Jurnal Biometrika dan Kependudukan

© 2019 Jurnal Biometrika dan Kependudukan
p-ISSN 2302–707X e-ISSN 2540–8828
9(1): 62–68, July 2020
DOI: 10.20473/jbk.v9i1.2020.62–68

IMPLEMENTATION OF CLUSTERING USING K-MEANS METHOD TO DETERMINE NUTRITIONAL STATUS

Stefanny Surya Nagari[1], Lilik Inayati[2]
[1] Department of Biostatistics and Population, Faculty of Public Health, Universitas Airlangga, 60115 Surabaya, East Java, Indonesia
[2]Mayangrejo Village Midwife, Kalitidu Health Center, 62152 Bojonegoro, East Java, Indonesia
Corresponding Author: Stefanny Surya Nagari
E-mail: stefannysnagari@gmail.com

ABSTRACT

Cluster analysis aims to classify data objects into two categories: objects that are similar in characteristics in one cluster and objects that are different in characteristics with the other objects of another cluster. K-Means is a method included in the distance-based clustering algorithm that starts by determining the number of desired clusters. Malnutrition is one of the biggest concerns in Indonesia. According to Riskesdas 2018 data, as many as 17.7% infants under 60-month-old are still having problems with nutrition intake while 3.9% are having malnutrition. This might result in higher death rate. This research was conducted to classify the nutritional status of infants under 60-month-old conducted by the C-Means Clustering method. This research is non-reactive, using secondary data in Ponkesdes Mayangrejo, Bojonegoro without direct interaction with the subject. This study concluded that the grouping of nutritional status is possible by using K-Means with 4 clusters formed which are 23 malnourished toddlers, 17 undernourished toddlers, 7 nourished toddlers, and 10 over-nourished toddlers.

Keywords: cluster analysis, k-means, nutritional status

ABSTRAK

Analisis kluster bertujuan untuk mengelompokan objek data yang mempunyai kemiripan karakteristik satu sama lain dalam kluster yang sama dan berbeda karakteristiknya terhadap objek yang berbeda kluster. K-Means adalah metode yang termasuk dalam algoritma klustering berbasis jarak yang dimulai dengan menentukan jumlah kluster yang diinginkan. Kekurangan gizi adalah salah satu masalah kesehatan yang masih menjadi perhatian di Indonesia, menurut data Riskesdas 2018 jumlah bayi usia dibawah 60 bulan sebanyak 17,7% masih mengalami masalah gizi dengan bayi yang mengalami gizi buruk sebanyak 3,9% sehingga akan menyebabkan tingkat kematian yang lebih tinggi. Penelitian ini dilakukan untuk mengelompokan status gizi anak usia dibawah 60 bulan yang dilakukan dengan metode C-Means Clustering. Penelitian ini adalah non-reaktif, menggunakan data sekunder di Ponkesdes Mayangrejo Kabupaten Bojonegoro, tanpa interaksi langsung dengan subjek. Penelitian ini menyimpulkan bahwa dapat dilakukan proses pengelompokan status gizi menggunakan K-Means dengan hasil 4 kluster yang terbentuk, terdiri dari 23 balita gizi buruk, 17 balita gizi kurang, 7 balita gizi baik dan 10 balita gizi lebih.

Kata kunci: *analisis kluster, k-means, status gizi*

Received: September 20, 2019 ***Accepted: December 20, 2020***

INTRODUCTION

The world's rapid development is followed by advances in information technology. This phenomenon occurs in all aspects of life so that it can help humans in carrying out activities. One example in the health sector is the process of storing data, processing data, and grouping data that will lead to the development of statistical theory by leading to the ability to calculate and reliability of computers.

The application of information technology in health sector can help health workers in grouping data so that it can facilitate the data processing and can be immediately allocated to the authorities so that it can be followed up.

