International Journal of
Advances in Scientific Research and Engineering (ijasre)

E-ISSN : 2454-8006

DOI: 10.31695/IJASRE.2019.33258

Volume 5, Issue 6
June - 2019

Implementation of the K-means Clustering Method on Stunting Case in Indonesia

Nailul Izza A.Md. S.KM[1], Dr. Windhu Purnomo M.S.[2] and Dr. Mahmudah, Ir. M. Kes. [2]

[1]Student Master of Public Health

[2] Department Biostatistics of Public Health Study Program

College of Public Health of Airlangga University

Surabaya, Jawa Timur

Indonesia

ABSTRACT

Method K - means clustering is one technique that will partition the data into groups, so that the data which have the same characteristics are grouped into the same group. Clustering can be done on various types of data including clustering in order to determine the province which is the priority for stunting reduction. This study aims to classify provinces based on the prevalence of stunting for infants 0-59 months, exclusive breastfeeding, weigh 4 times, adequacy of energy and protein using the K-Means clustering method. The results showed that the optimal cluster formed was 4 clusters, in which group 1 was a cluster whose province had good nutritional status with an alow prevalence of stunting with a percentage of exclusive breastfeeding, weighing 4 times, adequate energy and high protein. While cluster 4 is a cluster that needs priority attention since the average value of stunting high-value percentage of energy adequacy low protein,although it has a value of percentage of ASI exclusive and toddlers weighing 4 times high enough that consists of Nusa Tenggara Barat, Nusa Tenggara East, Aceh, South Sulawesi, West Sulawesi and Banten. Provinces with stunting values, exclusive breastfeeding, weighing toddlers more than 4 times, adequacy of energy and protein that are the same height tend to group together, while provinces with low scores also group themselves.

Keywords: *K-Means Clustering, Province, Stunting.*

1. BACKGROUND

Data mining refers to processes or methods that extract knowledge to find interesting patterns of large amounts of data [1]. Data mining techniques clusters use unsupervised learning method which means data mining does not need to do training first but can directly use it for grouping [2]. The process of grouping large amounts of data into several classes according to their respective characteristics is called clustering [3].

The K - means method is the most popular and is a simple solution for clustering . K-Means will partition the data into groups, so that data that has the same characteristics are grouped into the same group and data that has different characteristics will be grouped into other groups [1]. The results of the cluster formed from the K-means method are very dependent on the initiation of the initial center value of the cluster given and can be a weakness [2]. In addition, the k-means partition only finds clusters that are formed without regard to cluster size and density [4].

Stunting is a condition of failure to thrive in children under five (infants under five years) due to chronic malnutrition so children are too short for their age. Malnutrition occurs since the baby is in the womb and in the early days after the baby is born but thestunting condition only appears after the baby is 2 years old [5]. Toddlers are short (stunted) and very short (severely stunted) are toddlers with body length (PB/U) or height (TB/U) according to their age compared to WHO-MGRS standard [6].

The stunting prevalence in Indonesia is in the fifth largest in this world [7], besides that to prevalensi stunting is still above the national target in 2013 of 37,2 percent and for 2018 by 30,8 percent [8]. Data from these studies indicate that Indonesia is a country that still has public health problems, according to a statement from WHO to set limits on nutritional problems not more than 20 percent [9], [10].

Factors that are the cause of stunting due to poor parenting practices, where that 60% of children aged 0-6 months do not get breast milk exclusively (ASI) [7]. According to the Director General of Health and Community Empowerment [11] the cause of stunting one of them is non-exclusive breastfeeding. Anisa [12] research results that a low protein nutrient supply has a chance of stunting 5,775 times compared to toddlers in Kalibaru Sub-district whose protein intake is sufficient.

Determining the provinces priority in reducing stunting is certainly not only considering the high prevalence of stunting in the area, but also considering the causes of stunting . Clustering can be done on various types of data including clustering in order to determine the priority provinces . The aim of the study was to classify provinces based on the prevalence of stunting of children under five to five months, exclusive breastfeeding, weighing four times, adequacy of energy and protein using the K-Means clustering method.

