International Journal of Information Engineering and Science

E-ISSN: 3048-1953
P-ISSN: 3048-1902

Research Article

Grouping of Toddler Nutritional Status Based on Anthropometric Data in Pekan Kuala Village Using the K-Means Clustering Method

Dita Mawarni [1,*], Relita Buaton [2], Kristina Annatasia [3]

[1-3] Sekolah Tinggi Manajemen Informatika dan Komputer Kaputama, Indonesia
* Corresponding Author e-mail: mawarnid31@gmail.com

Abstract: Nutritional issues among toddlers continue to be a pressing public health challenge in Indonesia, including in Kelurahan Pekan Kuala, where, although anthropometric data have been systematically collected through the e-PPGBM application, they have not been thoroughly explored in terms of clustering patterns that may provide deeper insights. This study seeks to classify toddler nutritional status by applying the K-Means Clustering method to anthropometric indicators such as age, weight, height, and weight-to-height index. A dataset consisting of 648 entries recorded between January and March 2025 was processed using MATLAB R2014b with cluster variations set at 5, 7, and 9. The analysis revealed that the majority of toddlers were categorized as having good nutritional status, while a portion of the sample was identified as undernourished and some at risk of overnutrition, indicating the diverse nutritional challenges faced by this community. Furthermore, testing the variance across cluster configurations demonstrated that the 9-cluster model yielded the lowest variance score of 0.20, thereby representing the most optimal solution since it produced more homogeneous, balanced, and stable clusters compared to other configurations. These outcomes highlight the importance of data-driven approaches in public health planning, as the clustering results not only provide a clearer picture of nutritional distribution among toddlers but also serve as a foundation for more evidence-based and targeted intervention strategies. By offering a more granular understanding of nutritional variations, this research is expected to support local health authorities in developing customized nutrition programs, allocating resources more effectively, and ultimately improving child health outcomes in Kelurahan Pekan Kuala and similar communities across Indonesia, where malnutrition and overnutrition risks continue to coexist.

Keywords: Anthropometric Indicators; K-Means Clustering; Nutritional Status; Programs; Toddlers.

Received: July 04, 2025
Revised: July 20, 2025
Accepted: August 05, 2025
Published: August 20, 2025
Curr. Ver.: August 20,2025

1. Introduction

A crucial period in a child's growth and development is toddlerhood (children under five years old). Growth and development during this period determine a child's future success (World Health Organization, 2021). This golden age is a time of rapid growth and development that will never be repeated. Maintaining adequate nutrition for toddlers remains a challenge during this golden age. Malnutrition is highly likely to occur in toddlers who lack energy and nutrients (Dewi Marfuah et al., 2024).

Nutritional problems among toddlers in Indonesia remain high, with 18.4% experiencing malnutrition, 36.8% experiencing stunting, and 13.6% experiencing underweight (Nuzula et al., 2023). Therefore, monitoring the nutritional status of toddlers is a crucial step in supporting successful public health development. One method for assessing the nutritional status of toddlers is by using anthropometric measurements, specifically the weight-for-age index (BW/A), which is categorized into several levels: overnutrition, good nutrition, undernutrition, and poor nutrition (Novita Aryani & Henny Syapitri, 2021).

In Pekan Kuala Village, one effort to monitor the nutritional status of toddlers is the use of the e-PPGBM (Community-Based Electronic Nutrition Recording and Reporting) application, which records anthropometric data such as weight (BW), height (H), and age.

The data recorded in e-PPGBM is then used to determine the nutritional status of toddlers, whether they are undernourished, well-nourished, or at risk of overnutrition.

One way to analyze data more deeply is to use data grouping techniques, such as the K-Means Clustering method. The application of the K-Means Clustering method in Pekan Kuala Village is expected to provide a clearer picture of the nutritional status of toddlers in the area. This can help the integrated health post (Posyandu) in Pekan Kuala Village or community health centers (Puskesmas) plan more targeted interventions tailored to the needs of each toddler group.

Research conducted by (Siska Narulita et al., 2023) stated that the K-Means Clustering method can produce results that are expected to be used as a basis or recommendation for integrated health service posts (Posyandu) throughout Indonesia in determining the nutritional status of toddlers. (Ni Komang Sri Julyantari et al., 2021) In his writing, he explains that the K-Means Clustering method can produce useful information for health workers in efforts to improve the nutritional status of toddlers in the region. Research conducted by (Romadhoni et al., 2024) proves that the K-Means algorithm can be used as an effective analytical tool in identifying children's health patterns, thus helping in formulating more targeted health policies.

2. Theoretical Study

2.1 Toddler

Toddlers (0-59 months) are in a golden age of rapid physical, cognitive, and socio-emotional development, requiring high-quality nutrition to support optimal growth. However, this age group is vulnerable to malnutrition, which can hinder their growth and development potential (Kuliah, 2011).

