Jurnal Riset Informatika dan Teknologi Informasi (JRITI) ISSN (Print) : 3024-8167
Jejaring Penelitian dan Pengabdian Masyarakat (JPPM)
Vol. 2, No. 1, Agustus – November 2024, Hal. 126 - 132
Web : https://ejurnal.jejaringppm.org/index.php/jriti

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

[1]Ozzi Ardhiyanto, [2]M. Salam Asyidqi, [3]Ajif Yunizar Pratama Yusuf, dan [4]Tb. Ai Munandar
Informatics, Computer Science, Universitas Bhayankara Jakarta Raya
e-mail: [1]ozziardhi@gmail.com, [2]psalamassyidqi958@gmail.com, [3]ajifyunizar@dsn.ubharajaya.ac.id, [4]tb.aimunandar@dsn.ubharajaya.ac.id

ABSTRAKSI
Pengelompokan (clustering) nutrisi bayi berdasarkan berat badan, tinggi badan, dan usia merupakan metode analisis data yang digunakan untuk mengelompokkan status gizi bayi berdasarkan karakteristik tersebut. Penelitian tentang clustering nutrisi bayi bertujuan untuk menganalisis apakah masih terdapat banyak bayi di suatu daerah yang mengalami kekurangan atau kelebihan nutrisi, serta untuk mengidentifikasi kelompok bayi yang memerlukan perhatian khusus terkait asupan gizi mereka. Dalam analisis clustering nutrisi bayi, data berat badan, tinggi badan, dan usia bayi dikumpulkan, kemudian dikelompokkan berdasarkan kesamaan tinggi dan berat badan pada usia tertentu. Metode yang digunakan dalam penelitian ini adalah hierarchical clustering, yang dapat membantu mengelompokkan data tersebut secara efektif. Analisis clustering ini dapat membantu memahami bagaimana pola pemberian makan bayi bervariasi berdasarkan berat badan, tinggi badan, dan usia mereka. Hasil penelitian tentang clustering nutrisi bayi berdasarkan berat badan, tinggi badan, dan usia dapat memberikan wawasan yang berharga bagi ahli nutrisi, dokter anak, dan petugas kesehatan masyarakat dalam mengembangkan program intervensi yang tepat untuk memperbaiki pola pemberian makan bayi serta memenuhi kebutuhan gizi mereka. Selain itu, hasil clustering nutrisi bayi juga dapat digunakan untuk mengidentifikasi kelompok bayi yang memerlukan perhatian khusus dalam hal kebutuhan nutrisinya, sehingga risiko malnutrisi dan pertumbuhan yang tidak sehat pada bayi dapat diminimalisasi..

Kata Kunci: *pengelompokan,nutrisi bayi, pengelompokan aglomeratif, malnutrisi, pengelompokan nutrisi.*

ABSTRACT
Clustering infant nutrition based on weight, height, and age is a data analysis method utilized to categorize infants according to their nutritional status. Research involving infant nutrition clustering aims to determine whether a significant number of infants in a given area experience insufficient or excessive nutrition and to identify specific groups of infants requiring targeted nutritional interventions. In conducting this analysis, data on infants' weight, height, and age are collected and then grouped according to similarities in body measurements relative to age. The hierarchical clustering method is applied in this research to effectively organize and interpret the data. Through clustering analysis, researchers and practitioners can gain a deeper understanding of how feeding patterns and nutritional statuses vary among infants based on their physical growth parameters. The outcomes of clustering infant nutrition provide valuable insights to nutrition experts, pediatricians, and community health workers, assisting them in formulating appropriate intervention strategies aimed at enhancing feeding practices and meeting infants' nutritional requirements. Furthermore, clustering results help identify infants who need special nutritional attention, thus reducing the risks of malnutrition and unhealthy growth during early developmental stages.

Keywords: *clustering, infant nutrition, agglomerative clustering, malnutrition, nutritional grouping*

Penulis Korespondensi
Ozzi Ardhiyanto

Tanggal Submit : 03/07/2023
Tanggal Diterima : 13/01/2024
Tanggal Terbit : 30/08/2024

I. INTRODUCTION

Nutrition encompasses the essential components in food that provide health benefits when consumed properly [1]. These nutrients—ranging from macronutrients like proteins, carbohydrates, and fats to micronutrients such as vitamins and minerals—are vital for energy provision, tissue repair, immune function, and cognitive performance. A balanced intake of these components is particularly critical during early life to support rapid growth and prevent disease. Consequently, understanding the role and adequacy of nutrition lays the groundwork for assessing dietary status and designing interventions for young children.

