Faktor Exacta

Vol. 16, No. 2, June 2023, pp. 161~169
eISSN: 2502-339X, pISSN: 1979-276X, DOI: 10.30998/faktorexacta.v16i2.17186

Clustering Indonesian Provinces on Prevalence of Stunting Toddlers Using Agglomerative Hierarchical Clustering

Septian Wulandari
Department of Informatic Engineering, Universitas Indraprasta PGRI, Indonesia

Article Info

Article history:

Received Apr 05, 2023
Revised Jul 03, 2023
Accepted Jul 04, 2023

Keywords:

Prevalence of Stunting
Agglomerative Clustering
Ward's
Average Linkage

ABSTRACT

Stunting is a chronic nutritional problem caused by a lack of nutritional intake in toddlers. Indonesia is the 5th country with the highest cases of toddler nutrition experiencing stunting at 30.8% in 2018. The current problem, in Indonesia, is providing complete immunization and fulfilling child nutrition in each province is still low. Data obtained from 2018 to 2022 still toddlers who are malnourished and obese and there is no province grouping based on characteristics such as malnutrition, obesity, short toddlers, and complete basic immunization. Clustering is grouping objects into a group so that one cluster contains objects that are similar and different from other objects in other clusters. The agglomerative hierarchical clustering method can classify provinces based on the characteristics that cause stunting so that it can be used as a basis for early prevention for the Indonesian government to tackle stunting and can reduce stunting growth rates which continue to increase and can experience a decline. The agglomerative hierarchical clustering method used is the Average Linkage and Ward's algorithms with the data used is the prevalence of stunting taken in 34 provinces in Indonesia with 11 data attributes. The results of this study are that there are two clusters, namely Cluster 1 which has a relatively high prevalence of stunting with members of 13 provinces, and Cluster 2 which has a relatively low prevalence of stunting with members of 21 provinces. The highest chopenetic correlation value is in Ward's algorithm with a value of 0.8399978. So, it can be said that Ward's algorithm is better than the Average Linkage algorithm in clustering provinces in Indonesia on the prevalence of stunting toddlers.

Corresponding Author:

Septian Wulandari,
Department of Informatic Engineering,
Universitas Indraprasta PGRI,
Jl. Nangka No. 58 C, Tanjung Barat, Jagakarsa, Jakarta Selatan.
Email: septian.pmb09@rocketmail.com

1. INTRODUCTION

Stunting is a chronic nutritional problem caused by a lack of nutritional intake over a long period of time. Stunting generally occurs in toddlers (under the age of five) and is characterized by a short body shape compared to toddlers of the same age [1]. Stunting can occur while still in the womb or the form of a fetus and is only seen when the child is two years old. If a toddler has indicated stunting and is not balanced on catch-up growth, it can lead to decreased growth. Stunting is a public health problem that can result in a soaring risk of morbidity, mortality, and impairment in toddler growth both motorically and mentally [2]. Stunting or similar nutritional problems can be avoided or prevented, namely in the first 1,000 days of life or after birth, and other steps such as providing additional food and fortification of iron in food. Other stunting prevention can also be

done by fulfilling nutrition from the womb, providing exclusive breastfeeding from newborns to children aged 2 years, always monitoring child growth and development, including breast milk and complementary foods, and maintaining cleanliness in the environment.

Based on data on the prevalence of stunting in toddlers in 2020 according to WHO, 22% or approximately 149.2 million toddlers in the world experience stunting [3]. The percentage of stunting prevalence in toddlers in Indonesia in 2015 reached 36.4% [2]. This means that there is a problem of growth that is not maximized in approximately 8.9 million toddlers in Indonesia or about 1 in 3 toddlers experiencing stunting [4]. Moreover, according to the Basic Health Research (Riskesda) Indonesia is in the top five countries with nutritional cases of stunted toddlers at 30.8% in 2018, which has decreased compared to 2013, which was around 37.2% [5]. Stunting that occurs in toddlers has a considerable effect on children's health in the present and the future. Therefore, stunting is a serious problem that must be addressed immediately by the Indonesian government so that the increasing stunting growth rate can decrease. In this regard, the government has made regulations by making stunting the main priority in eradicating it among the community. Stunting must be addressed because indications of stunting have an impact on the development of human resources in Indonesia so that toddlers avoid stunting must be considered since they are born even from the womb. To overcome stunting in Indonesia, this study conducted clustering on stunting prevalence data in provinces in Indonesia so that it is expected to provide information for the Indonesian government in implementing programs related to handling stunting prevalence cases.

