Cendikia : Media Jurnal Ilmiah Pendidikan, 14 (5) (2024) pp. 503-510
Published by: IOCSCIENCE

Cendikia : Media Jurnal Ilmiah Pendidikan

Classification of villages in Tanimbar Islands based on stunting service packages using the K-Means Algorithm

Mesak Ratuanik[1], Samuel Urath[2], Pesparani Diana Jabar[3], Baceria Werluka[4], Inda A. Batbual[5], Thobias Melemabessy[6]

[1,2,4,5,6] Mathematics Education, Lelemuku Saumlaki University, Saumlaki, Indonesia
[3] English Language Education , Lelemuku Saumlaki University, Saumlaki, Indonesia

ARTICLE INFO

Article history:

Received Feb 26, 2024
Revised Mar 03, 2024
Accepted Apr 03, 2024

Keywords:

Cluster
K-Means Algorithm
Stunting
The Stunting Service Package

ABSTRACT

The Tanimbar Islands Regency still has a high prevalence of stunting in toddlers, so we must work together to eradicate it. According to WHO, the prevalence of stunting should not exceed 20%. According to data from the 2021 Indonesian Nutritional Status Survey (SSGI), the prevalence of stunting in toddlers is currently 25.1% in the Tanimbar Islands District, Maluku Province. The purpose of this study was to classify villages based on the indicators of the stunting service package in the Tanimbar Islands District. This research uses an analytic survey approach using secondary data obtained from the Central Bureau of Statistics (BPS) of the Republic of Indonesia in 2022 and the Tanimbar Islands District Health Office in 2022 by utilizing the K-Means Algorithm. The stages of data analysis in this study consisted of library research, data collection, data processing, the K-Means algorithm. Furthermore, the last stage is to verify the data consisting of analysis of findings based on the theory used. At this analysis stage, the K-Means Clustering Method was also applied to classify villages in the Tanimbar Islands District based on the stunting service package. Research results based on analysis using the K-Means algorithm (Number of causes in each cluster) provide an overview of the number of clusters that enter each cluster. Cluster 1 consists of 20 villages, cluster 2 consists of 66 villages, cluster 3 consists of 1 village and cluster 4 consists of 1 village.

Corresponding Author:

Mesak Ratuanik,
Mathematics Education,
Lelemuku Saumlaki University,
Jl. Prof. Dr Jl. Boediono, Lauran, Tanimbar Sel., Kab. Kepulauan Tanimbar, Maluku 97464, Indonesia.
Email: mratuanik83@gmail.com

INTRODUCTION

Stunting is a disorder where chronic malnutrition prevents children from growing normally, resulting in toddlers who are too short for their age. Stunting can disrupt brain growth, which can have long-term impacts including mental retardation, poor learning capacity, and increased risk of chronic diseases such as diabetes, hypertension, and obesity (Tello et al., 2022), (Yadika et al., 2019), (Fasha Alfarizi, 2022), dan (Fentiana et al., 2022). Six districts in Maluku Province have stunting rates in children under five that are higher than the national average. A total of 5 more cities/districts are

below the provincial average. The three districts that are in the spotlight in this effort are South Buru District which has the highest stunting rate among toddlers in Maluku in the 2022 SSGI, namely 41.6%. This figure represents an increase of 2.5 percentage points from 2021's 39.1%; Tanimbar Islands Regency is in second place with the highest prevalence of stunting, namely 31.5%; and Aru Islands Regency which is in third place with a percentage of 28.1% in Maluku Province. Stunting is a problem caused by several things, such as lack of infrastructure and sanitation services, scarcity of clean water sources, and lack of public awareness (Mustakim et al., 2022), (Thurstans et al., 2022) (Marcella Gloria Leto Bele et al., 2022), (Zubedi et al., 2021), (Salsabila, 2018), (Asroni et al., 2018). To combat stunting in the Tanimbar Islands Regency, attention must be paid to family planning, basic immunization, hygiene, and other factors.

