

Clustering of Child Nutrition Status using Hierarchical Agglomerative Clustering Algorithm in Bekasi City

¹Ozzi Ardhiyanto, ²Muhammad Salam Asyidqi, ³Ajif Yunizar Pratama Yusuf, S.Si, M.Eng, ⁴Dr. Tb. Ai Munandar, S.Kom., MT

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

^{1,2,3,4}*Informatics, Computer Science, Universitas Bhayangkara Jakarta Raya*

e-mail: ¹ozziardhi@gmail.com, ²salamassyidqi958@gmail.com,

³ajifyunizar@dsn.ubharajaya.ac.id, ⁴tb.aimunandar@dsn.ubharajaya.ac.id



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Corresponding Autor: Ozzi Ardhiyanto

Abstract

Clustering infant nutrition based on weight, height, and age is a data analysis method used to group infant nutritional status based on these characteristics. The research on clustering infant nutrition aims to analyze whether there are still many infants in the area with insufficient or excessive nutrition, and to identify groups of infants requiring special attention regarding their nutritional intake. In the analysis of infant nutrition clustering, data on weight, height, and age of infants are collected and then grouped based on similarities in body height and weight at certain ages. The method used in this research is hierarchical clustering, which can help in grouping the data. Clustering analysis can help understand how infants' feeding patterns vary based on their weight, height, and age. The results of research on clustering infant nutrition based on weight, height, and age can provide valuable insights for nutrition experts, pediatricians, and community health workers in developing appropriate intervention programs to improve infant feeding patterns and meet their nutritional needs. Additionally, the results of clustering infant nutrition can also be used to identify groups of infants requiring special attention regarding their nutritional needs, thus minimizing the risk of malnutrition and unhealthy growth in infants.

Keywords— Clustering, Infant nutrition, Agglomerative Clustering, Malnutrition, Nutritional grouping.

1 Introduction

Nutrition is an element contained in food that can provide a benefit to the body that when consuming it can be healthy [1]. The process of growth and development during infancy and toddlerhood is a very important process in determining the future of children both physically, mentally and behaviourally.[2] Attention to the physical growth of children under five from parents, child caregivers, teachers, cadres, midwives, to the level of health centres and health offices. The results of the National Nutrition Monitoring recorded in 2016 in Indonesia as many as 3.4% of toddlers had malnutrition status and 14.4% of toddlers had malnutrition status, while in the Special Region of Yogyakarta there were 2.1% experiencing malnutrition and 13.8% experiencing malnutrition.[3] One technique that can be used to determine the health level of infants and toddlers can be done with data mining techniques.[4] Of the several models that can be applied to data, one of them is population segmentation, in this case clustering (grouping)[5]. In the single linked (closest distance) method, the closeness between two groups is determined from the closest (smallest) distance between pairs of data from two different groups (one from the first group and one from the other group, or also called the maximum similarity).[6] In research on clustering baby nutrition based on anthropometry, the baby data is collected and then grouped based on similarity or similarity. The clustering method that will be used is the agglomerative clustering method with a linkage system. where the cluster will be determined between the closest data to determine the closest distance.[7] Therefore, research on infant nutrition clustering based on weight, height, and age is very important to ensure optimal health and development of infants in an area.

©2023 Ardhiyanto et al.

2 Literature Study

In this study, cluster analysis is an element of the data mining process that functions to group objects into a data cluster. Cluster is a group of data objects that have similarities in the same cluster. [8] the cluster method chosen is AHC (Agglomerative Hierarchical Clustering), there are several advantages, namely there is no need to determine the desired number of clusters because the process can be stopped immediately when the number of clusters is as desired. [9] The formation of a cluster has a method that is usually categorised based on the type of cluster produced [10]. AHC (Agglomerative Hierarchical Clustering), Hierarchical agglomerative clustering is a bottom-up characterised hierarchical clustering method in the form of combining n clusters until they are formed into one cluster.[8] clusters will be applied in clustering infant nutrition because, Malnutrition, namely malnutrition or Protein Energy Deficiency (PEM) and micronutrient deficiencies are problems that require special attention, especially in developing countries, which are important risk factors for morbidity and mortality in pregnant women and children under five.[11], Efforts in preventing the risk of malnutrition in toddlers can be done through improving aspects of maternal knowledge, attitudes and behaviour." [12] In this data clustering process, it starts from determining the number of clusters to be formed. Of the 150 toddler data, it will be grouped into 5 clusters namely Malnutrition, Under Nutrition, Good Nutrition, Over Nutrition and Obesity [6]. This research uses the Orange data science program Orange is an open source machine learning technology or data mining software [13]. Results and Discussion

3 Research methods

3.1 Data Gathering

The research was conducted by collecting data on toddlers from all posyandu in one of the health centres in Bekasi city. Then the data that has been obtained is processed to determine what data will be used as a cluster.

