same age. Several factors causing stunting can be identified so that the government can take appropriate steps to reduce stunting rates in Indonesia, especially in Banten province. Banten province consists of several districts/cities which have their own characteristics, so it is necessary to analyze the grouping of factors causing stunting based on districts/cities in Banten province using hierarchy cluster analysis. The cluster analysis method used in this research was Single Linkage method with Euclidian distance. The results of this research showed that two clusters were obtained based on the grouping results, where first cluster had a higher average percentage of babies born alive with Low Birth Weight (LBW) and percentage of average per capita monthly expenditure for food compared to second cluster. In other hand, percentage of households with access to adequate sanitation and drinking water in first cluster had smaller average value than second cluster.

Keywords: *Cluster Analysis; Single Linkage, Stunting*

: https://doi.org/10.30598/parameterv4i1pp329-337

1. INTRODUCTION

Stunting is a widely discussed health issue and has become the focus of attention for both the government and many people. Stunting is also known as a condition in which the growth and development of children aged 0-59 months are impaired or delayed. This period is the most crucial stage in human life. As is well known, the first 1,000 days of life (Hari Pertama Kelahiran/HPK) are critical phase that has long-term impacts and significantly influences the human life cycle. If these 1,000 days are successfully optimized, they will have a positive impact on the quality of human life, starting from early childhood, preschool age (5-6 years), adolescence, adulthood, and even old age [1].

The Indonesian Ministry of Health [2] states that stunting is a condition in which children under five years old (toddlers) have a length or height that is significantly lower than the average height of children of the same age. A normal toddler's length or height is typically above minus two standard deviations from the median growth standard set by World Health Organization (WHO). Children affected by stunting may face delays in intelligence, productivity, and academic achievement as they grow into adulthood. Even more concerning, stunting increases the likelihood of higher mortality rates, morbidity, and suboptimal cognitive and motor development [3]. Therefore, the issue of stunting is crucial and requires serious attention.

Based on global prevalence of stunted toddlers collected by WHO in 2020, there were 150.8 million cases, or 22.2% [4]. Indonesia is one of the Southeast Asian countries with the fifth-highest stunting prevalence in world and has a relatively high stunting rate compared to other middle-income countries, with a prevalence of 36.4% [5]. As of 2021, only six provinces had stunting rates lower than WHO's standard of 20%, such as Bangka Belitung Islands, Lampung, Riau Islands, Special Region of Yogyakarta, Jakarta, and Bali [6].

Banten Province is one of five provinces with the most severe stunting cases. According to the 2021 Indonesian Toddler Nutrition Status Survey (SSGBI), the stunting rate in Banten Province in 2021 was 24.5% [7]. Based on processed data from SSGBI 2021, Serang City, Cilegon City, Tangerang District, and Lebak District fall into the yellow stunting zone category, with a prevalence rate of 20% to 30%. South Tangerang City and Tangerang City are categorized as green stunting zones, with a prevalence of 10% to 20%. Meanwhile, Pandeglang District is classified as a red stunting zone, with prevalence rate of 37.8%, making that area is the highest number of stunting cases in Banten [8].

Several studies have identified factors that contribute to stunting in toddlers. One of them is studied by [9], which states that Low Birth Weight (LBW) is one of the causes of stunting. In that research, they concluded that with low birth weight (LBW) have undergone Intrauterine Growth Restriction, which leads to slower growth and development, often failing to meet the expected growth rate for their age after birth. This condition affects growth faltering, ultimately leading to stunting.

In addition to inherent factors such as birth weight, stunting can also be influenced by economic factors, which can be observed through consumption patterns and living environment conditions. As mentioned in the studied by [10], a significant variable affecting stunting is average value of household expenditure on food per capita. Moreover, access to safe drinking water and proper sanitation is also a contributing factor to stunting, as highlighted in studies by [11] and [12].

By identifying the factors causing stunting, a strategic approach can be developed to reduce stunting rates in Indonesia, particularly in Banten Province. Each region in Banten has its own unique characteristics, making it essential to analyze the grouping of

stunting-causing factors by districts/cities using cluster analysis. Cluster analysis is a technique used to categorize data into specific groups based on the similarity of objects within each group [13]. One of the methods in cluster analysis is the single linkage method, which calculates the minimum distance between two clusters or the nearest neighbor distance between an object in one cluster and an object in another cluster [14]. Through this grouping, the government is expected to gain a better understanding of the characteristics of each group, allowing for more precise and effective stunting prevention efforts.