Cluster analysis is an analysis that has the purpose of grouping data objects on the condition that they have similar characteristics and then becomes one cluster, but if objects with different characteristics will be grouped in

different clusters (Rohmawati, Defiyanti and Jajuli, 2015)

Cluster analysis is a multivariate technique where the analysis process uses a clustering algorithm rather than being done by human, with the main objective of this analysis is to sort data based on its characteristics so that the data that have the closest similarities to other data are grouped into one cluster. In other words, this analysis is useful in directing group findings that were previously unknown to the data (Han, Micheline, and Jian, 2011)

Nutritional status grouping can be done with the help of cluster analysis using data mining. The process of the analysis includes finding and grouping data on the condition that they have the same characteristics (Irfiani and Rani, 2018). Clustering analysis has many methods that can be done including methods of Fuzzy C-Means Clustering, Self-Organizing Maps, K-Means, etc.

K-Means is a method of clustering analysis that uses data mining in the process of grouping data (Bastian, Sujadi and Febrianto, 2018). K-Means method will only work on numeric attributes because this method is a distance-based algorithm from the way it works dividing data into several clusters (Ediyanto, Mara dan Satyahadewi, 2018). The purpose of the K-Means method is to minimize the variation of data in the same cluster while in different clusters the variation of data will be maximized (Witten, Frank and Hall, 2011). Therefore, it can be concluded that K-Means is a way to sort the data into several groups so that the group is homogeneous among its members or within the group that has formed the smallest variation of data.

K-Means is a method derived from a simple idea to minimize the double errors contained in the grouping problem (Singla and Karambir 2012). This method is a learning algorithm that is classified as simple since the initial purpose of this method is to determine how many clusters you want to form, then the first variable in the cluster can be selected to be the midpoint of the cluster and the process of repeating steps will be carried out until data stability occurs so that the data object will not be moved

Data grouping can be done if the data is non-hierarchical by using the K-Means method, but with the condition that they have the same characteristics (Efendi, Coastera dan Tanjung, 2019). The purpose of this method is that variations of data from each cluster will be maximized and objective functions that have been set during the clustering process will be minimized (Agusta, 2007).

One of the efforts to fulfil basic needs should be a concern because these efforts are also an investment in national development to improve the quality of human resources (Indonesian Ministry of Health, 2003). The main component of the HDI (Human Development Index) measurement process is health besides education and income (Sundari and Ariani, 2019). But in achieving these efforts, there are several obstacles including nutritional problems in infants (Prasetyo et al., 2014).

Indonesia's condition in 2018 experiences multiple nutritional problems, both excessive and malnutrition. Nutrition problems in infants under 60-month-old based on the results of the Basic Health Research in 2018 showed that as many as 17.7% of infants under 60-month-old experienced nutritional problems (Indonesian Ministry of Health, 2010). This figure consists of malnutrition problems in infants under 60-month-old by 3.9% and under-nutrition problems by 13.8%. Nutritional problems in infants under 60-month-old have several negative effects and interrelated factors that can affect nutritional problems, such as inhibiting the physical and mental growth. Therefore, the learning process of infants will be interfered for example in the process of learning to talk, walk, and so forth.

The direct cause of malnutrition is when toddlers do not get good nutrition and insufficient nutritious food provided by parents to toddlers and infectious diseases. Another cause of the emergence of malnutrition status is poverty in the availability and consumption of food in the family (Indonesian Ministry of Health, 2004). Lack of food availability and consumption in infants can also cause stunting and the most severe impact is that it can cause acceleration of death, disability and higher morbidity rates (Rahim, 2014).

The principle of balanced nutrition is to get used to eating a variety of foods according to the needs of each family member in a balanced way. Getting used to a healthy lifestyle regularly by maintaining a healthy diet and monitoring weight regularly will affect toddlers' nutritional status. Balanced nutrition is very important for toddlers' growth and development

since good nutrition will protect the body from diseases, infections, and chronic diseases.

The role of the Public Health Center (PHC) and Health Integrated Service Post (HISP) are needed to monitor the nutritional value of toddlers as the efforts to anticipate the increasing number of cases of malnutrition in children under 5-year-old. Moreover, the role of parents is also important in giving toddlers a balanced nutritional intake by choosing the type of food that fulfil the nutritional value for toddlers. Information and knowledge increase about proper parenting and eating patterns is needed.

One way to determine the nutritional status of infants is to take anthropometric measurements. Anthropometric measurements are often used in determining the nutritional status of toddlers because it is simple and safe and does not need special action during the process. The results of anthropometric measurements are accurate so that they can detect nutritional history in infants.