The period of infancy and early childhood represents a critical window for growth and development, during which physical, cognitive, and behavioral trajectories are established [2]. During this stage, rapid physiological changes occur that set the course for lifelong health, learning ability, and behavioral outcomes. Any disruption—such as nutrient deficiencies or imbalances—can have lasting effects on a child's motor skills, cognitive development, and susceptibility to illness. Therefore, continuous monitoring of growth indicators (e.g., length/height-for-age, weight-for-age, weight-for-length) is essential. This vital task demands coordinated efforts among parents, caregivers, healthcare providers, educators, and community health workers.

According to the 2016 National Nutrition Monitoring in Indonesia, 3.4% of toddlers were classified as malnourished and 14.4% were recorded as stunted across the country [3]. In the Special Region of Yogyakarta, the prevalence was slightly lower, with 2.1% of children experiencing malnutrition and 13.8% facing stunting. These statistics underscore the persistent challenge of undernutrition and impaired growth among young children at both national and regional levels. They also highlight the urgency of targeted nutritional surveillance and intervention programs. Without such measures, these conditions can compromise physical development and long-term health outcomes.

One promising approach to assess the health status of infants and toddlers involves applying data mining techniques to nutritional and anthropometric datasets [4]. Data mining enables the handling of large, complex data and uncovers hidden patterns that might not be evident through traditional analyses. Among these techniques, clustering—an unsupervised learning method—effectively segments the population into groups with similar characteristics [5]. This segmentation facilitates the identification of subgroups at risk of malnutrition or overnutrition and informs the design of tailored intervention strategies. By leveraging clustering, practitioners can optimize resource allocation and evaluate the effectiveness of nutritional programs in real time.

In hierarchical clustering, specifically the single-linkage (nearest neighbor) method, the similarity between two clusters is determined by the smallest Euclidean distance between any pair of data points—one from each cluster—thereby reflecting maximum similarity [6]. This approach tends to form elongated, chain-like clusters because it merges clusters based on their closest members. When applied to infant anthropometric measures (weight, height, and age), the method can reveal nuanced gradations in growth patterns and nutritional status. By iteratively merging clusters with minimal inter-cluster distances, researchers can construct a dendrogram that visualizes these hierarchical relationships. Such a dendrogram provides a clear overview of how infants group together based on their growth parameters.

This research employs an agglomerative clustering approach with a linkage criterion, wherein clusters are iteratively merged based on the minimum distance between data points [7]. Beginning with each infant as an individual cluster, the algorithm successively fuses the two closest clusters until the desired number of clusters is reached or a distance threshold is exceeded. The resulting clusters correspond to groups of infants with comparable weight-for-age and height-for-age growth trajectories. Consequently, clustering infant nutrition by weight, height, and age offers a robust framework for identifying distinct nutritional patterns and targeting interventions. Ultimately, these insights can support the development of evidence-based strategies to optimize early childhood health and developmental outcomes.

II. LITERATURE STUDY

Cluster analysis is a fundamental component of the data mining process, tasked with organizing data objects into clusters such that objects within the same cluster exhibit high degrees of similarity while differing markedly from objects in other clusters [8]. By leveraging similarity measures—such as Euclidean distance, Manhattan distance, or cosine similarity—cluster analysis uncovers inherent groupings in complex datasets without requiring predefined labels. In this way, cluster analysis supports exploratory data analysis, enabling researchers to detect patterns, anomalies, and natural segmentations. When applied correctly, it provides actionable insights that inform decision-making in diverse domains, including healthcare, marketing, and social sciences.

The Agglomerative Hierarchical Clustering (AHC) method was chosen for this study due to its flexibility and interpretability [9]. Unlike partition-based algorithms (e.g., k-means), AHC does not necessitate a preset number of clusters; researchers can halt the merging process once a clustering solution meets their criteria. Furthermore, cluster formation in AHC can be categorized by the linkage criterion used—such as single linkage, complete linkage, or average linkage—

which determines how inter-cluster distances are calculated and thus shapes the resulting cluster structure [10]. This capacity to adjust linkage strategies makes AHC highly adaptable to varying data characteristics and research objectives.