Clustering is one of the clustering techniques in data mining. Clustering is the grouping of objects into a group so that one cluster contains objects that have similarities and are different from other objects in other clusters [6]. Cluster analysis is a method of multivariate statistical analysis that is intended for clustering objects into a cluster based on the characteristics they have so that objects in one cluster have the same or homogeneous characteristics compared to objects in other clusters [7]. There are two types of methods in clustering, namely hierarchical clustering and partitioning clustering. In hierarchical clustering hierarchical clustering has the advantage of combining data and creating a hierarchy, data that has similarities will be placed in a hierarchy that has proximity and those that do not have similarities are placed in a hierarchy that is far apart [8]. The hierarchical clustering method is carried out using an approach by developing a binary tree base which is often referred to as a dendrogram. The clustering hierarchy is developed into two methods, namely agglomerative and divisive clustering. The agglomerative hierarchical clustering method starts by taking a single cluster (which contains only one data object per cluster) at the bottom level and continues to combine the two clusters at once and builds a bottom-up hierarchy of clusters. There are four algorithms in the agglomerative hierarchical grouping namely Single Linkage, Complete Linkage, Average Linkage, and Ward's [1]. The hierarchical clustering method is approached by developing a binary tree base which is often referred to as a dendrogram. Hierarchical clustering developed into two methods, namely agglomerative and divisive clustering. The agglomerative hierarchical clustering method starts by taking a single cluster (which contains only one data object per cluster) at the bottom level and continues to combine two clusters at a time and build a bottom-up hierarchy of clusters. There are four algorithms in agglomerative hierarchical clustering, namely Single Linkage, Complete Linkage, Average Linkage, and Ward's [9]. The advantages of Average Linkage are that it does not require determining the number of clusters, the dendrogram results provide a graphical representation, and our ability to detect various cluster shapes and sizes [10]. Meanwhile, the advantage of Ward's algorithm is that it provides results that facilitate clustering using hierarchical methods because it can Minimize the Number of Squares (SSE) is the sum of the squared differences between each observation and its group's mean [9].

There is another study on stunting clustering conducted by Maulina Rizky Anggraeni, et al in 2023 by clustering the prevalence of stunting in Tegalrejo Health Center using agglomerative hierarchal clustering Average Linkage algorithm and obtained there are additional villages that have a high prevalence of stunting [3]. Another study was also conducted by Musarrafah Paramadina, et al in 2019 by comparing the cluster analysis of the Average Linkage method and Ward's method in the case of South Sulawesi Province HDI and obtained 2 clusters with the highest Dunn index value occurring in the Average Linkage method [9]. Research conducted by M Rais Ridwan and Heri Retnawati in 2021 by conducting an application of cluster analysis using the agglomerative method resulted in 4 clusters with Ward's method being more accurate than the average linkage method [11]. So, in this study, clustering of provinces in Indonesia on the prevalence of stunting among children under five years of age was carried out using agglomerative hierarchical clustering with Ward's and Average Linkage algorithms