2. METHOD

This research is a non-reactive study because it only carries out secondary data collection obtained from research data "Monitoring of Nutritional Status" by the Directorate of Community Nutrition of the Ministry of Health of the Republic of Indonesia [6]. The data used in this study are prevailing stunting, exclusive breastfeeding, toddler weighing more 4 times, energy adequacy and protein sufficiency in a province in Indonesia. The unit of analysis of this study is the provinces in Indonesia, which number 34 provinces.

In this study, provinces in Indonesia will be grouped based on stunting prevalence data, the percentage of exclusive breastfeeding, the percentage of toddlers weighing more 4 times, the percentage of toddlers is enough energy and the percentage of toddlers is enough protein . The grouping method uses K-Means with the studio R program 1.1.423. The steps in the K-Means method are: [13]

1. Divide the items into k -group
2. Calculate the value of the centroid
3. Grouping items based on the nearest centroid, the distance used is euclidean distance .
4. Recalculate the group centroid when receiving new items or items that come out.
5. Iterates until there are no more items that can enter or exit again when the convergence criteria have been fulfilled.

The cluster formula used is as follows:

$$\chi = \mu + Z.\sigma$$

Description: X = average sample in the cluster
μ = population average
Z = standardization value
Σ = standard deviation

The clustering process produces several clusters, each of which has an average value used to determine the customer class at the data modeling stage. The cluster that has been formed is tested for its validity level to find out the best number of clusters using the silhouette index. The silhouette index value is between the value -1 to 1, if it is close to 1 it means that the number of clusters produced is optimal. Then in each cluster potential groups can be determined through the class produced [2].

3. RESULTS AND DISCUSSION

The clustering process produces several clusters, each of which has an average value used to determine the class at the data modeling stage. Determining the best number of clusters in the Kmeans method can be done through cluster validation by looking at the Silhouette index value .

DOI: 10.31695/IJASRE.2019.33258

Table 1. Silhouette Index Values For Optimal Cluster Selection

Number of cluster	1	2	3	4	5	6
Average Silhouette width	0	0,2281	0,1903	0,2306	0,2226	0,206

The results of cluster validation calculations with the R program indicate that the cluster is optimal for provincial grouping data based on the prevalence of stunting and the cause is 4 clusters. Based on the silhouette index value, it will be continued for the cluster distribution.

Table 2. Summary of 2017 Stunting Kmeans Clustering Data in 34 Provinces, Indonesia

Variable	Minimum	Maximum	Mean	Std. D eviation
Stunting	19,1	40,3	30,297	5,5566
Exclusive breastfeeding	10,7	61,4	33,235	10,4101
Weigh more 4 times	54,9	88,0	72,976	8,8622
Enough energy	17,10	43,50	29,0441	6,89524
Enough protein	36,10	66,20	55,0853	6,69614

The iteration process is done to get the right cluster . I n this study occurred 4 times iteration with a minimum distance between the center of the cluster that occurred at 4,191. The value of cluster center s in each cluster formed can be seen in table 3 . This can be any provincial reference which has the best achievement for the indicator category when the grouping results. The cluster center value which has the largest positive value indicates that the cluster is the best cluster.

Table 3. Clusters with the k-means method for all provinces

Indicator of causes of stunting	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Stunting	-1,18276	,21049	,10197	1,07433
Exclusive breastfeeding	,79391	-,04181	-,82165	,51053
Weigh more 4 times	,86728	-,47879	-,67438	,79817
Enough energy	,99067	-,90718	,26784	-,45115
Enough protein	,56222	-1,03655	,71088	-,49809

Cluster 1

Cluster 1 has a positive value in cluster center values for all indicators of the causes of stunting. The results of the calculation of the average value of the occurrence of stunting along with the indicators of the causes of stunting in the province in cluster 1 that is:

- The average value of the stunting percentage = (30,297) + (-1,1827 * 5.5566) = 23,7256
- The average percentage of exclusive breastfeeding = (33,235) + (0.7939 * 10,410) = 41,4996
- Average value of Weighing percentage 4 times = (72,976) + (0,8672 * 8,8622) = 80,6620
- Average percentage value Enough energy = (17,100) + (0,9906 * 6,8952) = 23,9308
- Average percentage value of enough protein = (36,100) + (0.5622 * 6.6961) = 39,8647

Based on the results of calculations in cluster 1, it shows that provinces that have an average stunting percentage value are low(23.72%), so the average percentage value for stunting is higher.