Hierarchical agglomerative clustering operates via a bottom-up approach, beginning with each data point as its own cluster and successively merging the two closest clusters at each iteration [8]. The "single linkage" criterion, also known as the nearest-neighbor method, merges clusters based on the smallest pairwise distance between their members, thereby yielding elongated, chain-like clusters that capture local similarities. As the algorithm progresses, a dendrogram—a tree-like diagram—is constructed to visualize the sequence of merges, offering both a qualitative and quantitative perspective on cluster relationships. Ultimately, the algorithm terminates when all data points have been merged into a single cluster or when a predefined stopping condition (e.g., distance threshold or desired cluster count) is met.

Clustering infant nutrition data is especially pertinent given the global burden of malnutrition, which encompasses Protein Energy Malnutrition (PEM) and micronutrient deficiencies. These conditions remain leading risk factors for morbidity and mortality among pregnant women and children under five in many developing regions [11]. By grouping infants according to weight, height, and age, AHC can identify subpopulations at heightened risk of undernutrition, overnutrition, or obesity. Such granular segmentation supports the design of targeted nutritional interventions and monitoring strategies tailored to the specific needs of each cluster.

Preventing malnutrition in toddlers entails more than data-driven detection; it also requires improving maternal knowledge, attitudes, and behaviors concerning infant feeding practices [12]. The clustering process in this study begins by determining the desired number of clusters—set at five—to delineate distinct nutritional statuses: Malnutrition, Undernutrition, Good Nutrition, Overnutrition, and Obesity [6]. By applying AHC to a dataset of 150 toddler records, researchers can map each child to one of these categories, facilitating the identification of at-risk groups and the evaluation of intervention outcomes over time.

All analyses and visualizations in this research are conducted using the Orange data mining platform, an open-source software suite that offers a user-friendly interface for machine learning and data analysis tasks [13]. Orange simplifies the implementation of hierarchical clustering through interactive workflows, enabling rapid prototyping and validation of clustering results. Its built-in widgets for data preprocessing, clustering, and visualization streamline the research process, allowing practitioners to focus on interpreting findings and translating them into practical public health recommendations.

III. RESEARCH METHODS

4.1. 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.

4.2. 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.

4.3. *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 :

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

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].

IV. RESULTS AND DISCUSSION

Figure 1 illustrates a Box Plot diagram representing the clustering results of children's nutritional status collected from various Posyandu (community health posts). Each distinct color in the diagram indicates a specific nutritional category: blue signifies good nutrition, red indicates poor nutrition, yellow represents obese nutrition, purple denotes risk of overnutrition, green signifies undernutrition, and orange represents overnutrition. From this visualization, clear differences are observable in the distribution of the data across each nutritional category in relation to the number of children attending each Posyandu.