2. RESEARCH METHODOLOGY

2.1. Data Sources and Research Variables

This study was quantitative research with secondary data gathered for analysis. The data was sourced from publication of the Statistics Indonesia (BPS) of Banten Province in 2022 and consisted of several variables. The variables used are shown in **Table 1**.

Table 1. Research Variables

Variable	Description
X_1	Percentage of live-born babies with Low Birth Weight (LBW)
X_2	Percentage of average per capita monthly expenditure on food
X_3	percentage of households with access to proper sanitation
X_4	Percentage of households with access to safe drinking water

2.2. Research Stages

The steps for data analysis in this study were as follows:

a. Collect data to be processed and conduct descriptive statistical analysis.
b. Investigate whether there is a significant correlation between variables using the Bartlett and KMO tests to assess data suitability.
c. Perform factor analysis to further examine variables that can represent groupings.
d. Calculate the distance of each variable using Euclidean distance. The Euclidean distance formula is:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2} \quad (1)$$

where $d(x,y)$ is the distance between objects x and y; x_i and y_i are the values of objects x and y for the i-th variable, and p is the number of variables.

e. Group districts/cities in Banten Province into clusters using the single linkage method.
f. Create a dendrogram generated from hierarchical cluster analysis using the single linkage method.
g. Calculate the average value of each variable for each cluster.
h. Interpret the results and make conclusions.

3. RESULTS AND DISCUSSION

The province of Banten consists of eight districts/cities, namely Tangerang City, Cilegon City, Serang City, and South Tangerang City, Pandeglang District, Lebak District, Tangerang District, and Serang District. The data used in this study was secondary data obtained from official website of BPS in Banten Province, consisting of variables in year 2022. Before conducting further data analysis, a summary of the descriptive statistics for each variable was presented. The summary of descriptive statistics is shown in **Table 2**.

Tablel 2. Descriptive Statistics

Variable	N	Minimum	Maximum	Mean
X_1	8	5.67	19.94	14.2600
X_2	8	41.16	63.90	53.7300
X_3	8	62.14	97.56	83.7263
X_4	8	73.27	99.08	91.4888

Based on **Table 2**, variable X_1 which represents the percentage of live-born babies with LBW in all districts/cities in Banten Province in 2022, had lowest value of 5.67% and highest value of 19.94%, with an average of 14.26%. Variable X_2 which represents the percentage of average monthly per capita expenditure on food in all districts/cities in Banten Province in 2022, had lowest value of 41.16% and highest value of 63.9%, with an average of 53.73%.

Meanwhile, variable X_3 which represents the percentage of households with access to proper sanitation, had lowest value of 62.14% and highest value of 97.56%, with an average of 83.7263%. Finally, variable X_4 which represents the percentage of households with access to safe drinking water, had lowest value of 73.27% and highest value of 99.08%, with an average of 91.4888%. Before conducting a cluster analysis, it was necessary to investigate whether the sample was sufficiently representative and adequate for further analysis, also to determine whether some variables had significant correlations. In this case, the KMO and Bartlett's tests were used, along with the Measure of Sampling Adequacy (MSA) values for each variable, to assess suitability of data. To determine presence of correlations, an independence test was conducted using Bartlett's test with following hypotheses:

Hypotheses:

H_0: $\rho = 1$, there is no correlation between variables.

H_1: $\rho \neq 1$, there is a correlation between variables.

The results of the KMO and Bartlett's tests are shown in **Table 3**.

Table 3. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.672
Bartlett's Test of Sphericity	Approx. Chi-Square	14.366
	Df	6
	Sig.	.026

Based on **Table 3**, the Kaiser-Meyer-Olkin (KMO) value was 0.672. Since the KMO value of 0.672 fell within range of 0.5 to 1, it could be concluded that the sample was representative for population, so that data was suitable for further analysis. Additionally,

the Bartlett's Test result showed a significant value of 0.026, which was less than $\alpha = 0.05$. This indicated that H_0 was rejected and H_1 was accepted, which led to conclusion that there was a correlation between variables. Since there was a correlation between variables, a Principal Component Analysis (PCA) was required to eliminate the correlation and ensure met assumptions for cluster analysis.

To determine whether variables were adequate for further analysis, this study used Measure of Sampling Adequacy (MSA) values derived from anti-image correlation matrix, which are shown in **Table 4**.