Cluster 2

P there are cluster 2 grouped regions with an average value of stunting percentage that is getting higher (31, 46 %) and has a low average value of the causes of stunting, namely the average value of exclusive ASI percentage is low (32.79 %), weighs more than 4 times lower (6 8.73 %), low energy adequacy (10.84 %) and low protein sufficiency (29.15 %).

DOI: 10.31695/IJASRE.2019.33258

Cluster 3

This is different from what happens in cluster 3 . In cluster 3 it contains provinces where the average stunting value is with and has the percentage value of exclusive breastfeeding and weighs more than 4 times lower, but has a high protein sufficiency value .

Cluster 4

In this cluster contains four provinces with an average value of stunting high pitch that has a percentage value of low-protein energy sufficiency, but has a value of more stunting causes a high percentage.

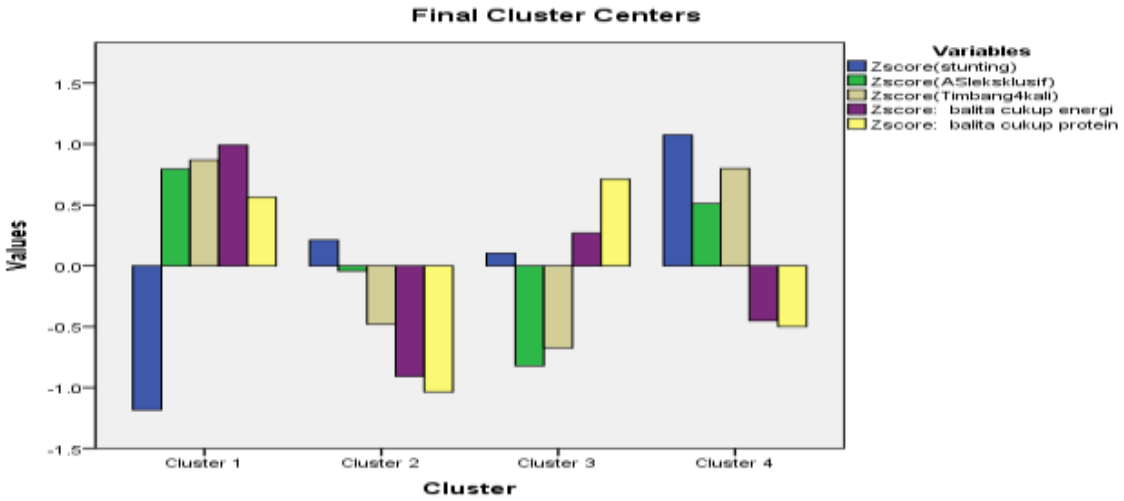

Fig 1. Graph of Final Cluster Center Average Value in 4 clusters

The variable that most shows the difference between the causes of stunting is the adequacy of energy indicated by the value of F = 13,993 (sig = 0,00), then the adequacy of the protein shown from the value F = 10,977 (sig = 0,00), and toddlers weigh 4 timesindicated by the value F = 10,504 (sig = 0,00) .