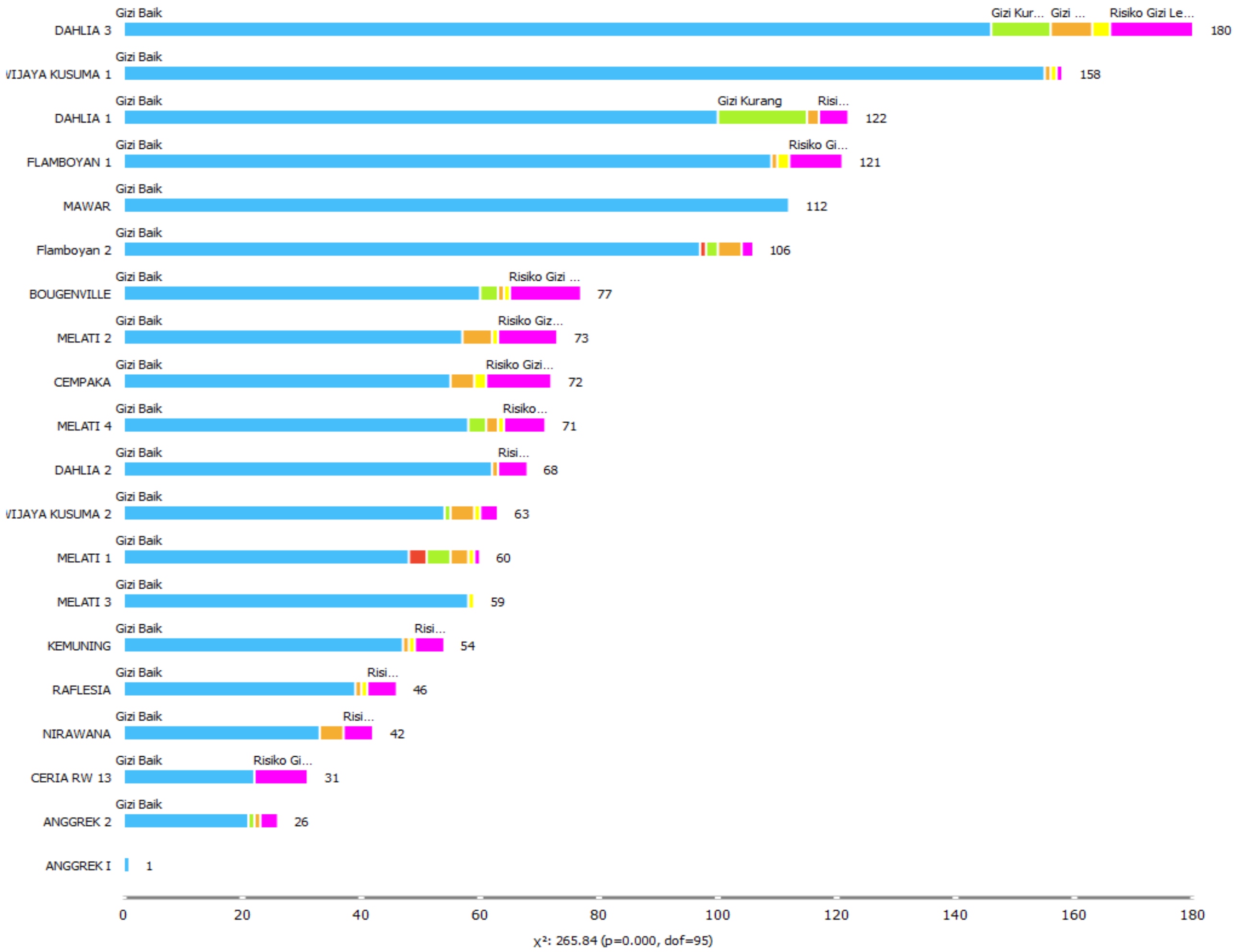

Figure 1 Box Plot cluster diagram result

Based on the box plot distribution, the good nutrition category (blue) exhibits a higher median and a consistently broader distribution across multiple Posyandu, suggesting a generally healthy nutritional status for most children. In contrast, categories indicating poor nutrition (red) and undernutrition (green) show comparatively lower numbers, yet still highlight several Posyandu with notably higher outliers. This pattern indicates significant nutritional variation across different locations, emphasizing the necessity for targeted health interventions to address localized nutritional deficiencies effectively.

Meanwhile, the categories for obese nutrition (yellow), risk of overnutrition (purple), and overnutrition (orange) display considerable variation among the Posyandu, with some showing extreme outliers. This suggests the existence of distinct groups of children who may face overnutrition issues, requiring tailored health intervention strategies different from those aimed at malnourished groups. Insights drawn from this diagram can provide valuable guidance for health workers and policymakers to design more focused and effective nutrition intervention programs.

Figure 2 presents a bar chart illustrating the nutritional status comparison between male and female children across the entire Posyandu dataset. The categories assessed include good nutrition, poor nutrition, undernutrition, overnutrition, obesity, and risk of overnutrition. This visualization effectively highlights the gender-based distribution of nutritional conditions within the studied population, revealing notable differences and similarities between male and female children.

From the data presented, it is clear that in the categories of good nutrition, poor nutrition, and undernutrition, males outnumber females. Specifically, the prevalence of good nutritional status is notably higher among males, suggesting relatively favorable conditions for boys within the assessed communities. However, this positive trend is offset by the fact that more males also fall into the poor nutrition and undernutrition categories. Such findings imply that while a greater proportion of boys maintain good nutritional levels, a significant number still experience substantial nutritional deficiencies.