Anthropometric measurements can be done by parents to monitor their toddlers' nutrition, but if parents cannot take measurements because they do not have supporting tools, parents can regularly visit the HISP called "Posyandu" every month. Posyandu does not only provide maternal and child health services, but also their height and weight will be measured to know the child's growth and development, especially regarding the development of the nutritional value.

In order to determine the nutritional status of children under five-year-old in Mayangrejo Ponkesdes, Kepmenkes 2010 classification standard is used as a benchmark for determining nutritional status to monitor children's nutritional status by looking at clusters as homogeneous as possible using K-means method.

METHODS

This research is a non-reactive study, which uses secondary data (Kuntoro, 2011). This study does not require a response from the respondent or the respondent does not realize that they are being studied (secondary data). The data was obtained from the results of the measurement and monitoring of nutritional status at the Mayangrejo Ponkesdes, Bojonegoro from January to June 2019. Nutritional status is measured according to Indonesian Ministry of Health Regulation Number 155/Menkes/Per/I/2010 Concerning Penggunaan Kartu Menuju Sehat (KMS) Bagi Balita classification standard, namely body weight by age.

The sampling process was carried out randomly and as many as 57 toddlers were obtained as samples. This study aims to classify the nutritional status of children under 5-year-old by grouping respondents into relatively homogeneous clusters using the C-Means Clustering method.

RESULT

Table 1. Descriptive Analysis of Body Weight and Age

	Weight	Age
N	57	57
N Missing	0	
Average	11.22	32.53
Mode	8.8	48
Median	11.3	36
Standard Deviation	3.17	17.07

Based on Table 1, a descriptive analysis of nutritional status according to the age of respondents shows the average age of toddlers 32.53 months, with a standard deviation of 17.07 months. While the body weight of the respondents indicated the average underweight of 11.22 kg with a standard deviation of 3.17 kg. While the distribution of age and weight according to the nutritional status based on the anthropometric book of the Minister of Health Decree 2010 has a fairly diverse nutritional status.

Table 2. Descriptive Analysis Based on the Anthropometry Pocket Book of the Ministry of Health 2010

Nutritional status	Weight		Age	
	Average	Standard Deviation	Average	Standard Deviation
Good nutritional status	11.84	3.33	31.5	17.25
Poor nutritional status	10.42	2.10	36	16.93
Poor nutritional status	7.6	0.80	31.5	18.81

Cluster Analysis on Toddlers' Nutrition Status Using the K-Means Method

The non-hierarchical clustering method is K-Means which main purpose is to group data into clusters. The data mining algorithm, which works by beginning with determining the number of initial groups as desired, then the object to be studied included in the group is a K-means cluster method. The results of the center end of the cluster can be seen in table 3.

Table 3 shows the updated centroid results until there were no changes in the cluster. This centroid shows the average number of children under 5-year-old joined in a cluster.

Table 3. End Center Cluster Results

	1	**2**	**3**	**4**
Age	44,13	21.18	56.71	8,20
Weight	13.69	8.88	12.91	8.34

Table 4 shows the results of the number of members of each cluster from the grouping process of nutritional status in the Mayangrejo Ponkesdes in 4 clusters, where cluster 1 with 23 children, cluster 2 with 17 children, cluster 3 with 7 children and cluster 4 with 10 children.

Table 4. Results Number of Each Cluster's Member

Cluster	1	23
	2	17
	3	7
	4	10
Valid		57
Missing		0

Based on table 5, it can be seen that the grouping members are based on nutritional status in each cluster with a total of 4 (four) clusters namely clusters 1, 2nd toddlers, 3, 11, 12, 13, 14, 19, 22, 25, 27, 28, 29, 30, 35, 36, 37, 40, 41, 46, 47, 49, 54 and 55. Cluster 2, toddlers 5, 8, 9, 10, 15, 16, 17, 20 , 23, 31, 33, 42, 43, 45, 48 and 53. Cluster 3, 6th, 26th, 32nd, 39th, 51st, 52nd, and 56th Cluster 4, 1st, 4th, 7th Toddler , 18, 21, 24, 34, 38, 44 and 57.