3.2 Preprocessing Data

Preprocessing is a stage carried out to extract or retrieve information from unstructured data [14].

There are several things that researchers do before doing clusters:

1. In the data obtained there are some that have no value or are empty in this case the researcher fills in the empty value with the adjusted value.
2. selecting data to be clustered the data selected for clustering.

3.3 Determining Single Linkage Distance

The approach in measuring the similarity of an object with other objects is to use the distance between pairs of objects. Pairs of objects that have a closer distance will be more 'similar' in characteristics than pairs of objects that have a greater distance. One method to measure the distance between objects is by using Euclidian Distance. Suppose there are two objects (u and v) and as many as p variables, then the Euclidian Distance is :

$$d_{u,v} = \sqrt{(u_1^2 - v_1^2) + (u_2^2 - v_2^2) + \dots + (u_p^2 - v_p^2)}$$

Then after knowing the Euclidian Distance of each pair of objects in the data, clustering can be done based on the following methods:

Single Linkage

Clusters are formed based on pairs of objects with the closest distance.

$$d_{u,v}w = \min(d_{u,w}, d_{v,w})$$

Single Linkage Method or known as single link is an AHC method used to form clusters based on the closest distance between objects to one another [15].

4 Results and Discussion

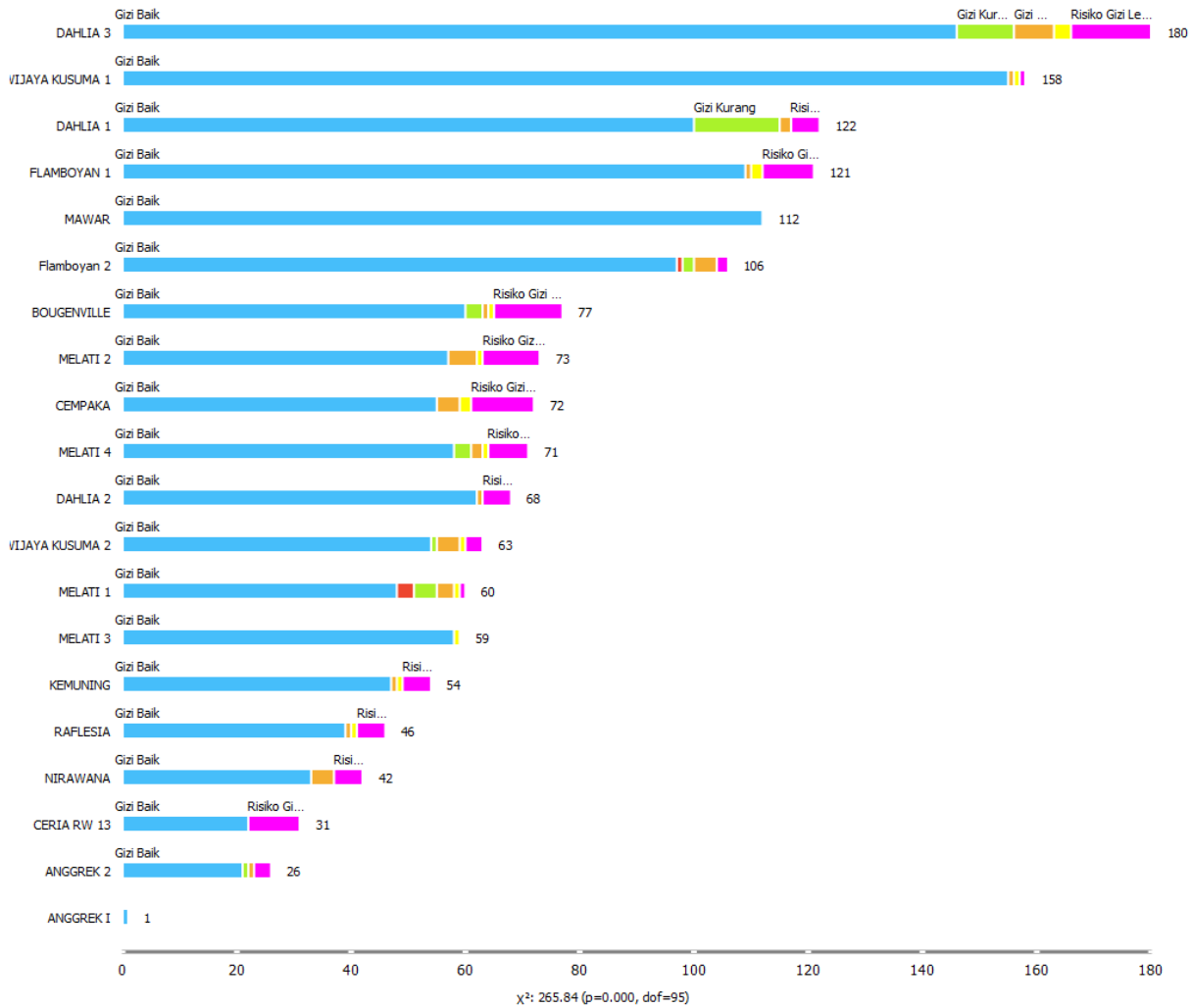


Figure 1 Box Plot cluster diagram result

In the Box Plot Diagram above, the number of children from each posyandu with nutritional information is shown.

- Blue coloured diagram Good nutrition
- Red coloured diagram Poor nutrition
- Yellow coloured diagram Obese Nutrition
- Purple coloured diagram Risk of over nutrition
- Green coloured diagram Undernutrition
- Orange coloured diagram Overnutrition

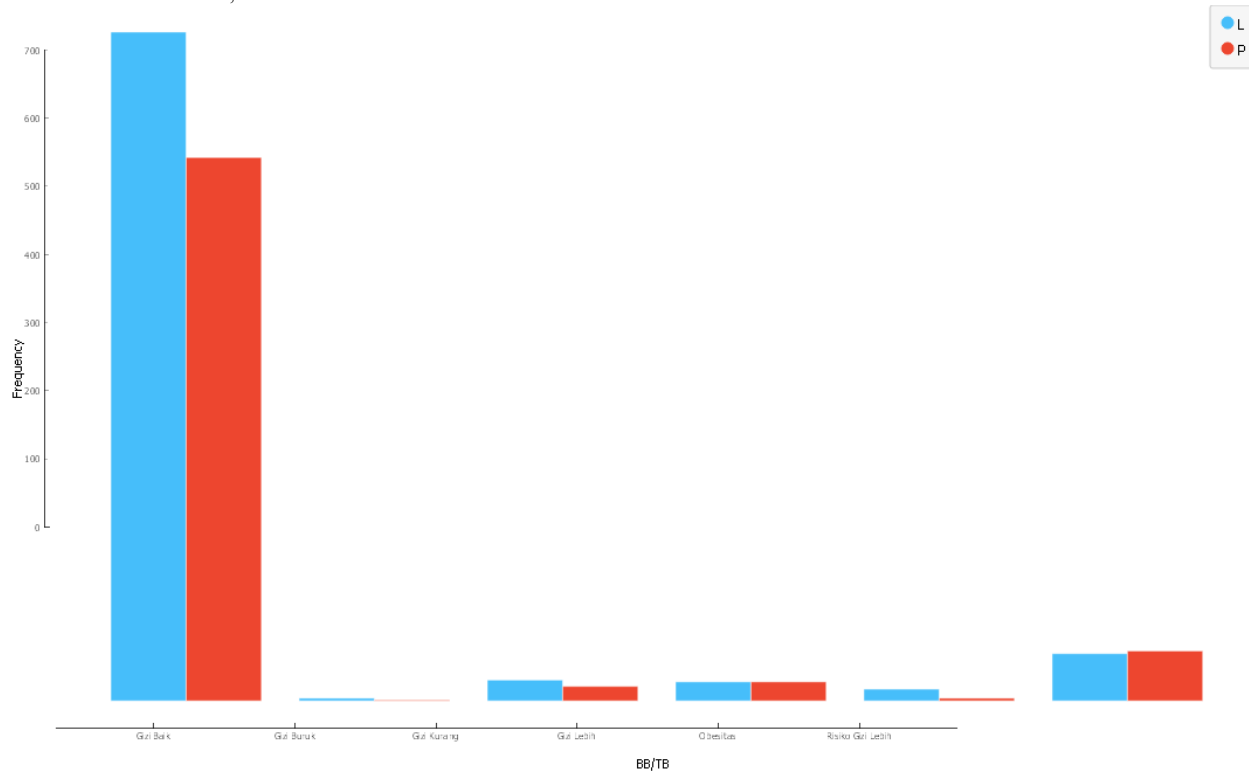


Figure 2 Comparison of male and female nutrition

The bar chart above shows each nutritional status of the entire posyandu based on male and female gender.