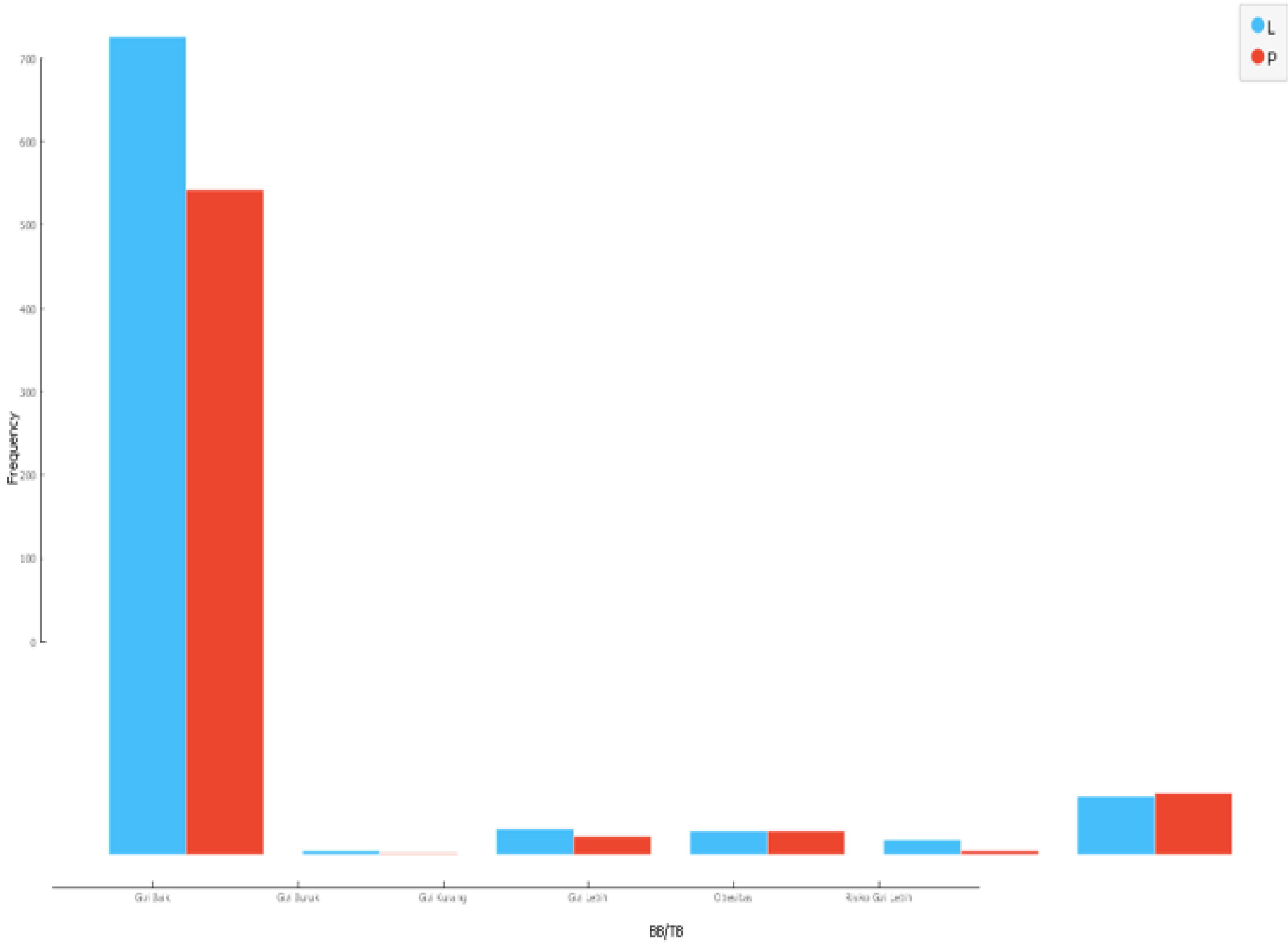

Figure 2 Comparison of male and female nutrition

Conversely, the category labeled "at risk of overnutrition" shows a distinctly higher representation of females than males. This suggests that girls are more susceptible to approaching a condition of excessive nutrition or being overweight compared to boys within the analyzed groups. On the other hand, the overnutrition category appears equally distributed between genders, indicating a balanced occurrence of this nutritional status among male and female children, highlighting a shared vulnerability or lifestyle factor influencing nutritional excess.

Finally, the obesity category demonstrates a higher frequency among males compared to females, suggesting that males are more prone to severe forms of overnutrition within the population analyzed. This result underscores the importance of targeted health interventions and nutritional guidance, particularly for boys, who appear at greater risk of obesity and its associated health complications. Overall, the gender-based differences identified through this bar chart emphasize the need for nuanced and gender-sensitive nutrition programs to adequately address the diverse nutritional needs of children in the studied communities.