Table 4. Anti-image Matrices

		Measure of Sampling Adequacy
Anti-image Correlation	X_1	0.938
	X_2	0.771
	X_3	0.604
	X_4	0.646

Based on **Table 4**, the MSA values for each variable could be observed. If the MSA value is greater than 0.5, the variable is considered adequate for further analysis. Therefore, it could be concluded that all four variables were suitable for PCA. In PCA, one method for determine number of principal components is by selecting eigenvalues that greater than 1 **[14]**. The results of PCA are shown in **Table 5**.

Table 5. Principal Component Analysis

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	2.758	68.947	68.947
2	0.825	20.621	89.569
3	0.354	8.841	98.409
4	0.064	1.591	100.000

Based on **Table 5**, number of principal components with an eigenvalue greater than 1 was only one, namely the first principal component with an eigenvalue of 2.758. After conducting PCA, the principal component scores are obtained. These scores served as new variables that are uncorrelated, allowed to proceed to cluster analysis. The cluster analysis used in this study fell under hierarchical cluster analysis, specifically using the single linkage method. This hierarchical cluster analysis was employed because the dataset in this study is relatively small **[15]**.

The single linkage method is one of the approaches in hierarchical cluster analysis, based on the shortest distance between objects (minimum distance). The distance used in this method was Euclidean distance. **Table 6** presented a summary of the matrix from the Euclidean distance calculations.

Table 6. **The Euclidean Distance Calculations**

Case	1	2	3	4	⋮	8
1	0.000	1.352	2.918	2.289	⋮	4.169
2	1.352	0.000	3.085	2.772	⋮	4.498
3	2.918	3.085	0.000	1.724	⋮	2.045
4	2.289	2.772	1.724	0.000	⋮	3.266
⋮	...	...	...	...	⋱	...
8	4.169	4.498	2.045	3.266	⋮	0.000

The clustering process using the single linkage method was a grouping procedure based on the minimum distance. Therefore, the step that must be taken before this clustering process is to measure similarity or proximity using a predetermined distance metric. The next step was to find the closest object or data in the Euclidean distance matrix and then merge them into a new cluster. The results of new distance matrix are shown in **Table 7**.

Table 7. **The New Distance Matrix Single Linkage**

	12	376548
12	0.000	25.416
376548	25.416	0.000

The clustering process illustrated in the form of a dendrogram showed clustering results of the factors causing stunting in infants in the districts/cities of Banten Province in year 2022 using the single linkage method as shown in **Figure 1**.

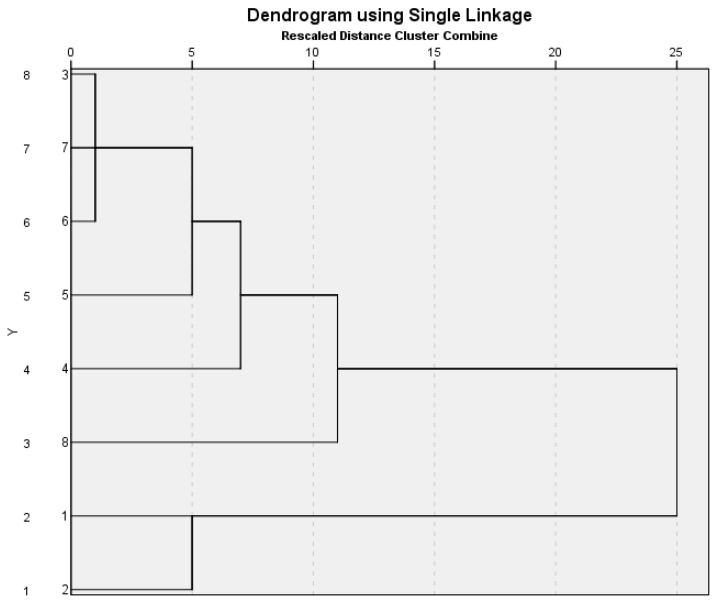

Figure 1. **Dendrogram Single Linkage**

Based on **Figure 1**, it could be seen that using the single linkage method resulted in two clusters from the clustering process. Cluster 1 consists of 2 districts, while Cluster 2 consists of 6 districts/cities. The details of the districts/cities in each cluster were as follows:

- Cluster 1: Pandeglang District and Lebak District.
- Cluster 2: Tangerang District, Serang District, Tangerang City, Cilegon City, Serang City, and South Tangerang City.

After determining the number of clusters and their members, next step is to identify the distinctive features or characteristics of each cluster to describe its contents. Identification in this study was conducted through descriptive analysis by examining the average values of variables in each cluster. The results of average values for each cluster are shown in **Table 8**.