2.2 Nutritional Status of Toddlers

Undernutrition is a condition in which a toddler experiences a weight deficit relative to their height. This is often associated with inadequate food intake or short-term health problems. Meanwhile, good nutrition is a condition in which a toddler has a weight proportional to their height, indicating adequate nutritional intake to support healthy growth and development (Fajar et al., 2022). As for the risk of overnutrition, this condition indicates a toddler's tendency to be overweight compared to their height. If not promptly addressed, this condition risks developing into overweight or obesity (Gizi, 2023).

2.3 Anthropometric Data

Anthropometric data generally includes several data points resulting from the anthropometric measurements themselves, namely the results of measurements of BB, TB, or PB and age measurements in general (Al-Rahmad & Fadillah, 2023). Toddler Anthropometry Standards must be used as a reference for health workers, program managers, and related stakeholders to assess the nutritional status of toddlers. The Body Weight Index according to Height (BB/TB) of children aged 0 (zero) to 60 (sixty) months is used to determine the categories: severe wasting, undernutrition (wasted), good nutrition (normal), at risk of overweight, overweight, and obesity (Permenkes RI No. 2, 2020).

Table 1. Categories of Toddler Nutritional Status

Body Weight according to Body Length or Height (BB/PB or BB/TB) for children aged 0 - 60 months	Malnutrition (severely *wasted*)	<-3 SD
	Malnutrition (*wasted*)	-3 SD sd <-2 SD
	Good nutrition (*normal*)	-2 SD sd +1 SD
	risk *of* overweight	> + 1 SD sd + 2 SD
	Overweight	> + 2 SD sd + 3 SD
	Obesity (*obese*)	> + 3 SD

2.4 Understanding Data Mining

According to (Dicha Mutia Dhani et al., 2024), data mining is the process of automatically extracting data from large or complex data sets to identify important patterns or trends that often go unnoticed. The goal is to discover hidden patterns or trends in the data that can provide valuable insights and support decision-making (Rahman et al., 2022).

2.5 Understanding KDD

Knowledge Discovery in Databases (KDD) is the entire complex process of searching for and identifying patterns in data. The patterns obtained must be valid, new, useful, and understandable (Sundari et al., 2021).

2.6 Understanding Clustering

Clustering is a data mining technique that groups objects based on the information available in the data, describing their characteristics and the relationships between them (Kumar et al., 2023). A cluster is a collection of data objects that are similar to each other but different from objects in other groups or clusters (Juliawati et al., 2023).

2.7 Understanding the K-Means Algorithm

According to (Ramadhayanti et al.., 2022), the K-Means algorithm is a non-hierarchical clustering method that aims to divide data into several clusters, so that data with similar characteristics are grouped in the same cluster, while data with different characteristics are put into other clusters.

The grouping process using K-Means Clustering can be carried out through several stages as follows (Juliawati et al., 2023) :

a. Determine the number of Clusters K

b. Determining the central value (centroid) is generally done randomly.

c. Calculate the distance of the object to the centroid based on the data in each cluster using Euclidean Distance with the following formula:

$$Dij = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + ... + (Xni - Xnj)^2}$$

Information :

D(i,j)= distance of data i to the center of Cluster j

Xni = i data in n data attribute

Xnj= jth center point of nth attribute

d. Move each data point to the cluster that has the closest centroid and update the cluster centers by calculating the average of the values in each cluster.

e. Return to step number 3 if there is still data that has moved clusters. If the cluster center has not changed, then the grouping process is complete, or until the cluster center has not changed again.

2.8 Understanding MATLAB

According to (Febrianti & Harahap, 2021), it is an abbreviation of Matrix Laboratory, which is software used for programming, analysis, and technical and mathematical calculations based on matrices.

3. Research Methods

Research methods are systematic steps used to obtain data in conducting research. In this thesis research, several stages of the method were applied.

These stages are as follows:

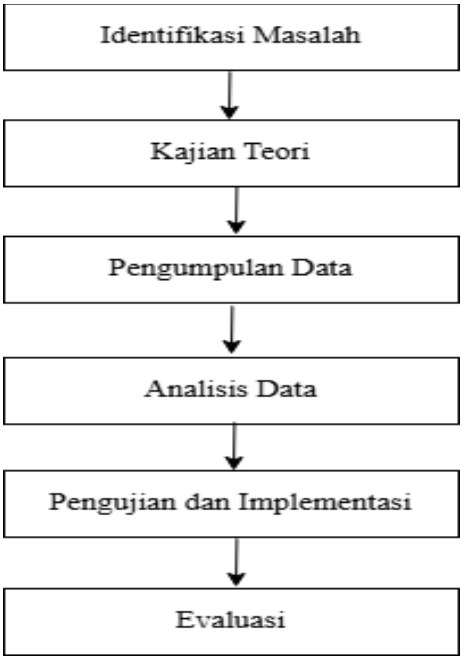

Figure 1. Research Method Structure

Based on the established method structure, the researcher explains the research stages as follows:

a. Identify the Problem
This stage is the initial stage carried out, where at this stage the problem identification is carried out for Pekan Kuala Village, which is the object of the research.