Table 4. Division of Cluster Province and distance to centroid

Class	Number of members (N = 34)	Name of Province	Distance to centroid
Cluster 1	8	Riau islands	1,79681
		Bali	1,50681
		South Sumatra	1,39923
		Central Java	1,24944
		DKI Jakarta	1,46378
		In Yogyakarta	2,19933
		West Java	1,36393
		East Java	1,58165
Cluster 2	9	Central Kalimantan	1,80553
		Maluku	0,79434
		West Papua	1,39773
		Lampung	1,21534
		West Sumatra	0,64514
		Jambi	1,57235
		West Kalimantan	1,24089
		South Kalimantan	1,06909
		Papua	2,43613
Cluster 3	1 1	North Sumatra	2,05093
		Riau	1,08349
		Bengkulu	0,49282
		Kep,Bangka Belitung	1,82274
		North Sulawesi	1,17915
		Southeast Sulawesi	1,49547
		Gorontalo	1,18844
		Central Sulawesi	1,36938

DOI: 10.31695/IJASRE.2019.33258

		North Maluku East Kalimantan North Kalimantan	1,75809 1,21127 1,49649
Cluster 4	6	West Nusa Tenggara East Nusa Tenggara Aceh South Sulawesi West Sulawesi Banten	1,87377 1,83231 1,97622 1,09360 1,72160 1,45727

Clustering using the K-means method obtained cluster 1 as the best provincial cluster with an average low stunting value with exclusive breastfeeding rates, weighing toddlers 4 times, energy sufficiency and high protein . While cluster 4 is a cluster that need priority attention since the average value of stunting high value percentage of energy adequacy low protein, although it has a value of percentage of ASI e ksklusif and toddlers weighing 4 times high enough that consists of Nusa Tenggara Barat, Nusa Tenggara East , Aceh , South Sulawesi , West Sulawesi and Banten. In addition, provincial groups that need to be prioritized are in cluster 2 which although the stunting average is not too high but the percentage value of the causes of stunting is low.

4. CONCLUSION

Provinces with stunting values, exclusive breastfeeding, weighing toddlers more than 4 times, adequacy of energy and protein that are the same height tend to group together, while provinces with low scores also group themselves.

REFERENCES

1. J, Han and and M, Kamber, Data Mining : Concepts and Techniques, San Fransisco California: Morgan Kauffman Publisher, 2006, p. 5-7, 79-95 .
2. C, C, Aggarwal, Data Mining, 2015, p . 171-182.
3. C, J, Matheus and P, K, Chan, “Systems for Knowledge Discovery in Databases,” IEEE Trans, Knowl, Data Eng,, 1993, p. 903-913.
4. A, Musdholifah, S, Hashim, and S, Zaiton, “Cluster Analysis on High-Dimensional Data: A Comparison of Density-based Clustering Algorithms,” Aust, J, Basic …, vol, 7, no, 2, 2013, p. 380–389.
5. National planning and development agency, "Human Resources Stunting And Development Improved Economic Situation and Maintained Stability," 2018.
6. Directorate Community of Nutrition, Nutrition Status Monitoring Handbook (PSG) in 2017.
7. National team accelerates poverty reduction, 100 Priority Districts / Cities for Stunting Children: Summary . Jakarta: Deputy President of the Republic of Indonesia Secretariat, 2017,
8. Health research and development agency, “The main results of 2018 Riskesdas,” 2018.
9. Aryastami N, K, dan I, T, , Policy Study and Management of Stunting Nutrition Problems in Indonesia," Bulettin Researcher System.Health. , vol. 45, no. December 4, 2017, pp. 233–240, 2017.
10. M, de Onis and F, Branca, “Childhood stunting: A global perspective,” Matern, Child Nutr,, vol, 12, pp, 12–26, 2016,
11. D, P, dan P, M, Promkes, Behavior change communication strategy, November, 2018,
12. P, Anisa, “Final Project: Factors - Factors Associated with the occurrence of Stunting in Toddlers Aged 25 - 60 Months in Kalibaru Village Depok in 2012 Universitas Indonesia Factors Associated with Stunting Events in Toddlers Aged 25 - 60 Months in Kal Village , "Universitas Indonesia, 2012.
13. Johnson D, W, and W, Richard A, , Applied Multivariate Statistical Analysis, Sixth edit, NerJersey: Pearson, 2007.

DOI: 10.31695/IJASRE.2019.33258