2. METHODS

In this research, the approach used is a quantitative descriptive literature study. Literature studies were carried out by collecting literature references that will support the completion of this research. In the

quantitative descriptive stage, this research is carried out by processing data, analyzing data, and interpreting data according to the needs of this research. Meanwhile, the technique used in this study was direct observation In the first stage, data collection in this study was obtained from BPS data which can be accessed at https://www.bps.go.id. Stunting prevalence data are taken from 2018 to 2022 in 34 provinces in Indonesia with 11 data attributes, namely:

x_1: Special index for handling stunting
x_2: Percentage of infants less than 6 months of age who receive exclusive breastfeeding
x_3: Percentage of short and very short toddler
x_4: Percentage of provinces achieving 80% complete basic immunization in infants
x_5: Percentage of obese toddlers (bb/tb) aged 0-59 months
x_6: Prevalence of malnourished toddler (0-59 months)
x_7: Prevalence of malnourished toddler (0-23 months)
x_8: Prevalence of underweight toddler (0-59 months)
x_9: Prevalence of underweight toddler (0-23 months)
x_{10}: Prevalence of undernourished children (0-59 months)
x_{11}: Prevalence of undernourished children (0-23 months)

There are two assumptions that must be met in clustering analysis such as representative samples and there must be no multicollinearity between each variable [12]. To test for multicollinearity, it can be done by calculating the Variance Inflation Factors (VIF) value [13]. Here is the formula for calculating the value of Variance Inflation Factors (VIF) [12]:

$$VIF_i = \frac{1}{(1-R_i^2)} \quad (1)$$

with R_i^2 is the coefficient of determination of the variable i. It is said that there is multicollinearity when the value $VIF > 10$ or if the Tolerance value $> 0{,}10$ then there is no multicollinearity.

To determine the adequacy of the data, the Kaiser Meyer Olkin (KMO) test was conducted. Then, the Bartlett Spherecity Test is carried out to test whether there is a relationship between each variable. The following is the formula for calculating KMO [14]:

$$KMO = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2}{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}^2} \quad (2)$$

with:
i: 1,2,3, … , p
j: 1,2,3, … , p, where $j \neq i$
r_{ij}^2: simple correlation coefficient between variables i and j
a_{ij}^2: partial correlation coefficient between variables i and j

Meanwhile, the method used in this research is the agglomerative hierarchical clustering method. The hierarchical clustering method has the advantage of being able to combine data and create a hierarchy, data that has similarities will be placed in a hierarchy that has proximity and those that do not have similarities are placed in a hierarchy that is far apart [8].

Hierarchical clustering method was developed to overcome the weakness of partitioning clustering. Hierarchical clustering was developed into two methods, namely agglomerative and divisive clustering. The agglomerative hierarchical clustering method starts by taking a single cluster (which contains only one data object per cluster) at the bottom level and continues to merge the two clusters together and build a bottom-up hierarchy of clusters. A cluster hierarchy can be interpreted using binary tree terminology. The root represents the entire set of data objects to be clustered and forms the base of the hierarchy (level 0). at each level, child entries (nodes) that are subsets of the entire dataset correspond to clusters. The entries in each of these clusters can be determined by traversing the tree from the current cluster node to the base singleton data point. Each level in the hierarchy corresponds to some set of clusters. The base of the hierarchy consists of all singleton points that are leaves of the tree. This cluster hierarchy is also called a dendrogram. Algorithms that belong to agglomerative hierarchical clustering are Single Linkage, Complete Linkage, Average Linkage, and Ward's [1]. These four algorithms have differences in the clustering process. Ward's algorithm is used to calculate the distance between two clusters during agglomerative hierarchical lustering. Stages in Ward's algorithm every number of groups will be reduced by one group by combining the two groups that show the smallest progress or increase in the total number of squares in the group. Ward's algorithm uses a calculation by maximizing the homogeneity of The Sum of Squares Error (ESS) in one group, namely the formula [13]:

$$ESS = \sum_{j=i}^{k}\left(\sum_{i=1}^{n_j} x_{ij}^2 - \frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}^2\right)\right) \quad (3)$$

where:
x_{ij}: object value in i
i : $1, 2, 3, \ldots, n$ in the j group

k : number of groups per stage
n_j : number of group of i in a group of j

The average linkage algorithm is a grouping of objects based on the average distance between each pair of objects. The stages of the Average Linkage algorithm are calculating the matrix $D = \{d_{ik}\}$ to obtain objects that have the most similarity between one another such as U and V. Then, calculate the distance between UY and W with the formula [13]:

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{UV} N_W} \quad (4)$$

with $d_{ik}$is the distance between objects to i object in the cluster UV with the object k in the cluster W, while N_{UV} and N_W are the numbers of attributes in the cluster (UV) and W.