b. Theoretical Study
At this stage, the theoretical basis supporting the research problem is gathered. Theoretical sources are drawn from literature such as books, scientific journals, articles, and other academic references.

c. Data Collection
This stage is the process of collecting toddler data needed for the research. The data collected includes variables such as age, weight (BW), height (H), and the BW/H ratio as a reference for determining nutritional status. All data was obtained from the Kuala Community Health Center, which covers the Pekan Kuala sub-district.

d. Data Analysis
This stage is the process of processing and analyzing data that has been collected previously, so that the data can be grouped according to the variables that have been determined in this research.

e. Testing and implementation
This stage is the stage where validation testing and implementation of data that has been previously analyzed are carried out.

f. Evaluation
This is the final stage, which involves concluding. These conclusions provide a clear understanding of the overall research results, and it is hoped that the suggestions provided will lead to improvements and benefit others.

This study used data from 648 toddlers from the Kuala Community Health Center (Puskesmas) with variables of age, weight, height, and weight/height index. The data were analyzed using MATLAB R2014b and the K-Means Clustering method after transforming categorical variables to numerical values (Putra et al., 2024). The results are expected to help integrated health posts (Posyandu) and community health centers (Puskesmas) design targeted nutrition interventions.

4. Results and Discussion

4.1 Data Input

The input data for the system is toddler anthropometric data, consisting of age, weight (BW), height (H), and the BW/H ratio. This data is the result of measurements that have been transformed into numerical form according to the analysis requirements. The transformation was performed to equalize the scale between variables and facilitate the grouping process using the K-Means Clustering method.

1. Input Data
   Number of data : 648 data
   Variable : W: Toddler Age
   X: Toddler's Weight
   Y: Toddler's Height
   Z: BB/TB Ratio (Nutritional Status of Toddlers)
2. Cluster Grouping: 3 Clusters (5, 7, and 9)
3. Transformation Value

Table 2. Data Transformation Values

No	Variables	Transformation	Transformation Values
1	Toddler Age (Months)	6 - 12 months	1
		13 - 24 months	2
		25 - 36 months	3
		37 - 48 months	4
		49 - 59 months	5
2	Weight	5.8 - 12 kg	1
		12.1 - 15.3 kg	2
		15.4 - 18.3 kg	3
		18.4 - 21.2 kg	4
		21.3 - 24.2 kg	5
3	Height	61.2 - 81.5 cm	1
		81.6 - 93.9 cm	2
		94 - 103.5 cm	3
		103.6 - 111.7 cm	4
		111.8 - 119.5 cm	5
4	BB/TB (Nutritional status)	Malnutrition	1
		Good Nutrition	2
		Risk of Overnutrition	3

4.2 Clustering Results with Matlab

Data grouping using Matlab is done based on the proximity between variables using the Euclidean distance approach. The clustering process is carried out with varying numbers of clusters, namely 5, 7, and 9 clusters.

Grouping Results with 5 Clusters

```
Centroid Coordinates of Each Cluster:
Centroid 1: Age=4.17, BB=2.67, TB=3.11, BB/TB=2.00
Centroid 2: Age=3.17, BB=1.00, TB=2.19, BB/TB=1.29
Centroid 3: Age=1.85, BB=1.00, TB=1.26, BB/TB=2.00
Centroid 4: Age=5.00, BB=3.30, TB=4.08, BB/TB=2.00
Centroid 5: Age=2.91, BB=2.00, TB=2.37, BB/TB=2.00

-----------------------------------------
SUMMARY OF CLUSTERING RESULTS:
>> Cluster 1 contains 106 data
>> Cluster 2 contains 42 data
>> Cluster 3 contains 199 data
>> Cluster 4 contains 193 data
>> Cluster 5 contains 108 data

-----------------------------------------
```

Figure 2. Results of Grouping 5 Clusters

Based on the image above, it can be seen that:

- **a.** Cluster 1, with a total of 106 data, obtained the results of grouping toddlers aged "37 - 48 months" with BB "15.4 - 18.3 kg", TB "94 - 111.7 cm", and having TB/BB that is "Good Nutrition"
- **b.** Cluster 2, with a total of 42 data, obtained the results of grouping toddlers aged "25 - 36 months" with BB "5.8 - 12 kg", TB "81.6 - 93.9 cm", and having TB/BB namely "Undernourished"
- **c.** Cluster 3, with a total of 199 data points, obtained the results of grouping toddlers aged "6 - 12 months" with BB "5.8 - 12 kg", TB "61.2 - 81.5 cm", and having TB/BB of "Good Nutrition"
- **d.** Cluster 4, with a total of 193 data, obtained the results of grouping toddlers aged "49 - 59 months" with BB "15.4 - 18.3 kg", TB "103.6 - 111.7 cm", and having TB/BB that is "Good Nutrition"
- **e.** Cluster 5, with a total of 108 data, obtained the results of grouping toddlers aged "25 - 36 months" with BB "12.1 - 15.3 kg", TB "81.6 - 93.9 cm", and having TB/BB that is "Good Nutrition"