- Good nutrition, more males than females
- Poor nutrition, more males than females
- Undernutrition, more males than females
- Overnutrition, equal number of males and females
- Obesity, more males than females
- At risk of overnutrition, more females than males

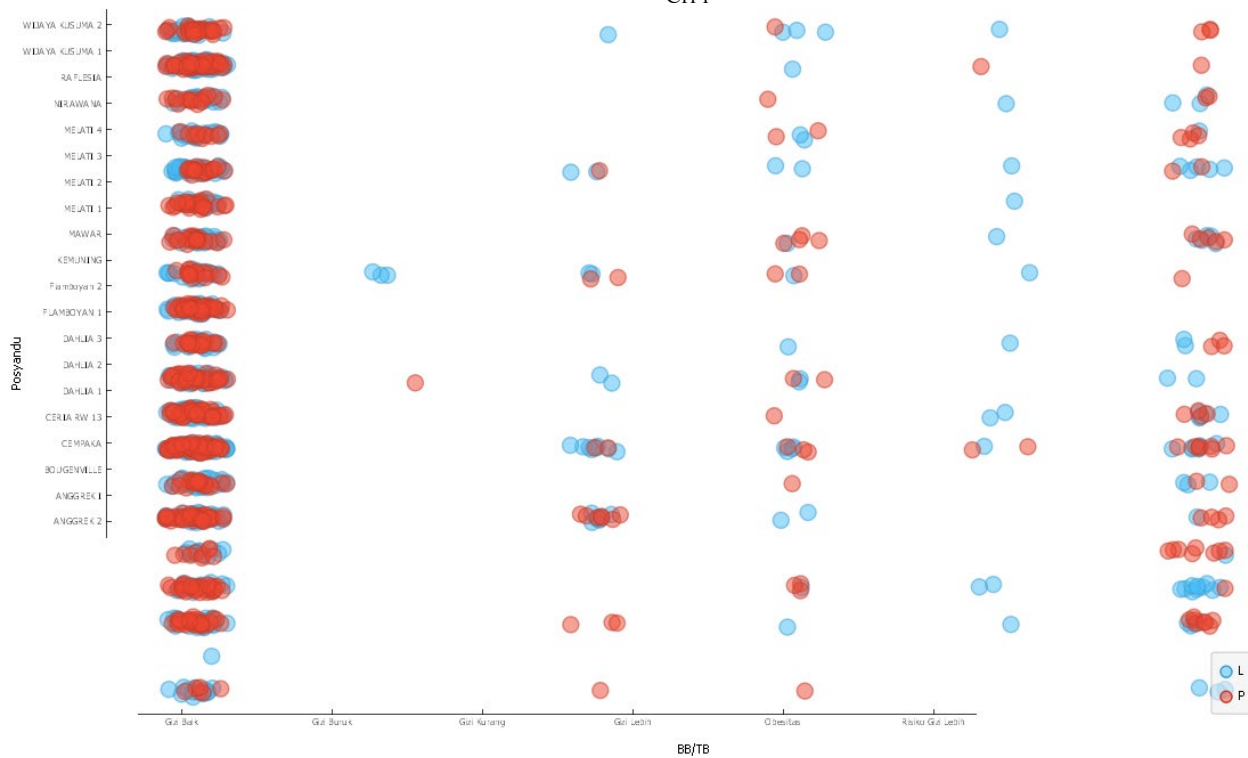


Figure 3 scatter plot of nutrition, JK, and posyandu

The scatter plot diagram above shows the nutritional status based on posyandu and gender.

- All Posyandu have good nutrition
- Malnutrition is found in Posyandu Flamboyant 2 with female gender and MELATI 1 with male gender.
- Undernutrition in wijaya kusuma 2 and flamboyant 2 posyandu is only male and for bougenville and orchid 2 posyandu is only female.
- Overnutrition, the number is equal between men and women
- Obesity from all posyandu is dominated by men
- Risk of overnutrition, more women than men and almost in every posyandu



Figure 4 Dendrogram Single linkage

5 Conclusion

Based on the analysis and clustering done:

1. Single linkage agglomerative hierarchical clustering algorithm is an unsupervised learning algorithm by combining two clusters and so on into a new cluster based on similarity determined using Euclidean distance with the average distance criteria of all individuals in the cluster.
2. From the analysis that has been done From the results of the calculation and clustering above, it can be seen that the comparison based on nutritional status between gender, and posyandu, the results obtained are more "Good Nutrition" status than other nutritional status in one of the sub-districts in the city of Bekasi.