The Regional Government of Tanimbar Islands Regency (KKT) is increasingly aggressive and committed to reducing the stunting rate by involving all parties. Based on facts and findings obtained from several previous studies regarding the factors that influence stunting in KKT. The Regional Government is more focused during the First 1,000 Days of Birth (HPK) on preventing and handling cases of stunting in pregnant women and toddlers under 2 years old through specific and sensitive nutritional interventions. Based on the findings of the situation study, the regional government has decided which villages in the KKT will be the locus for rapid reduction in stunting in 2022 and 2023. This has been proven or realized by the Regional Government by holding stunting consultation activities with the theme "Together Cutting Down Stunting" which was held on November 1, 2022. This is the reason for researchers to further classify villages in KKT based on stunting service packages using the K-Means algorithm.

RESEARCH METHODOLOGY

This research uses the K-Means algorithm to classify stunting in villages in KKT based on stunting service packages. Data obtained from primary data in BPS KKT 2022 and the KKT Health Service. The research location was carried out at KKT. The data collection methods used in this research are BPS KKT data for 2022 and KKT Health Service data for 2022. After the data is collected, the data is selected according to the stunting service package used, then data preprocessing is carried out. The aim is to ensure there is no duplication of data, no missing values and correct errors in the dataset. This stage carries out data cleaning, so that the data can be processed and the data mining process is carried out (Hoogeveen et al., 2022), (Priyatman et al., 2019), (Hutagalung, 2022), (Irfiani & Rani, 2018). Using the K-Means algorithm, the stages will be repeated until stability occurs (BASTIAN, 2018), (Kamila et al., 2019), (Prayoga et al., 2019), (Bahauddin et al., 2021), (Rohmah et al., 2021). The stages are as follows:

1. Select the desired number of Clusters (k) in the data set
2. Determine the center point (Centroid) randomly at the initial stage
3. Calculate the closest distance of each data to the Centroid. To calculate the closest distance to the Centroid, it is Euclidean distance
4. You can use the formula below:
   $de = \sqrt{(xi - si)^2 + (yi - ti)^2}$ (1)
   Information:
   (x, y) = Coordinate Object
   (s, t) = Centroid Coordinates
   i = Number of Objects
5. Recalculate the Cluster center with the current Cluster membership. Cluster Center is the average of data from a cluster (Goletti et al., 2022). Can be calculated using the formula:
   $$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

Information:
V_{ij} = Average centroid on Cluster k-i for the ke-j variable
N_i = Number of members of the i Cluster
i, k = Cluster Index
X_{kj} = The ke-k data value of the ke-j variable for the Cluster.

6. Recalculate each object using a new Cluster center (new Centroid), this is the initial stage of opening a new iteration. If the Cluster members do not experience any further Cluster movement, then the Clustering process is declared complete (Cardozo & Mens, 2022). However, if the Cluster members experience any further movement, then return to step c until the Cluster members do not experience any further movement (Martou et al., 2022).

RESULTS AND DISCUSSIONS

1. Designing/Designing Research

a. Data Standardization

Data standardization is carried out if there are significant unit differences between the variables studied. Because in Appendix 1 the units of the variables are different, in this research data standardization is carried out first using the data standardization formula:

$$Z = \frac{x_i - \bar{x}}{s} \quad (3)$$

Based on data collection at the Tanimbar Islands District Health Service, there are 2 variables with 88 villages where data from each object in the variable will be standardized. The first step taken in data standardization is to find the average and standard deviation. The following are the results of standardization of the variables used in the research.

Table 1. Standardization of Research Variables

Destinasi Ekowisata	X1	X2
Saumlaki	1.54417	0.74475
Saumlaki Utara	-0.23601	-0.43764
Matakus	0.95077	1.58931
Oltim	0.95077	1.25149
⋮	⋮	⋮
Adodo Molu	-0.23601	-0.43764

b. Outlier Detection

Based on standardized research data, if there is data whose value is not within ± 2.5, it means the data is an outlier. The results of observations on outlier data are shown in Table 1 previously. Based on the previous Table 1, it can be seen that there are no values that exceed ± 2.5. So it can be concluded that the data used in this research does not contain outliers.

2. Test Assumptions

There are assumptions that must be met in cluster analysis, namely:

a. Assumption of Sample Adequacy

To find out whether the sample used is sufficient for analysis, it can be seen from the Kaiser Meyer Olkin (KMO) value.