Figure 3 presents a scatter plot that illustrates the nutritional status of children, differentiated by Posyandu location and gender. The diagram clearly highlights the distribution of various nutritional statuses across the community health centers, revealing distinct patterns related to gender and specific Posyandu locations.

All Posyandu locations are represented as having children with good nutritional status, indicating that overall nutrition levels across these communities are relatively favorable. However, concerning cases of malnutrition, specific instances are notable; female malnutrition occurs exclusively at Posyandu Flamboyan 2, while male malnutrition is specifically observed at Posyandu Melati 1. This demonstrates localized nutritional deficiencies, highlighting the necessity for targeted nutritional interventions in these specific health centers.

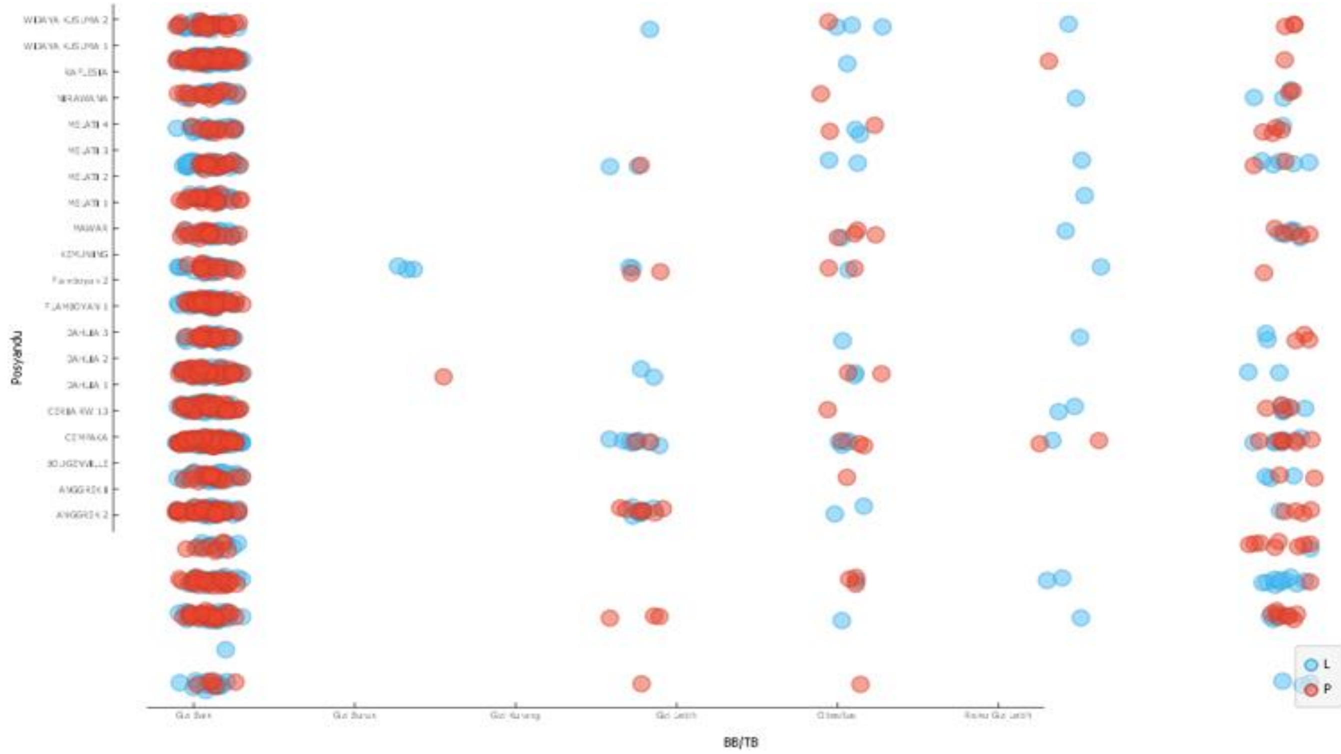

Figure 3 scatter plot of nutrition, JK, and posyandu

Undernutrition also exhibits a clear gender-based division among the Posyandu. Posyandu Wijaya Kusuma 2 and Flamboyan 2 exclusively record male children experiencing undernutrition, whereas Bougenville and Orchid 2 show only female cases. These gender-specific patterns in undernutrition suggest potential cultural or behavioral factors influencing dietary intake and health practices, warranting further investigation and specialized attention.