Table 5. Results Grouping Based on Nutrition Status

A toddler	**Cluster**	**Distance**
1	4	5,276
2	1	4,054
3	1	2,190
4	4	6,496
5	2	2,449
6	3	4,794
7	4	8,789
8	2	8,839
9	2	6,961
10	2	2,941
11	1	4,025
12	1	8,249
13	1	4,079
14	1	7,919
15	2	2,892
16	2	3,228
17	2	5,214
18	4	7,293
19	1	3,432
20	2	3,795
21	4	5,017
22	1	1,475
23	2	5,193
24	4	3,915
25	1	3,971
26	3	3,548
27	1	2,916
28	1	8.475
29	1	4,085
30	1	7,288
31	2	3,248
32	3	3,648
33	2	0.827
34	4	2,960
35	1	3,432
36	1	3,882
37	1	6,161
38	4	2,169
39	3	0.769
40	1	3,999
41	1	5,091
42	2	3,003
43	2	1,487
44	4	4,879
45	2	3,188
46	1	4,497
47	1	1,172
48	2	3,177
49	1	4,112
50	2	3,295
51	3	3,312

52	3	2,783
53	2	3,281
54	1	8,171
55	1	3,931
56	3	2,923
57	4	0.923

Table 6. Results of Cluster Analysis using the K-Means Method

		N	The mean	Std. Deviation
1	Age	23	44,13	4.7
	Weight	23	13.69	2.19
2	Age	17	21.18	4.14
	Weight	17	8.88	1.07
3	Age	7	56.71	2.87
	Weight	7	12.91	2.13
4	Age	10	8.2	4.76
	Weight	10	8.34	2.89

Based on table 6, the K-Means grouping shows the poor nutritional status, including the cluster with the most members showing an average body weight of 13.69 kg with a standard deviation of 2.19 kg, and an average age of toddlers 44.13 months with a standard deviation of 4.70 months. While the least number of cluster members is good nutritional status, which shows an average body weight of 12.91 kg with a standard deviation of 2.90 kg, and an average toddler age of 56.71 months with a standard deviation of 2.87 months.

Accuracy of Nutrition Status Classification Results Using Anthropometric Handbooks with Classification Using K-Means Method

Table 7. Results of the Cross Tabulation of Cluster Analysis using K-Means Method and the Anthropometry Book

	Anthropometry Book			
K-Means	2	3	4	Total
1	19	4	0	23
2	9	5	3	17
3	3	3	1	7
4	9	1	0	10
Total	40	13	4	57

Based on table 7 on the results of cross-tabulation from cluster analysis with the anthropometric pocketbook with the results of grouping using the K-Means method, the Total Accuracy Rate (TAR) value is 0.44, calculating the total accuracy rate = $\frac{19+5+1}{57}$= 0.44. These results indicate the level of accuracy is still lacking with a benchmark of 70%.

DISCUSSION

Descriptive analysis of the nutritional status of children under five in the Mayangrejo Ponkesdes, Bojonegoro in 2019, the average age of toddlers is 32.53 months and the average underweight of children under 5-year-old is 11.22 kg. Based on the Kepmenkes 2010 standard weight according to age, the category of nutritional status of children under 5-year-old in Mayangrejo Ponkesdes, Bojonegoro District is good.

Cluster analysis using the K-means method on anthropometric data at Mayangrejo Ponkesdes can be done by grouping nutritional status with (weight/age) parameters. The cluster analysis model supports grouping of nutritional status, so that toddlers with nutritional status: gradation 1 (poor, far from anthropometric standards), gradation 2 (lacking, close to anthropometric standards), grade 3 (good, approaching anthropometric standards), gradation 4 (more, very close to anthropometric standards). Based on the results of the final centroid center, it was found that cluster 1 was nutritional status in gradation 3, cluster 2 had nutritional status in gradation 2, cluster 3 had nutritional status in gradation 4, and cluster 4 had nutritional status in gradation 1.