H0: The sample is not sufficient for further analysis
H1: The sample is sufficient for further analysis
Test Statistics

$$KMO = \frac{\sum_{i=1}^{\rho}\sum_{j=1}^{\rho} r_{ij}^2}{\sum_{i=1}^{\rho}\sum_{j=1}^{\rho} r_{ij}^2 + \sum_{i=1}^{\rho}\sum_{j=1}^{\rho} \rho_{ij}^2} \quad (4)$$

Test Criteria:

If the KMO value is > 0.5 then it fails to accept H0 or the sample is suitable for further analysis.

The KMO test results can be seen in Table 2 below:

Table 2 Test KMO and Barlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.550
Bartlett's Test of Sphericity	Approx. Chi-Square	18.984
	df	1
	Sig.	.712

Based on Table 2, the KMO and Barlett's tests are useful for determining the suitability of a variable, whether it can be processed further using this cluster analysis technique or not. The way to do this is by looking at the KMO MSA value. If the KMO MSA value is greater than 0.50 then the cluster analysis technique can be continued. Based on the table above, it can be seen that the KMO MSA value is 0.628 > 0.50 and the Barlett's Test of Sphericty (Sig.) value is 0.000 < 0.05, so the cluster analysis in this research can be continued because it meets the first requirement.

b. Multicollinearity Assumption

The second assumption is multicollinearity. To determine whether there is multicollinearity, you can look at the correlation values in the correlation matrix. It is said to be multicollinearity if the correlation value is greater than 0.80. Next, the results of the multilinearity test can be seen in Table 3 below.

Tabel 3. Test Multicollinearity

		0-23	24-59
0-23	Pearson Correlation	1	.446**
	Sig. (2-tailed)		.000
	N	88	88
24-59	Pearson Correlation	.446**	1
	Sig. (2-tailed)	.000	
	N	88	88

Based on Table 3, it shows that all variables have a Pearson correlation value of less than 0.80, so it can be concluded that there are no cases of multicollinearity in the research variables.

3. K-Means Cluster Analysis

a. Stunting Clusterization at Village Level in Tanimbar Islands Regency

K-means clustering is a non-hierarchical cluster analysis method that attempts to partition existing objects into one or more clusters or groups of objects based on their characteristics, so that objects that have the same characteristics are grouped in the same cluster and objects that have different characteristics. grouped into another cluster. The K-Means Clustering method attempts to group existing data into several groups, where the data in one group has the same characteristics as each other and has different characteristics from the data in other groups. The following is the first view of the data clustering process before iteration is carried out.

Tabel 4. Iteration History[a]

	Change in Cluster Centers			
Iteration	1	2	3	4
1	2.834	4.313	.000	.000
2	.202	.071	.000	.000
3	.203	.067	.000	.000
4	.451	.132	.000	.000
5	.000	.000	.000	.000

Based on Table 4, it shows that the iteration process in grouping clusters from the initial table resulted in an iteration process 4 times. In iterations 1, 2, 3, there were several centeroids that were not significant and in iteration 4 there were significant centeroids. So, all clusters have been formed and the iteration stops at iteration 4.

Table 5. Final Cluster Centers

	Cluster			
	1	2	3	4
0-23	2.30	1.02	6.00	4.00
24-59	10.95	1.88	25.00	36.00

Based on Table 5, it shows the results of the final process in clustering which forms 4 clusters for each variable. The variables in the final cluster centers table are the results for standardized values with positive numbers meaning that the data is above the total average.

To see the significance level test between clusters and find out the differences in each cluster, it is necessary to carry out an ANOVA test. The conditions for using the F number in cluster analysis are that the greater the calculated F number (if a hypothesis test is carried out, the calculated F will be greater than the F table) and the significance level (sig) < 0.05; the greater the differences between the three clusters formed. The following are the results of the Anova test as follows.

Table 6. ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
0-23	17.965	3	2.300	84	7.811	.000
24-59	899.097	3	4.190	84	214.569	.000

Based on Table 6, the decision rule is as follows.
H0 = There are no significant differences between the four clusters
H1 = The four clusters have significant differences
Decision Rule: Reject the Null Hypothesis (H0) if the p-value <0.05

Based on Table 7, it shows the p-value for the two variables used in the research, namely variables 0-23 and 24-59 which have a p-value (Sig.) <0.05. So it can be concluded that there is a significant difference between the two variables because the p-value obtained is <0.05.