6 Suggestion

Suggestions for future research

1. can use more new data.
2. future research should take data from several health centres to make comparisons
3. it is advisable to do with manual calculations with the Python or R programming language

BIBLIOGRAPHY

- [1] A. Ayu Fitriana STKIP Modern Ngawi Jl Ir Soekarno Kabupaten Ngawi, “Fitriana, Sumiatun: Pemahaman Orang Tua Mengenai Gizi Anak Pemahaman Orang Tua Mengenai Gizi Anak.”
- [2] A. Audilla, Stik. Utama Abdi Husada, and K. Penulis, “Jurnal Ilmiah Pamenang-JIP HUBUNGAN PENGETAHUAN IBU TENTANG NUTRISI DENGAN STATUS GIZI BALITA DI POSYANDU DESA WAJAK LOR KECAMATAN BOYOLANGU KABUPATEN TULUNGAGUNG TAHUN 2022,” vol. 4, no. 1, pp. 22–26, 2022, doi: 10.53599.
- [3] U. Jenderal, A. Yani, and Y. Abstrak, “ANTROPOMETRI PENGUKURAN STATUS GIZI ANAK USIA 24-60 BULAN DI KELURAHAN BENER KOTA YOGYAKARTA Reni Merta Kusuma, Rizki Awalunisa Hasanah,” 2018.
- [4] G. Tapanuli, U. Oleh, D. W. Sitohang, and A. Rikki, “Pulo Brayan Darat I, Kec,” 2019.
- [5] G. Abdurrahman, “Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage”.
- [6] S. Gizi Balita Di Kelurahan Jumpat Rejo Sukodono Sidoarjo Amir Ali, R. Medis dan Informasi Kesehatan Stikes Yayasan Rumah Sakit DrSoetomo DrMoestopo, and A. Surabaya, “Clustering Data Antropometri Balita Untuk Menentukan,” 2020. [Online]. Available: <http://jurnal.mdp.ac.id>
- [7] H. Februariyanti and D. B. Santoso, HIERARCHICAL AGGLOMERATIVE CLUSTERING UNTUK PENGELOMPOKAN SKRIPSI MAHASISWA. 2017.
- [8] E. Widodo, P. Ermayani, L. N. Laila, and A. T. Madani, “Pengelompokan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering (Indonesian Province Grouping Based on Poverty Level Using Hierarchical Agglomerative Clustering Analysis).”
- [9] L. Zahrotun, “ANALISIS PENGELOMPOKAN JUMLAH PENUMPANG BUS TRANS JOGJA MENGGUNAKAN METODE CLUSTERING K-MEANS DAN AGGLOMERATIVE HIERARCHICAL CLUSTERING (AHC),” 2015.
- [10] H. Februariyanti et al., “I N F O R M A T I K A ANALISIS KECENDERUNGAN INFORMASI MENGGUNAKAN ALGORITMA HIERARCHICAL AGGLOMERATIVE CLUSTERING,” Jurnal Informatika, Manajemen dan Komputer, vol. 13, no. 1, 2021.
- [11] T. Regina, A. W. Program, S. Keuangan, and P. P. Swadharma, “TATA KELOLA PEMERIKSAAN GIZI BAYI, BALITA DAN USIA LANJUT.”
- [12] D. P. Lestari, “Upaya Pencegahan Risiko Gizi Buruk pada Balita: Literature Review,” Jurnal Ilmiah Universitas Batanghari Jambi, vol. 22, no. 1, p. 532, Mar. 2022, doi: 10.33087/jiubj.v22i1.1828.
- [13] S. Indah Nurhafida and F. Sembiring, “ANALISIS TEXT CLUSTERING MASYARAKAT DI TWITER MENGENAI MCDONALD’SXBTS MENGGUNAKAN ORANGE DATA MINING,” 2021.
- [14] 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>
- [15] R. T. Aldisa, “JURNAL MEDIA INFORMATIKA BUDIDARMA Data Mining Penentuan Jurusan Siswa Menggunakan Metode Agglomerative Hierarchical Clustering (AHC),” 2023, doi: 10.30865/mib.v7i2.6092.