There are no specific rules for determining the number of clusters in agglomerative hierarchical clustering, but there are some guidelines that can be used, namely [15]

1. If the objective of clustering is only to identify a section of the market, management may want to cluster a certain number of clusters such as 3, 4, and 5.
2. The relative size of the cluster should be useful and beneficial.
3. In the cluster results that have been formed, further testing is carried out, namely with the silhouette coefficient to be able to determine the strength and superiority when objects are placed in a particular cluster. The size of the silhouette coefficient value can be seen in Table 1 [16].

Table 1. Measure of Sillhouette Coefficient Value

Scale	Description
$0{,}7 < SC \leq 1$	Strong Structure Medium
$0{,}5 < SC \leq 0{,}7$	Medium Structure
$0{,}25 < SC \leq 0{,}5$	Weak Structure
$SC \leq 0{,}25$	No Structure

3. RESULTS AND DISCUSSION

Based on the descriptive analysis in this study, it was found that the highest stunting index was in DI Yogyakarta province, which amounted to 92.94% and the lowest was in Maluku province, which amounted to 50.91%. The percentage of infants aged less than 6 months who received exclusive breastfeeding was highest in North Maluku province at 64.8% and lowest in North Sumatra province at 25.69%. The percentage of short and very short toddlers was highest in West Sulawesi province at 16.2% and lowest in Bali province at 0.25%. The percentage of districts/cities achieving 80% complete basic immunization in infants was highest in Lampung, Riau Islands, DKI Jakarta, Central Java, Yogyakarta, Bali, and West Nusa Tenggara provinces at 100% and lowest in East Nusa Tenggara province at 4.55%. The percentage of obese children under five (BB/TB) by age group 0-59 months was highest in Papua province at 13.2% and lowest in West Nusa Tenggara province at 3.3%.

Descriptive analysis in this study for the percentage of prevalence of malnourished toddlers based on the age group 0-59 months was highest in Maluku province at 7.4% and lowest in Bali province at 2%. Meanwhile, the prevalence of malnourished children under five years of age 0-23 months was highest in Maluku province at 10.3% and lowest in DI Yogyakarta province at 1.4%. The prevalence of underweight children aged 0-59 months was highest in East Nusa Tenggara province at 22.2% and lowest in Bali province at 9.8%. Meanwhile, the prevalence of underweight children aged 0-23 months was highest in East Nusa Tenggara province at 17.6% and lowest in Bali province at 7.4%. The prevalence of undernutrition among children under five years of age 0-59 months was highest in East Nusa Tenggara province at 29.5% and lowest in Riau Islands province at 13%. Meanwhile, the prevalence of under-five malnutrition by age group 0-23 months was highest in East Nusa Tenggara province at 24.5% and lowest in West Java province at 10.6%.

In cluster analysis, the clustering process must fulfill the variables to be observed. Therefore, a multicollinearity test must be performed by calculating the Variance Inflation Factors (VIF) value. Multicollinearity test processing results are obtained in Table 2

Table 2. Multicollinearity Test

Independent variable	Collinearity Statistics	
	Tolerance	VIF
x_1	0,723	1,383
x_2	0,778	1,285
x_3	0,654	1,528

x_4	0,675	1,482
x_5	0,943	1,061
x_6	0,550	1,818
x_7	0,751	1,332
x_8	0,642	1,559
x_9	0,731	1,367
x_{10}	0,585	1,711
x_{11}	0,884	1,132

Based on Table 2, it can be seen that all independent variables have the following values $VIF < 10$ and Tolerance Value $> 0{,}10$. So it can be said that there is no multicollinearity between these variables.