The analysis further reveals distinct trends for cases of overnutrition, obesity, and risk of overnutrition. Overnutrition appears equally distributed between male and female children, suggesting shared dietary or lifestyle factors across genders. Conversely, obesity is notably dominated by male children across all Posyandu, signaling that boys may face greater risks related to excessive nutritional intake or sedentary lifestyles. Additionally, the "risk of overnutrition" category predominantly features females and is observed consistently in nearly all Posyandu, underlining the urgency of preventive interventions specifically tailored for girls to mitigate progression toward obesity or related health conditions. See also Figure 4 for dendrogram cluster visualization.

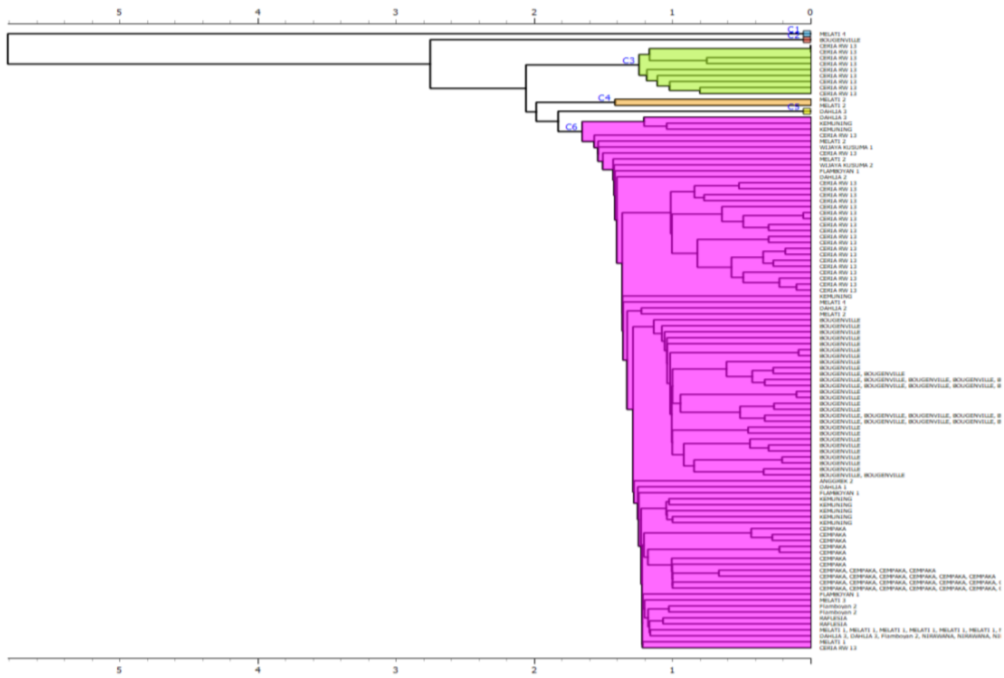

Figure 4 Dendrogram Single linkage

V. KESIMPULAN

Based on the clustering analysis conducted using the Single Linkage Agglomerative Hierarchical Clustering algorithm, this study successfully grouped infants' nutritional status by merging clusters based on similarities determined through Euclidean distances. This method effectively highlighted distinct nutritional groups within the observed data, enabling clear insights into variations across different Posyandu locations and between genders.

The results revealed that most infants were classified under the "Good Nutrition" category, indicating generally favorable nutritional conditions in the studied sub-district in Bekasi. However, the presence of specific cases such as malnutrition, undernutrition, overnutrition, obesity, and risk of overnutrition underscores the ongoing need for targeted health interventions and customized nutritional programs aimed at addressing these varied nutritional concerns within the community.

For future research, it is recommended to utilize more recent and extensive datasets to enhance the accuracy and representativeness of clustering results, as well as to collect data from multiple health centers (Posyandu or Puskesmas) to facilitate comparative analysis across different communities. Additionally, incorporating manual calculations complemented by programming languages such as Python or R is advisable to validate and streamline the clustering analysis process.

DAFTAR PUSTAKA

[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. Hutama 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 Jumput 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, “Pengelompokkan 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.