Table 8. Average Cluster Values Using the Single Linkage Method

Variable	Cluster 1	Cluster 2
X_1	17.56	13.16
X_2	60.99	51.31
X_3	64.77	90.045
X_4	75.27	96.895

The average calculations for each cluster indicated differences between the two clusters formed using the single linkage method. Based on **Table 8**, each cluster has distinct characteristics. Cluster 1 had a higher average value for the percentage of live-born babies with LBW (X_1) and the percentage of average per capita monthly expenditure on food (X_2) compared to Cluster 2. However, in terms of the percentage of households with access to proper sanitation (X_3) and the percentage of households with access to safe drinking water (X_4), Cluster 1 had lower average values than Cluster 2.

Based on these characteristics, Cluster 1 was identified as a cluster with a high stunting factor. This was also supported by data from the 2022 BPS Banten Province publication, which stated that the districts included in that cluster had higher prevalence rate compared to other districts and cities in Banten Province.

Given these shared characteristics, local governments in Cluster 1 can implement policies such as early education for pregnant women to maintain a balanced diet and monitor fetal growth and development to ensure a normal birth weight. The government can also implement the Supplementary Feeding (Pemberian Makanan Tambahan/PMT) program for pregnant women, as research by **[16]** suggests that pregnant women who do not receive PMT have a 10.046 times higher risk of having stunted toddlers compared to those who receive PMT, with a classification accuracy of 74%.

In addition, the government can focus on improving sanitation and access to safe drinking water for local communities. This program can be implemented through the optimization of village funds for sanitation improvements and increased accessibility to clean water **[17]**.

Cluster 2 is characterized by a relatively low percentage live-born babies with LBW, indicating that pregnant women in the districts and cities within this cluster receive adequate attention and can maintain their pregnancy health. In addition, access to drinking water and sanitation is already considered adequate, as reflected in the high average values for both variables. Although some districts and cities within this cluster are still classified as yellow zones for stunting in Banten Province, with proper efforts and policies from the government to address infant stunting, the prevalence in these areas can be reduced. In the future these areas could transition into the green zone, like Tangerang City and South Tangerang City.

4. CONCLUSION

The results of this study concluded that the implementation of cluster analysis using the single linkage method on districts/cities in Banten Province based on the factors causing stunting can be described as follows:

1. Using the single linkage method, two clusters were formed from the clustering process.
2. Cluster 1 consists of two districts: Pandeglang Districts and Lebak districts, while Cluster 2 consists of six districts/cities: Tangerang Districts, Serang Districts, Tangerang City, Cilegon City, Serang City, and South Tangerang City.
3. Cluster 1 has higher average values for the percentage of live-born babies with low birth weight (BBLR) and the percentage of average per capita monthly food expenditure compared to Cluster 2. However, in terms of the percentage of households with access to proper sanitation (X_3) and the percentage of households with access to safe drinking water (X_4), Cluster 1 has lower average values than Cluster 2. Based on these characteristics, Cluster 1 is identified as the cluster with a high stunting factor.
4. Meanwhile, Cluster 2 is characterized by a relatively low percentage of babies with BBLR and adequate access to drinking water and sanitation. Therefore, Cluster 2 can be considered to have relatively good characteristics in minimizing the factors that contribute to stunting in toddlers.

REFERENCES

[1] C. K. Herbawani, U. Q. Karima, M. N. H. Syah, A. N. Hidayati, and B. Aprianto, "Analisis Determinan Kejadian Stunting di Wilayah Kerja Puskesmas Cinangka, Kota Depok," *Ghidza J. Gizi dan Kesehat.*, vol. 6, no. 1, pp. 64–79, 2022, doi: 10.22487/ghidza.v6i1.518.

[2] Kemenkes RI, "Infodatin : Pusat Data Dan Informasi Kementerian Kesehatan RI Situasi Balita Pendek," 2016.

[3] A. Adityaningrum, N. Arsad, and H. Jusuf, "Faktor Penyebab Stunting di Indonesia : Analisis Data Sekunder Data SSGI Tahun 2021," *Jambura J. Epidemiol.*, vol. 3, no. 1, pp. 1–10, 2023.

[4] Panigoro, "Upaya Pencegahan Dan Penanggulangan Stunting Pada Balita Di Wilayah Kerja Puskesmas Tilongkabila," *J. Ilmu Kesehat. dan Gizi*, vol. 1, no. 1, pp. 79–91, 2020.