To determine the adequacy of the data, the Kaiser Meyer Olkin (KMO) test was conducted. This is done to determine whether all the data that has been taken is sufficient for factoring and a homogeneity test is carried out using the Barlett Spherecity Test to determine the relationship between each variable. If the KMO value is> 0.5, it can be assumed that the data processed or analyzed has met the assumptions on data sufficiency. Based on the KMO test, the KMO value is 0.523, meaning that the data analyzed has met the data sufficiency.

Table 3. Barlett Spherecity Test

	Chi-Square	p-value
Barlett Spherecity	69,789	0,351

Based on Table 3, the Chi-square value is 69.789 at a significant level of 5%, the p-value is 0.351. If the significant value of p-value > 0.05, it can be concluded that the data variance is homogeneous, meaning that the homogeneity assumption test is fulfilled.

The next step is to cluster provinces using agglomerative hierarchical clustering on the prevalence of stunting among children under five in Indonesia. The algorithm used is Ward's algorithm. The clustering process is done with the help of R Studio software. Clustering results using Ward's algorithm and Average Linkage. Clustering is done by calculating the Euclidean distance using Ward's algorithm and average linkage which produces chopenetic correlation with the highest value between the two algorithms. Based on the calculation of chopenetic correlation, the results are shown in Table 4.

Table 4. Chopenetic Correlation Value

Algorithm	Chopenetic Correlation
Average Linkage	0.5103928
Ward's	0.8399978

Table 4 shows that the highest chopenetic correlation value is in Ward's algorithm with a value of 0.8399978. Therefore, Ward's algorithm is used in this study. Furthermore, clustering from 2 clusters up to 10 clusters and the Sillhouette Coefficient value is calculated and the calculation results are obtained as in Table 5 and Figure 1.

Table 5. Sillhouette Coefficient Calculation Results

Numbers of Clusters	2	3	4	5	6	7	8	9	10
Sillhouette Coefficient Value	0,317	0,185	0,158	0,117	0,103	0,06	0,066	0,059	0,058

 DOI: 10.30998/faktorexacta.v16i2.17186

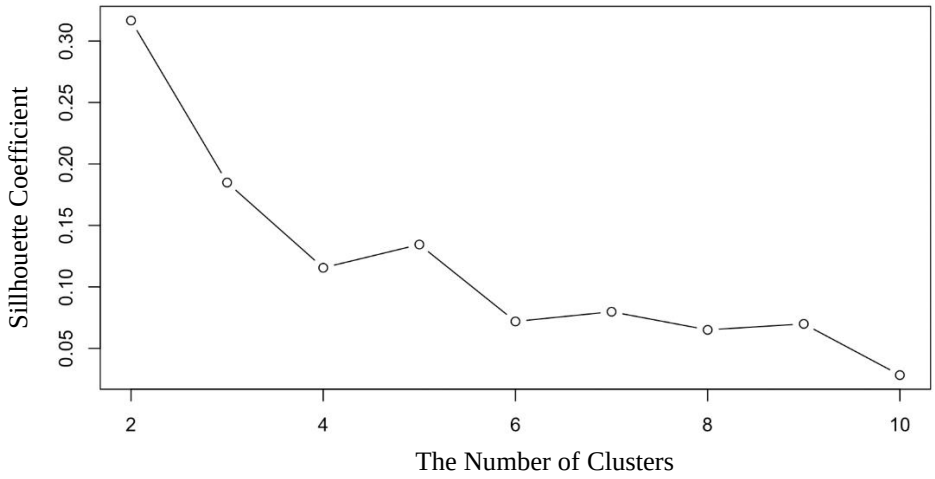

Figure 1. Graph of Sillhouette Coefficient Calculation Results

In Table 5 and Figure 1, it is clear that the highest Sillhouette Coefficient value is in the number of clusters, namely 2 clusters with a Sillhouette Coefficient value of 0.317 which is included in the weak structure category, meaning that there is a weak bond between each object and the cluster that has been formed. So, the clustering of provinces in Indonesia on the prevalence of stunting among children under five is grouped into 2 clusters. To facilitate the clustering process, labels are given to each province which can be seen in Table 6.