Grouping Results with 7 Clusters

```
Centroid Coordinates of Each Cluster:
Centroid 1: Age=5.00, BB=3.30, TB=4.08, BB/TB=2.00
Centroid 2: Age=1.78, BB=1.00, TB=1.00, BB/TB=2.00
Centroid 3: Age=3.01, BB=2.01, TB=2.40, BB/TB=2.00
Centroid 4: Age=4.17, BB=2.67, TB=3.11, BB/TB=2.00
Centroid 5: Age=2.00, BB=1.16, TB=2.02, BB/TB=1.98
Centroid 6: Age=3.73, BB=1.00, TB=2.73, BB/TB=1.00
Centroid 7: Age=3.00, BB=1.00, TB=1.94, BB/TB=1.44

-----------------------------------------
SUMMARY OF CLUSTERING RESULTS:
>> Cluster 1 contains 193 data
>> Cluster 2 contains 145 data
>> Cluster 3 contains 97 data
>> Cluster 4 contains 106 data
>> Cluster 5 contains 64 data
>> Cluster 6 contains 11 data
>> Cluster 7 contains 32 data

-----------------------------------------
```

Figure 3. Results of Grouping 7 Clusters

Based on the image above, it can be seen that:

- **a.** Cluster 1, with a total of 193 data, obtained the results of grouping toddlers aged "49 - 59 months" with BB "15.4 - 18.3 kg", TB 103.6 - 111.7 cm", and having TB/BB of "Good Nutrition"
- **b.** Cluster 2, with a total of 145 data, obtained the results of grouping toddlers aged "13 - 24 months" with BB "5.8 - 12 kg", TB "61.2 - 81.5 cm", and having TB/BB that is "Good Nutrition"
- **c.** Cluster 3, with a total of 97 data, obtained the results of grouping toddlers aged "25 - 36 months" with BB "12.1 - 15.3 kg", TB "81.6 - 93.9 cm", and having TB/BB of "Good Nutrition"
- **d.** Cluster 4, with a total of 106 data, obtained the results of grouping toddlers aged "37 - 48 months" with BB "15.4 - 18.3 kg", TB "94 - 103.5 cm", and having TB/BB that is "Good Nutrition"
- **e.** Cluster 5, with a total of 64 data, obtained the results of grouping toddlers aged "13 - 24 months" with BB "5.8 - 12 kg", TB "81.6 - 93.9 cm", and having TB/BB of "Good Nutrition"
- **f.** Cluster 6, with a total of 11 data, obtained the results of grouping toddlers aged "37 - 48 months" with BB "5.8 - 12 kg", TB "94 - 103.5 cm", and having TB/BB namely "Undernourished"
- **g.** Cluster 7, with a total of 32 data, obtained the results of grouping toddlers aged "25 - 36 months" with BB "5.8 - 12 kg", TB "81.6 - 93.9 cm", and having TB/BB namely "Undernourished"

Grouping Results with 9 Clusters

```
Centroid Coordinates of Each Cluster:
Centroid 1: Age=5.00, BB=2.99, TB=3.91, BB/TB=2.00
Centroid 2: Age=2.01, BB=1.00, TB=1.34, BB/TB=1.94
Centroid 3: Age=4.02, BB=2.61, TB=3.13, BB/TB=2.00
Centroid 4: Age=1.00, BB=1.00, TB=1.00, BB/TB=1.87
Centroid 5: Age=2.89, BB=1.85, TB=2.00, BB/TB=2.00
Centroid 6: Age=5.00, BB=4.00, TB=4.24, BB/TB=2.00
Centroid 7: Age=3.28, BB=1.00, TB=2.28, BB/TB=1.00
Centroid 8: Age=1.85, BB=1.00, TB=1.00, BB/TB=3.00
Centroid 9: Age=2.98, BB=2.00, TB=3.00, BB/TB=2.00
----------------------------------------
SUMMARY OF CLUSTERING RESULTS:
>> Cluster 1 contains 151 data
>> Cluster 2 contains 157 data
>> Cluster 3 contains 90 data
>> Cluster 4 contains 30 data
>> Cluster 5 contains 80 data
>> Cluster 6 contains 58 data
>> Cluster 7 contains 29 data
>> Cluster 8 contains 13 data
>> Cluster 9 contains 40 data
----------------------------------------
```

Figure 4. Results of Grouping 9 Clusters

Based on the image above, it can be seen that:

- **a.** Cluster 1, with a total of 151 data, obtained the results of grouping toddlers aged "49 - 59 months" with BB "15.4 - 18.3 kg", TB 103.6 - 111.7 cm" and having TB/BB that is "Good Nutrition"
- **b.** Cluster 2, with a total of 157 data, obtained the results of grouping toddlers aged "13 - 24 months" with BB "5.8 - 12 kg", TB 61.2 - 81.5 cm" and having TB/BB namely "Good Nutrition"
- **c.** Cluster 3, with a total of 90 data, obtained the results of grouping toddlers aged "37 - 48 months" with BB "15.4 - 18.3 kg", TB 94 - 103.5 cm" and having TB/BB, namely "Good Nutrition"

d. Cluster 4, with a total of 30 data, obtained the results of grouping toddlers aged "6 - 12 months" with BB "5.8 - 12 kg", TB 61.2 - 81.5 cm" and having TB/BB of "Good Nutrition"
e. Cluster 5, with a total of 80 data, obtained the results of grouping toddlers aged "25 - 36 months" with BB "12.1 - 15.3 kg", TB 81.6 - 93.9 cm" and having TB/BB of "Good Nutrition"
f. Cluster 6, with a total of 58 data, obtained the results of grouping toddlers aged "49 - 59 months" with BB "18.4 - 21.2 kg", TB 103.6 - 111.7 cm" and having TB/BB that is "Good Nutrition"
g. Cluster 7, with a total of 29 data, obtained the results of grouping toddlers aged "25 - 36 months" with BB "5.8 - 12 kg", TB 81.6 - 93.9 cm" and having TB/BB namely "Undernourished"
h. Cluster 8, with a total of 13 data, obtained the results of grouping toddlers aged "13 - 24 months" with BB "5.8 - 12 kg", TB 61.2 - 81.5 cm" and having TB/BB that is "Risk of Over Nutrition"
i. Cluster 9, with a total of 40 data, obtained the results of grouping toddlers aged "25 - 36 months" with BB "12.1 - 15.3 kg", TB 94 - 103.5 cm" and having TB/BB that is "Good Nutrition"

4.3 Test Results

Next, once the data has been grouped into their respective clusters, the results are tested. This cluster test is performed to ensure the quality of the clusters obtained and to ensure the representativeness of the existing data. Cluster test results are performed using cluster analysis. This cluster analysis measures the distribution of data points within each cluster (variance). Cluster analysis involves calculating Vminimum, Vmaximum, and cluster variance. To calculate the variance for each cluster, use the following formula :

$$Variance_k = \frac{1}{nk}\sum_{i=1}^{nk}(X_i - C_k)^2$$

Where :

nk : Number of data points in a cluster
xi : i-th data point in cluster
ck : Centroid of cluster

a. Vminimum (Minimum Variance)
Vminimum or Minimum Variance is the lowest (slightest) variation of all clusters formed in the data set. After the variance is calculated, the minimum variance value of all clusters can be taken.
Formula: Vminimum = min(Variance1, Variance2, …, Variancek)

b. Vmaximum (Maximum Variance)
Maximum Variance or Maximum Variance is the highest (most) variation of all clusters formed in the data set. After the variance is calculated, the maximum variance value of all clusters can be taken.
Formula: Vmaximum = max(Variance1, Variance2, …, Variancek)

c. Cluster Variance
Cluster Variance is the mean (average) of all clusters formed in a data set. Cluster variance serves to describe the extent of the spread of data points within a cluster.
Cluster variance formula :

$$ClusterVariance = \frac{1}{k}\sum_{i=1}^{k} Variance_k$$

d. Test Results for 5 Clusters, 7 Clusters, and 9 Clusters

Table 3. Test Results for 5 Clusters, 7 Clusters, and 9 Clusters

Cluster	Centroid	*Variance*	*Vmin*	*Vmax*	*Cluster Variance*
5	4.17; 2.67; 3.11; 2.00	0.46	0.15	1.27	0.43
	3.17; 1.00; 2.19; 1.29	0.54	0.15	1.49	

Cluster	Centroid	*Variance*	*Vmin*	*Vmax*	*Cluster Variance*
	1.85; 1.00; 1.26; 2.00	0.47	0.09	1.79	
	5.00; 3.30; 4.08; 2.00	0.29	0.10	0.70	
	2.91; 2.00; 2.37; 2.00	0.35	0.15	2.23	
7	5.00; 3.30; 4.08; 2.00	0.29	0.10	1.70	0.32
	1.78; 1.00; ; 1.00; 2.00	0.35	0.05	1.61	
	3.01; 2.01; 2.40; 2.00	0.26	0.16	1.34	
	4.17; 2.67; 3.11; 2.00	0.46	0.15	1.27	
	2.00; 1.16; 2.02; 1.98	0.16	0.03	0.99	
	3.73; 1.00; 2.73; 1.00	0.40	0.15	0.61	
	3.00; 1.00; 1.94; 1.44	0.30	0.20	1.20	
9	5.00; 2.99; 3.91; 2.00	0.12	0.01	1.19	0.20
	2.01; 1.00; 1.34; 1.94	0.30	0.12	1.32	
	4.02; 2.61; 3.13; 2.00	0.37	0.17	1.35	
	1.00; 1.00; 1.00; 1.87	0.12	0.02	0.76	
	2.89; 1.85; 2.00; 2.00	0.25	0.03	1.25	
	5.00; 4.00; 4.24; 2.00	0.18	0.06	0.58	
	3.28; 1.00; 2.28; 1.00	0.40	0.16	1.04	
	1.85; 1.00; 1.00; 3.00	0.13	0.02	0.72	
	2.98; 2.00; 3.00; 2.00	0.07	0.00	1.96	