The results showed in cluster 1, the grouping of nutritional status based on body weight according to the age, obtained 23 toddlers. One of the toddlers in cluster 1 was taken as an example to discuss, namely, the 2nd toddler weighing 14.9 kg and 48-month-old. The second toddlers according to the results of grouping by the K-Means method including cluster 1, which classified as toddlers with good nutritional status. Thingit, a researcher, states that the average weight of a toddler is classified into the group of good nutritional status and grouping this is in line with the anthropometric book which explains that children with good nutritional status will weigh between 14.5-19 kg in the 48-72 month age group (Indonesian Ministry of Health Decree Number 1995 / MenKes / SK / XII / 2010 Concerning Standar Antropometri Penilaian Status Gizi Anak).

The results showed in cluster 2, the grouping of nutritional status based on body weight according to the age are 17 toddlers. One of the toddlers in cluster 2 was taken as an example to be discussed, namely the 16th toddler weighing 8.3 kg and 18-month-old. The 16th toddlers according to the results of the grouping by the K-Means method are included in cluster 2, which classified as toddlers with malnutrition status. Thingit, a researcher, states that the average weight of a toddler belongs to the group of malnutrition is in line with the anthropometric book that explains that children with poor nutritional status will weigh between 7.8-8.9 kg in the 18-19 months age group (Indonesian Ministry of Health Decree Number 1995 / MenKes / SK / XII / 2010 Concerning Standar Antropometri Penilaian Status Gizi Anak). This can occur because the growth process of infants is influenced by age factors, including biological risk factors. The younger the age, the more chances that the immune system has not yet developed resulting in toddlers easily affected by malnutrition (Lestari, 2016).

The results showed in cluster 3, the grouping of nutritional status based on body weight according to the age obtained 7 toddlers. One of the toddlers in cluster 3 was taken as an example to be discussed, namely, the 26th toddler weighing 15.2 kg and 54-month-old. The 26th toddler according to the results of grouping by K-Means method is included in cluster 3, which classified as over-nutritional status. Thingit states that the average weight of a toddler belongs to over-nutritional status and grouping is not in line with the anthropometric book which explains that children who weigh 15.2 kg and 54-month-old is considered as having good nutritional status (Indonesian Ministry of Health Decree Number 1995 / MenKes / SK / XII / 2010 Concerning Standar Antropometri Penilaian Status Gizi Anak). More nutritional status conditions can occur from the effects of malnutrition during pregnancy due to metabolic tissues, such as hypothalamus, which is inaccurate programming that inhibits appetite control in infants, resulting in obesity (Utami, Putri and Rosa, 2014).

The results showed in cluster 4, the grouping of nutritional status based on body weight according to the age obtained 10 toddlers. One of the toddlers in cluster 4 was taken as an example to be discussed, namely the 4th toddler with a bodyweight of 6.4 kg and 2-month-old. The 4th toddler, according to the results of grouping with the K-means method, is included in cluster 4, which is classified as toddler with poor nutritional status. It states that the average underweight for children under the category of malnutrition status and grouping is not in line with the anthropometric book which explains that children who weigh between 3.3-7.9 kg in the age group 0-6 months are considered having good nutritional status (Indonesian Ministry of Health Decree Number 1995 / MenKes / SK / XII / 2010 Concerning Standar Antropometri Penilaian Status Gizi Anak).

From the results of the cluster analysis, it can be concluded that K-Means method is still not informative enough and does not provide a good standard deviation estimation so that it still needs to be done to add another algorithm to provide an accurate one. This research is in line with the opinion of Efron and Robert (1998) that cluster analysis will produce a good standard deviation estimation when using the Bootstrap resampling approach for each cluster.

The measuring accuracy of the toddlers' nutritional status with anthropometric pocketbooks compared to the K-Means method showed an accuracy of 0.44. This accuracy value is still not quite right with a benchmark of 70% because there is still a mistake in the classification so that it still has not been able to prove a good level of accuracy in the process of grouping nutritional status using K-Means method.

CONCLUSIONS AND SUGGESTIONS

Conclusion

The results of this study indicate that the nutritional status of children under five-year-old in the Mayangrejo Ponkesdes can be clustered using K-Means method through weight parameters according to age in clusters 4, namely cluster 1 with 23 toddlers in poor nutritional status, cluster 2 with 17 toddlers in poor nutritional status, cluster 3 with 7 toddlers in good nutritional status and cluster 4 with 10 toddlers in over-nutritional status.