Table 6. Labeling of Provinces in Indonesia

No	Province	No	Province	No	Province	No	Province
1	Aceh	10	Riau Islands	19	East Nusa Tenggara	28	Southeast Sulawesi
2	North Sumatra	11	DKI Jakarta	20	West Kalimantan	29	Gorontalo
3	West Sumatra	12	West Java	21	Central Kalimantan	30	West Sulawesi
4	Riau	13	Central Java	22	South Kalimantan	31	Maluku
5	Jambi	14	DI Yogyakarta	23	East Kalimantan	32	North Maluku
6	South Sumatra	15	East Java	24	North Kalimantan	33	West Papua
7	Bengkulu	16	Banten	25	North Sulawesi	34	Papua
8	Lampung	17	Bali	26	Central Sulawesi		
9	Kep. Bangka Belitung	18	West Nusa Tenggara	27	South Sulawesi		

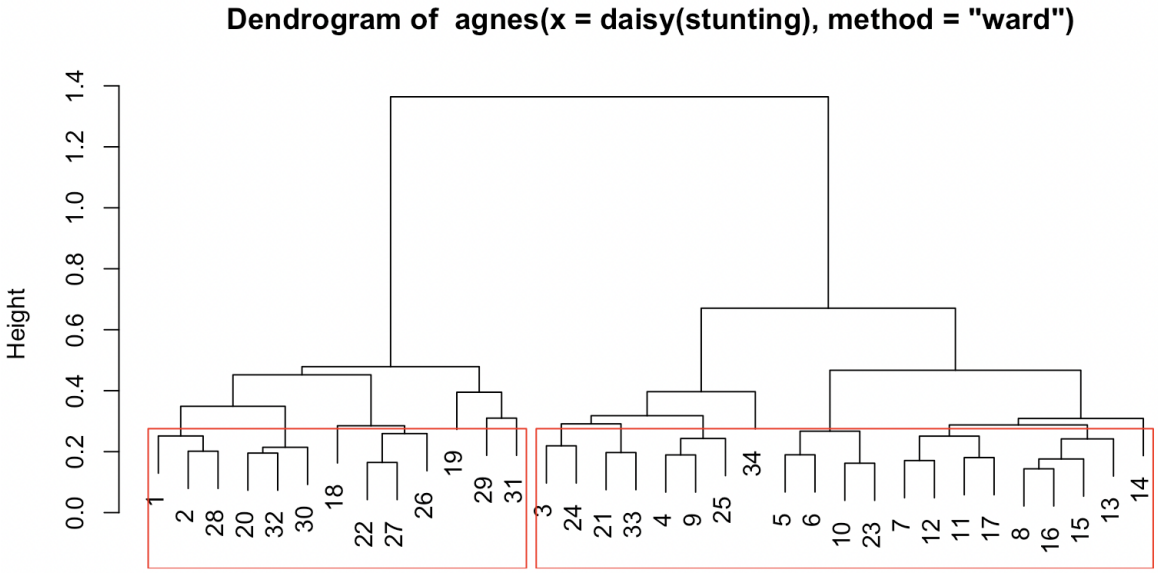

Figure 2. Ward's Algorithm Dendrogram

Figure 2 is the result of clustering using agglomerative hierarchical clustering using Ward's algorithm. the dendrogram above shows that there are 2 clusters with cluster 1 as many as 13 provinces and cluster 2 as many as 21 provinces. Cluster members in each cluster can be seen in Table 5.