Based on the table above, it can be seen that:

5 Cluster Testing

a. Cluster 1 has a variance value of 0.46, indicating that the data is fairly homogeneous and distributed close to the centroid. With a Vmin value of 0.15 and a Vmax of 1.27,

the data distribution range remains within stable limits and supports the quality of the clustering.

b. Cluster 2 has the highest variance value, at 0.54, indicating that the data in this cluster is relatively more varied than the other clusters. The Vmin value of 0.15 and Vmax of 1.49 confirm the wider distribution.

c. Cluster 3 has a variance value of 0.47, indicating a fairly even data distribution but still within acceptable limits. With Vmin values of 0.09 and Vmax of 1.79, this indicates some data diversity but remains within a reasonable range.

d. Cluster 4 has a variance value of 0.29, which is the lowest value in this test, indicating that the data is very dense and uniform. Vmin values of 0.10 and Vmax of 0.70 indicate the narrowest and most stable distribution.

e. Cluster 5 has a variance value of 0.35, which is also considered low. The Vmin value of 0.15 and Vmax of 2.23 indicate that most of the data is quite close to the cluster center, but there is still some variation in the data.

Cluster Variance of 0.43 indicates that the 5-cluster grouping model is quite good with a relatively homogeneous distribution of data between clusters.

7 Cluster Testing

a. Cluster 1 has a variance value of 0.29, which is one of the lowest values. The Vmin value of 0.10 and Vmax of 1.70 indicate a fairly stable distribution despite some data deviations.

b. Cluster 2 has a variance of 0.35, indicating moderate data distribution. A Vmin value of 0.05 and a Vmax of 1.61 indicate a higher range of variation than cluster 1.

c. Cluster 3 has a variance of 0.26, indicating a fairly dense data distribution. Vmin values of 0.16 and Vmax of 1.34 support a good level of homogeneity.

d. Cluster 4 has the highest variance, at 0.46, indicating that the data in this cluster is quite spread out. Vmin values of 0.15 and Vmax of 1.27 indicate higher heterogeneity.

e. Cluster 5 has a variance of 0.16, which is the lowest value in this test. Vmin values of 0.03 and Vmax of 0.99 indicate that the data is highly concentrated around the centroid.

f. Cluster 6 has a variance of 0.40, indicating moderate to high data dispersion. A Vmin value of 0.15 and a Vmax value of 0.61 indicate a relatively low dispersion, despite the relatively high variance.

g. Cluster 7 has a variance of 0.30, which is considered quite stable. A Vmin value of 0.20 and a Vmax of 1.20 indicate moderate data distribution.

Cluster Variance for the 7 cluster test was 0.32, lower than that of 5 clusters, indicating improved clustering quality.

9 Cluster Testing

a. Cluster 1 has a variance of 0.12, indicating that the data in this cluster is very dense. A Vmin of 0.01 and a Vmax of 1.19 indicate that most of the data is close to the centroid.

b. Cluster 2 has a variance of 0.30, which is considered moderate. Vmin of 0.12 and Vmax of 1.32 indicate a fairly wide distribution of data.

c. Cluster 3 has the highest variance, namely 0.37, indicating that the data is quite heterogeneous. Vmin of 0.17 and Vmax of 1.35 indicate a large spread.

d. Cluster 4 has a variance of 0.12, indicating that the data is quite homogeneous. Vmin of 0.02 and Vmax of 0.76 support this.

e. Cluster 5 has a variance of 0.25, which is still quite stable. Vmin of 0.03 and Vmax of 1.25 indicate moderate data distribution.

f. Cluster 6 has a variance of 0.18, indicating that the data is quite homogeneous. Vmin of 0.06 and Vmax of 0.58 indicate a narrow data distribution.

g. Cluster 7 has a variance of 0.40, which is considered high. Vmin of 0.16 and Vmax of 1.04 indicate significant data deviation.

h. Cluster 8 has a variance of 0.13, which is quite low. Vmin of 0.02 and Vmax of 0.72 indicate that the data is quite concentrated.

i. Cluster 9 has the lowest variance, at 0.07, indicating that the data in this cluster is highly homogeneous. Vmin of 0.00 and Vmax of 1.96 indicate that although one data item is quite far apart, the average distribution is very dense.

Cluster Variance for the 9 cluster test is 0.20, which is the lowest value compared to other tests, which indicates that the 9 cluster grouping is the most optimal, with the most homogeneous data distribution and the most stable data distribution.