Table 7. Number of Cases in each Cluster

Cluster	1	20.000
	2	66.000
	3	1.000
	4	1.000
Valid		88.000
Missing		.000

Based on Table 7, Number of caes in each cluster, provides an overview of the number of clusters included in each cluster. Cluster 1 consists of 20 villages, cluster 2 consists of 66 villages, cluster 3 consists of 1 village and cluster 4 consists of 1 village.

CONCLUSION

The results of this research show that, Sample Adequacy Assumption/Kaiser Meyer Olkin (KMO), it can be seen that the KMO MSA value is 0.628 > 0.50 and the Barlett's Test of Sphericty (Sig.) value is 0.000 < 0.05, so the cluster analysis in this study can be continued because it has met first requirement. The multicollinearity assumption for all variables has a Pearson correlation value of less than 0.80 so it can be concluded that there are no cases of multicollinearity in the research variables. K-Means Cluster Analysis shows that the iteration process in grouping clusters from the initial table results in an iteration process 4 times. In iterations 1, 2, 3, there were several centeroids that were not significant and in iteration 4 there were significant centeroids. So, all clusters have been formed and the iteration stops at iteration 4. Regency/City Clusters Formed, shows that there are 4 clusters formed in stunting clustering at the village level in Tanimbar Islands Regency with the following details: Cluster 1 consists of 20 Villages, Clusters 2 consists of 66 villages, Cluster 3 consists of 1 village, Cluster 4 consists of 1 village. Based on the Anova test, it shows that the p-value for the two variables used in the research, namely variables 0-23 and 24-59, has a p-value (Sig.) <0.05. So it can be concluded that there is a significant difference between the two variables because the p-value obtained is <0.05. Suggestions in this research are to consider adding qualitative elements to gain deeper insights. The hope is to involve other experts with different expertise to get diverse perspectives and a deeper understanding of research on stunting.

ACKNOWLEDGEMENTS

1. Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi sebagai pemberi dana Pengabdian kepada Masyarakat Pemula Tahun Anggaran 2023.
2. Lembaga Layanan Pendidikan Tinggi (LLDIKTI) Wilayah XII Maluku dan Maluku Utara
3. Tim Reviewer Monev Internal
4. Universitas Lelemuku Saumlaki
5. LPPM Universitas Lelemuku Saumlaki
6. Dinas Kesehatan Kabupaten Kepulauan Tanimbar
7. Program Studi Pendidikan Matematika, Universitas Lelemuku Saumlaki

To all parties who have helped in this research process and hopefully it will become input in eradicating stunting in Indonesia and especially in the Tanimbar Islands Regency.

References

Asroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*, *21*(1), 60–64. https://doi.org/10.18196/st.211211

Bahauddin, A., Fatmawati, A., & Permata Sari, F. (2021). Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means. *Jurnal Manajemen Informatika Dan Sistem Informasi*, *4*(1), 1–8. https://doi.org/10.36595/misi.v4i1.216

BASTIAN, A. (2018). Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). *Jurnal Sistem Informasi*, *14*(1), 28–34. https://doi.org/10.21609/jsi.v14i1.566

Cardozo, N., & Mens, K. (2022). Programming language implementations for context-oriented self-adaptive

systems. *Information and Software Technology*, *143*(October). https://doi.org/10.1016/j.infsof.2021.106789

Fasha Alfarizi, T. (2022). Literature Review : Hubungan Kebijakan dan Pelayanan Kesehatan dengan Kebijakan dan Pelayanan Kesehatan Kejadian Stunting. *Borneo Student Research*, *3*(3), 2949–2955.