Table 7. Results of Provincial Cluster Members

Cluster	Province	Total
1	Aceh, North Sumatra, Southeast Sulawesi, West Kalimantan, North Maluku, West Sulawesi, West Nusa Tenggara, South Kalimantan, South Sulawesi, Central Sulawesi, East Nusa Tenggara, Gorontalo, Maluku	13
2	West Sumatra, North Kalimantan, South Kalimantan, West Papua, Riau, Kep. Bangka Belitung, North Sulawesi, Papua, Jambi, South Sumatra, Kep. Riau, East Kalimantan, Bengkulu, West Java, DKI Jakarta, Bali, Lampung, Banten, East Java, Central Java, DI Yogyakarta	21

Growth in the prevalence of stunting will decrease if the special index for stunting treatment is high, infants aged less than 6 months who receive exclusive breastfeeding are high, the percentage of short and very short toddlers is high, complete basic immunization for infants is high, toddlers are obese (bb/tb), low nutrition low malnutrition, low malnutrition, and low malnutrition. The following is the interpretation obtained from this study as a whole:

1. Cluster 1 is a category of provincial groups that have a low special index for handling stunting, babies aged less than 6 months who get low breastfeeding, a high percentage of short and very short toddlers, low complete basic immunization for babies, obese toddlers (bb/tb) is high, malnutrition is high, malnutrition is high, and malnutrition is high. The provinces in this group are 13 provinces with members of Aceh, North Sumatra, Southeast Sulawesi, West Kalimantan, North Maluku, West Sulawesi, West Nusa Tenggara, South Kalimantan, South Sulawesi, Central Sulawesi, East Nusa Tenggara, Gorontalo, and Maluku.
2. Cluster 2 is a category of provincial groups that have a high special index for handling stunting, babies aged less than 6 months who get high breastfeeding, low percentages of short and very short toddlers, complete basic immunization for tall babies, obese toddlers (bb/tb) is low, malnutrition is low, malnutrition is low, and malnutrition is low. The provinces in this group are 21 provinces with members of West Sumatra, North Kalimantan, South Kalimantan, West Papua, Riau, Kep. Bangka Belitung, North Sulawesi, Papua, Jambi, South Sumatra, Kep. Riau, East Kalimantan, Bengkulu, West Java, DKI Jakarta, Bali, Lampung, Banten, East Java, Central Java, and DI Yogyakarta.

4. CONCLUSION

Based on the results of provincial clustering research in Indonesia using agglomerative hierarchical clustering produces 2 clusters, namely 13 provinces in Cluster 1 and 21 provinces in Cluster 2. Based on the comparison of the Average Linkage and Ward's algorithms the most optimal clustering result is to use Ward's algorithm with a Chopenetic Correlation value of 0.8399978. In determining the number of clusters, the Sillhouette Coefficient calculation is used by looking at the highest value on the Sillhouette Coefficient and obtaining 2 clusters of 0.317 even though based on the size of the Sillhouette Coefficient value there is a weak bond between each object and the cluster that has been formed. Namely, Cluster 1 has a fairly high prevalence of stunting, and Cluster 2 has a fairly low prevalence of stunting. So, it can be said that Ward's algorithm is better than the Average Linkage algorithm in clustering provinces in Indonesia on the prevalence of stunting toddlers. For further research, it is hoped that it will not only compare the Average Linkage and Ward's algorithms but add several other agglomerative algorithms such as Single Linkage and Complete Linkage.

REFERENCES

[1] M. Rizky Anggraeni and U. Yudatama, "JURNAL MEDIA INFORMATIKA BUDIDARMA Clustering Prevalensi Stunting Balita Menggunakan Agglomerative Hierarchical Clustering," 2023, doi: 10.30865/mib.v7i1.5501.

[2] Kinanti Rahmadhita, "Permasalahan Stunting dan Pencegahannya Stunting Problems and Prevention," *Jurnal Ilmiah Husada*, vol. 11, no. 1, pp. 225–229, 2020, doi: 10.35816/jiskh.v10i2.253.