[5] I. S. Sukamto, S. Juwita, and N. B. Argaheni, "Upaya pencegahan dan penanganan stunting dengan pengenalan program stunting melalui kader di Kota Surakarta," *JMC J. Midwifery Community*, vol. 1, no. 2, pp. 11–23, 2023, [Online]. Available: https://jurnal.uns.ac.id/jmc/index

[6] Kemenkes RI, "Profil Kesehatan Indonesia 2021," 2022. [Online]. Available: Pusdatin.Kemenkes.Go.Id

[7] A. Setiawan and R. Muttaqin, "Sinergitas dalam Percepatan Penanganan Kasus Stunting di Kota Serang," *JIIP J. Ilm. Ilmu Pemerintah.*, vol. 8, no. 2, 2023.

[8] Kompas.com, "Banten Duduki Peringkat Kelima Angka Stunting Terbanyak di Indonesia." [Online]. Available: https://regional.kompas.com/read/2022/03/07/191640578/banten-duduki-peringkat-kelima-angka-stunting-terbanyak-di-indonesia-ada

[9] T. R. Sinaga, S. D. Purba, M. Simamora, J. A. Pardede, and C. Dachi, "Berat Badan Lahir Rendah dengan Kejadian Stunting pada Batita," *J. Ilm. Permas J. Ilm. STIKES Kendal*, vol. 11, no. 3, pp. 493–500, 2021, doi: 10.32583/pskm.v11i3.1420.

[10] A. C. A. Pangaribuan, K. D. Dhanutama, M. O. Wijaya, P. T. Navasha, and R. Nooraeni, "Analisis Determinan Balita Pendek Dan Sangat Pendek Di Indonesia 2015-2018 Dengan Regresi Data Panel," *J. Pendidik. Ekon.*, vol. 8, no. 2, pp. 56–61, 2020, doi: 10.26740/jupe.v8n2.p56-61.

[11] A. J. Pitoyo, A. Saputri, R. E. Agustina, and T. Handayani, "Analysis of Determinan of Stunting Prevalence among Stunted Toddlers in Indonesia," *Populasi*, vol. 30, no. 1, p. 36, 2022, doi: 10.22146/jp.75796.

[12] R. A. Wicaksono, K. S. Arto, E. Mutiara, M. Deliana, M. Lubis, and J. R. L. Batubara, "Risk factors of stunting in indonesian children aged 1 to 60 months," *Paediatr. Indones. Indones.*, vol. 61, no. 1, pp. 12–19, 2021, doi: 10.14238/pi61.1.2021.12-9.

[13] R. A. Faujia, E. S. Setianingsih, and H. Pratiwi, "Analisis Klaster K-Means Dan Agglomerative Nesting Pada Indikator Stunting Balita Di Indonesia," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 1249–1258, 2022, doi: 10.34123/semnasoffstat.v2022i1.1511.

[14] D. R. Thamrin and D. Murni, "Analisis Cluster Hierarki Metode Single Linkage Pada Kabupaten/Kota di Provinsi Sumatera Barat Berdasarkan Indikator Kesehatan," *J. Math. UNP*, vol. 7, no. 3, p. 45, 2022, doi: 10.24036/unpjomath.v7i3.12988.

[15] C. Wirdiastuti and H. Helma, "Pengelompokkan Sembilan Kabupaten/Kota di Provinsi Sumatera Barat Berdasarkan Tingkat Kriminalitas dengan Menggunakan Analisis Gerombol," *UNP J. Math.*, vol. 4, no. 2, pp. 1–4, 2019, [Online]. Available: http://ejournal.unp.ac.id/students/index.php/mat/article/view/6296

[16] D. C. Mardlatilla and I. D. Ratih, "Analisis Faktor-Faktor yang Berpengaruh terhadap Stunting Menggunakan Regresi Data Logistik Biner (Studi Kasus : Desa Jongbiru Kecamatan Gampengrejo Kabupaten Kediri)," *J. Sains dan Seni ITS*, vol. 11, no. 2, 2022, doi: 10.12962/j23373520.v11i2.75814.

[17] D. Satriawan, "Pengelompokan Provinsi Di Indonesia Berdasarkan Faktor Penyebab Balita Stunting," *Bul. Penelit. Sist. Kesehat.*, vol. 24, no. 4, pp. 308–317, 2021, doi: 10.22435/hsr.v24i4.4341.