Fentiana, N., Achadi, E. L., Besral, Kamiza, A., & Sudiarti, T. (2022). A Stunting Prevention Risk Factors Pathway Model for Indonesian Districts/Cities with a Stunting Prevalence of ≥30%. *Kesmas*, *17*(3), 175–183. https://doi.org/10.21109/kesmas.v17i3.5954

Goletti, O., Mens, K., & Hermans, F. (2022). An Analysis of Tutors' Adoption of Explicit Instructional Strategies in an Introductory Programming Course. In *ACM International Conference Proceeding Series* (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3564721.3565951

Hoogeveen, S., Sarafoglou, A., Aczel, B., Aditya, Y., Alayan, A. J., Allen, P. J., Altay, S., Alzahawi, S., Amir, Y., Anthony, F. V., Kwame Appiah, O., Atkinson, Q. D., Baimel, A., Balkaya-Ince, M., Balsamo, M., Banker, S., Bartoš, F., Becerra, M., Beffara, B., … Wagenmakers, E. J. (2022). A many-analysts approach to the relation between religiosity and well-being. *Religion, Brain and Behavior*. https://doi.org/10.1080/2153599X.2022.2070255

Hutagalung, J. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(1), 606–620. https://doi.org/10.35957/jatisi.v9i1.1516

Irfiani, E., & Rani, S. S. (2018). Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, *6*(4), 161. https://doi.org/10.26418/justin.v6i4.29024

Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, *5*(1), 119. https://doi.org/10.24014/rmsi.v5i1.7381

Marcella Gloria Leto Bele, Elvira Mustikawati Putri Hermanto, & Fenny Fitriani. (2022). Pemodelan Geographically Weighted Regression pada Kasus Stunting di Provinsi Nusa Tenggara Timur Tahun 2020. *Jurnal Statistika Dan Aplikasinya*, *6*(2), 179–191. https://doi.org/10.21009/jsa.06204

Martou, P., Mens, K., Duhoux, B., & Legay, A. (2022). Generating Virtual Scenarios for Cyber Ranges from Feature-Based Context-Oriented Models: A Case Study. In *ACM International Conference Proceeding Series* (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3570353.3570358

Mustakim, M. R. D., Irwanto, Irawan, R., Irmawati, M., & Setyoboedi, B. (2022). Impact of Stunting on Development of Children between 1-3 Years of Age. *Ethiopian Journal of Health Sciences*, *32*(3), 569–578. https://doi.org/10.4314/ejhs.v32i3.13

Prayoga, Y., Tambunan, H. S., & Parlina, I. (2019). Penerapan Clustering Pada Laju Inflasi Kota Di Indonesia Dengan Algoritma K-Means. *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, *1*(1), 24–30. https://doi.org/10.30645/brahmana.v1i1.4

Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *5*(1), 62. https://doi.org/10.26418/jp.v5i1.29611

Rohmah, A., Sembiring, F., & Erfina, A. (2021). Implementasi Algoritma K-Means Clustering Analysis untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung). *Seminar Nasional Sistem Informasi Dan Manajemen Informatika*, 290–298. https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/32

Salsabila, N. (2018). Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang. *Central Library Of Maulana Malik Ibrahim State Islamic University Of Malang*, 89. http://etheses.uin-malang.ac.id/16985/1/14650031.pdf

Tello, B., Rivadeneira, M. F., Moncayo, A. L., Buitrón, J., Astudillo, F., Estrella, A., & Torres, A. L. (2022). Breastfeeding, feeding practices and stunting in indigenous Ecuadorians under 2 years of age. *International Breastfeeding Journal*, *17*(1), 1–15. https://doi.org/10.1186/s13006-022-00461-0

Thurstans, S., Sessions, N., Dolan, C., Sadler, K., Cichon, B., Isanaka, S., Roberfroid, D., Stobaugh, H., Webb, P., & Khara, T. (2022). The relationship between wasting and stunting in young children: A systematic

review. *Maternal and Child Nutrition*, *18*(1). https://doi.org/10.1111/mcn.13246

Yadika, A. D. N., Berawi, K. N., & Nasution, S. H. (2019). Pengaruh stunting terhadap perkembangan kognitif dan prestasi belajar. *Jurnal Majority*, *8*(2), 273–282.

Zubedi, F., Oroh, F. A., & Aliu, M. A. (2021). Pemodelan Stunting Dan Gizi Kurang Di Kabupaten Bone Bolango Menggunakan Regresi Pisson Generalized Modeling. *Jurnal Matematika Dan Pendidikan Matematika*, *6*(2), 113–128.