[3] M. Rizky Anggraeni and U. Yudatama, "JURNAL MEDIA INFORMATIKA BUDIDARMA Clustering Prevalensi Stunting Balita Menggunakan Agglomerative Hierarchical Clustering," 2023, doi: 10.30865/mib.v7i1.5501.

[4] F. Mahanani Mulyaningrum and M. Mulya Susanti, "FAKTOR-FAKTOR YANG MEMPENGARUHI STUNTING PADA BALITA DI KABUPATEN GROBOGAN," *Jurnal Keperawatan dan Kesehatan Masyarakat*, vol. 10, no. 1, pp. 74–84, 2021.

[5] M. W. D. R. N. R. , A. U. N. W. M. Zainul Rahman, "ANALISIS KEBIJAKAN PENCEGAHAN STUNTING DAN RELEVANSI PENERAPAN DI MASYARAKAT (Studi Kasus: Desa Donowarih)," *Karta Raharja*, vol. 2, no. 1, pp. 27–33, 2021.

[6] S. Wulandari, "CLUSTERING KECAMATAN DI KOTA BANDUNG BERDASARKAN INDIKATOR JUMLAH PENDUDUK DENGAN MENGGUNAKAN ALGORITMA K-MEANS," in *Seminar Nasional Riset dan Teknologi (SEMNAS RISTEK) 2020*, 2020.

[7] Fithriyah Azzahrah, S. Annas, and Z. Rais, "Hybrid Hierarchical Clustering dalam Pengelompokan Daerah Rawan Bencana Tanah Longsor di Sulawesi Selatan," *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 4, no. 3, pp. 153–161, Dec. 2022, doi: 10.35580/variansiunm38.

[8] R. P. Justitia, N. Hidayat, and E. Santoso, "Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus : RichDjoe Barbershop Malang)," 2021. [Online]. Available: http://j-ptiik.ub.ac.id

[9] M. Paramadina, S. Sudarmin, and M. K. Aidid, "Perbandingan Analisis Cluster Metode Average Linkage dan Metode Ward (Kasus: IPM Provinsi Sulawesi Selatan)," *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 1, no. 2, p. 22, Jul. 2019, doi: 10.35580/variansiunm9357.

[10] D. Widyadhana, R. B. Hastuti, I. Kharisudin, and F. Fauzi, "Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 584–594, 2021, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[11] M. Rais Ridwan and H. Retnawati, "Numerical: Jurnal Matematika dan Pendidikan Matematika Application of Cluster Analysis Using Agglomerative Method," *Numerical: Jurnal Matematika dan Pendidikan Matematika*, vol. 5, no. 1, pp. 33–48, 2021, doi: 10.25217/numerical.v5i1.1358.

[12] R. A. Prabowo *et al.*, "Simulasi Pemilihan Metode Analisis Cluster Hirarki Agglomerative Terbaik Antara Average Linkage Dan Ward Pada Data Yang Mengandung Masalah Multikolinearitas," 2020.

[13] A. F. Dewi and K. Ahadiyah, "Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan," *Zeta - Math Journal*, vol. 7, no. 2, pp. 57–63, Nov. 2022, doi: 10.31102/zeta.2022.7.2.57-63.

[14] A. Sauddin, "Analisis Faktor-Faktor yang Mempengaruhi Kepatuhan Wajib Pajak Orang Pribadi dalam Memenuhi Kewajiban Membayar Pajak Kendaraan Bermotor (PKB)."

[15] S. Hanada *et al.*, "Penggunaan Analisis Cluster dalam Pengelompokan Kecamatan di Kabupaten Karawang Berdasarkan Metode Kontrasepsi Peserta KB Aktif," in *Prosiding Statistika*, 2021, pp. 42–29. doi: 10.29313/.v7i1.25510.

[16] R. Hidayati, A. Zubair, A. Hidayat Pratama, L. Indana, P. Studi Sistem Informasi, and F. Teknologi Informasi, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering," 2021.