5. Conclusion and Suggestions

Based on the results of research on the grouping of nutritional status of toddlers in Pekan Kuala Village using the K-Means Clustering method, it can be concluded:

The nutritional status of 648 toddlers in Pekan Kuala Village was successfully classified using the K-Means algorithm using weight, height, age, and weight/height variables. Experiments were conducted with 5, 7, and 9 clusters using MATLAB R2014b.

Grouping data with 5 clusters produces cluster 1: 106 data, which can be grouped based on toddlers aged "37 - 48 months" with BB "15.4 - 18.3 kg", TB "94 - 111.7 cm", and have TB/BB that is "Good Nutrition". Cluster 2: 42 data, can be grouped based on toddlers aged "25 - 36 months" with BB "5.8 - 12 kg", TB "81.6 - 93.9 cm", and have TB/BB that is "Undernourished". Cluster 3: 199 data, can be grouped based on toddlers aged "6 - 12 months" with BB "5.8 - 12 kg", TB "61.2 - 81.5 cm", and have TB/BB that is "Good Nutrition". Cluster 4: 193 data, can be grouped based on toddlers aged "49 - 59 months" with BB "15.4 - 18.3 kg", TB "103.6 - 111.7 cm", and have TB/BB that is "Good Nutrition". Cluster 5 108 data can be grouped based on toddlers aged "25 - 36 months" with BB "12.1 - 15.3 kg", TB "81.6 - 93.9 cm", and have TB/BB that is "Good Nutrition".

Grouping data with 7 clusters produces Cluster 1: 193 data, which can be grouped based on toddlers aged 49–59 months with BB 15.4 -- 18.3 kg, TB 103.6 - 111.7 cm, and have BB/TB that is Good Nutrition. Cluster 2: 145 data, can be grouped based on toddlers aged 13–24 months with BB 5.8 - 12 kg, TB 61.2 - 81.5 cm, and have BB/TB that is Good Nutrition. Cluster 3: 97 data, can be grouped based on toddlers aged 25 - 36 months with BB 12.1 - 15.3 kg, TB 81.6 - 93.9 cm, and have BB/TB that is Good Nutrition. Cluster 4: 106 data, can be grouped based on toddlers aged 37-48 months with a body weight of 15.4-18.3 kg, height of 94-103.5 cm, and have a body weight/height of Good Nutrition. Cluster 5: 64 data, can be grouped based on toddlers aged 13-24 months with a body weight of 5.8-12 kg, height of 81.6-93.9 cm, and have a body weight/height of Good Nutrition. Cluster 6: 11 data, can be grouped based on toddlers aged 37-48 months with a body weight of 5.8-12 kg, height of 94-103.5 cm, and have a body weight/height of Undernutrition. Cluster 7: 32 data, can be grouped based on toddlers aged 25-36 months with a body weight of 5.8-12 kg, height of 81.6-93.9 cm, and have a body weight/height of Undernutrition.

Grouping data with 9 clusters produces Cluster 1: 151 data, which can be grouped based on toddlers aged 49 - 59 months with BB 15.4 - 18.3 kg, TB 103.6 - 111.7 cm, and have BB/TB that is Good Nutrition. Cluster 2: 157 data, can be grouped based on toddlers aged 13 - 24 months with BB 5.8 - 12 kg, TB 61.2 - 81.5 cm, and have BB/TB that is Good Nutrition. Cluster 3: 90 data, can be grouped based on toddlers aged 37 - 48 months with BB of 15.4 - 18.3 kg, TB 94 - 103.5 cm, and have BB/TB that is Good Nutrition. Cluster 4: 30 data, can be grouped based on toddlers aged 6 - 12 months with BB 5.8 - 12 kg, TB 61.2 - 81.5 cm, and have BB/TB that is Good Nutrition. Cluster 5: 80 data, can be grouped based on toddlers aged 25 - 36 months with BB 12.1 - 15.3 kg, TB 81.6 - 93.9 cm, and have BB/TB that is Good Nutrition. Cluster 6: 58 data, can be grouped based on toddlers aged 49 - 59 months with BB 18.4 - 21.2 kg, TB 103.6 - 111.7 cm, and have BB/TB that is Good Nutrition. Cluster 7: 29 data, can be grouped based on toddlers aged 25 - 36 months with BB 5.8 - 12 kg, TB 81.6 - 93.9 cm, and have BB/TB that is Undernourished. Cluster 8: 13 data, can be grouped based on toddlers aged 13 - 24 months with a body weight of 5.81 - 2 kg, height of 61.2 - 81.5 cm, and have a body weight/height that is at risk of overnutrition. Cluster 9: 40 data, can be grouped based on toddlers aged 25 - 36 months with a body weight of 12.1 - 15.3 kg, height of 94 - 103.5 cm, and have a body weight/height that is well nourished.