Suggestion

Based on the results of the study, it is recommended that further research be done by trying other algorithms to examine better

accuracy in grouping the children's nutritional status to provide accurate estimation and good information so that the method can be used as an alternative to determine the classification.

REFERENCES

Agusta, Y., 2007. K-Means-Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*. 3 (February): pp.47–60.

Bastian, A., Sujadi, H., and Febrianto, G., 2018. Penerapan Algoritma K-Means Clustering Analysis pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). *Jurnal Sistem Informasi*. 14 (1), pp. 26–36

Ediyanto, Mara, M.N., and Satyahadewi, N., 2013. Pengklasifikasian Karakteristik dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah*. 2 (2), pp. 133–136.

Efendi, R., Coastera, F.F., and Tanjung, F.R., 2019. Pengelompokan dan Pemetaan Derajat Kesehatan Kota Bengkulu dengan Metode K-Means Clustering. *Jurnal Rekursif*. 7 (1), pp. 91–97.

Efron, B., and Robert J. T., 1998. *An Introductio to the Bootstrap*. Second Edition. United States of America: Chapman & Hall.

Han, J., Kamber, M., and Pei, J. 2011. *Data Mining. Concepts and Techniques*. 3rd Edition. Waltham: The Morgan Kaufmann Series in Data Management Systems.

Indonesian Ministry of Health, 2003. *Indikator Indonesia Sehat 2010*. Jakarta.

Indonesian Ministry of Health, 2004. *Petunjuk Teknis Standar Pelayanan Minimal (SPM) Penyelenggaraan Perbaikan Gizi Masyarakat*. Jakarta.

Indonesian Ministry of Health, 2010. *Hasil Utama Riskesdas 2018*. Jakarta.

Indonesian Ministry of Health Decree Number 1995/MenKes/SK/XII/2010 Concerning Standar Antropometri Penilaian Status Gizi Anak.

Indonesian Ministry of Health Regulation Number 155/Menkes/Per/I/2010 Concerning Penggunaan Kartu Menuju Sehat (KMS) Bagi Balita.

Irfani, E., and Rani, S.S., 2018. Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita. *Jurnal Sistem dan Teknologi Informasi (JUSTIN).* 6(4), pp.165–172.

Kuntoro. 2011. *Metode Statistik*. Revised Edition. Surabaya: Melati.

Lestari, N.D., 2016. Analisis Determina Gizi Kurang pada Balita di Kulon Progo, Yogyakarta. *Indonesia Journal of Nursing Practices*. 1 (1): pp. 15–21.

Prasetyo, H., Kuntoro, Purnomo, W., Melaniani, S., Adriani, M., and Otok, B.W., 2014. Penerapan Clustering Bootstrap dengan Metode K-Means. *Jurnal Biometrika dan Kependudukan*. 3 (1), pp. 43–49.

Rahim, F. K. 2014. Faktor Risiko Underwight Balita Umur 7-59 Bulan. *Jurnal Kesehatan Masyarakat*. 9 (2), pp. 115–121.

Rohmawati, N., Defiyanti, S., Jajuli, M., 2015. Implementasi Algoritma K-Means dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jitter*. I (2): pp. 62–68.

Singla, A., and Karambir. 2012. Comparative Analysis & Evaluation of Euclidean Distance Function and Manhattan Distance Function Using K-means Algorithm. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2 (7), pp. 63–68.

Sundari, S.S., and Ariani, N., 2019. Penerapan Data Mining untuk Pengelompokan Penyakit dengan Algoritma Fuzzy C-Means (Studi Kasus: UPT Puskesmas Salawu). *Jurnal Voice of Informatics*. 8 (2), pp. 63–67.

Utami, N.H., Putri, D.S. ., and Rosa, B.C., 2014. Kejadian Pendek Gemuk pada Anak Usia Bawah Dua Tahun Berhubungan dengan Konsumsi Lemak dan Pendidikan Ibu. *Penel Gizi Makan*. 37 (1): pp. 1–10.

Witten, I. H., Frank, E., and Hall, M.A., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Third Edition. San Francisco: The Morgan Kaufmann Series in Data Management Systems.