Based on the 5, 7, and 9 clusters, the majority of toddlers aged 13–59 months were well-nourished. Undernutrition was predominant among toddlers aged 25–36 and 37–48 months, while the risk of overnutrition among toddlers aged 13–24 months appeared in nine clusters. These findings emphasize the importance of routine monitoring and targeted nutritional interventions.

The test shows 9 most optimal clusters with the lowest variance value (0.20), indicating the most homogeneous data distribution and the most stable distribution compared to 5 and 7 clusters.

Further research is suggested to use clustering and evaluation methods other than K-Means and cluster variance, as well as adding variables, so that the clustering results are more optimal and can be used as a comparison for this research.

References

Al-Rahmad, A. H., & Fadillah, I. (2023). *Assessment of nutritional status and growth of toddlers: New anthropometry standards WHO-2006 multicentre growth reference study (MGRS)* (pp. 1–37). Department. https://gizipoltekkesaceh.ac.id/wp-content/uploads/2023/03/Modul_-Penilaian-Perkembangan-BALITA.pdf

Aryani, N., & Syapitri, H. (2021). The relationship between feeding patterns and nutritional status of toddlers in the Percut Chart. *Priority Nursing Journal, 4*(1), 135–145. https://doi.org/10.34012/jukep.v4i1.1402

Dhani, D. M., Buaton, R., & Prahmana, I. G. (2024). Application of the K-means clustering method to determine student satisfaction with campus facilities and infrastructure at STMIK Kaputama Binjai. *Bridge: Journal of Information Systems and Telecommunications Publication, 2*(3), 229–243. https://doi.org/10.62951/bridge.v2i3.170

Fajar, S. A., Anggraini, C. D., & Husnul, N. (2022). Effectiveness of providing additional food on the nutritional status of toddlers at the Citeras Health Center, Garut Regency. *Nutrition Scientific Journal, 1*(1), 30–40. https://doi.org/10.37058/nsj.v1i1.5975

Febrianti, T., & Harahap, E. (2021). Use of MATLAB applications in learning linear programming. *Jurnal Matematika, 20*(1), 1–7.

Gizi, S., Purba, A., Siregar, R. N., Simanjuntak, N., & Simbolon, H. (2023). Factors related to the nutritional status of toddlers (6–24 months) at the Buhit Samosir Community Health Center. *Indonesian Health Scientific Journal, 9*(1), 266–273.

Juliawati, F., Buaton, R., Saragih, R., & Kaputama, S. (2023). Data mining grouping of non-cash food assistance (BPNT) recipients using the clustering method (Case study: Payabakung Hamparan Perak Village Office). *Journal of Computer Science and Information Technology, 3*(2), 69. https://doi.org/10.47065/explorer.v3i2.793

Julyantari, N. K. S., Budiarta, I. K., & Putri, N. M. D. K. (2021). Implementation of K-means for grouping nutritional status of toddlers (Case study of Banjar Titih). *Jurnal Janitra Informatika dan Sistem Informasi, 1*(2), 92–101. https://doi.org/10.25008/janitra.v1i2.134

Kuliah, M. (2011). *Assignment letter*. STTKAO. https://sttkao.ac.id/storage/penelitian/5g_penelitian_bersama_finall_210722080700.pdf

Kumar, V., Singh, A., & Sharma, S. (2023). Clustering-based approach for nutritional status identification among under-five children using health survey data. Healthcare Analytics, 3, 100160. https://doi.org/10.1016/j.health.2023.100160

Marfuah, D., Sarifah, S., Khotimah, S. K., & Hatifah, D. K. (2024). Anthropometric measurements and determination of nutritional status of toddlers at the Posyandu Balita Bina Sejahtera Kadipiro Banjarsari Surakarta. *ALKHIDMAH: Journal of Community Service and Partnership, 2*(3), 138–149. https://doi.org/10.59246/alkhidmah.v2i3.983

Ministry of Health of the Republic of Indonesia. (2020). *Children's anthropometry: Regulation of the Minister of Health of the Republic of Indonesia Number 2 of 2020* (pp. 7, 16).

Narulita, S., Prihati, P., Oktaga, A. T., & Widyantoro, A. E. (2023). Performance of the K-means clustering algorithm for determining malnutrition status in toddlers. *Journal of Information, Science and Technology, 6*(1), 188–202. https://doi.org/10.55606/isaintek.v6i02.128

Nuzula, R. F., Arfan, N. A., & Ningrum, S. (2023). The role of cadres in efforts to improve the nutritional status of toddlers at integrated health posts. *Jurnal Kesehatan Samodra Ilmu, 14*(1), 18–21. https://doi.org/10.55426/jksi.v14i01.246

Putra, R. D., Santoso, H., & Wijaya, D. (2024). Data mining approach for early detection of stunting in toddlers using K-means clustering. International Journal of Advanced Computer Science and Applications, 15(2), 325–331.

Ramadhayanti, L. T., Buaton, R., & Maulita, Y. (2022). Application of data mining to group Covid-19 vaccination data using the clustering method. *BIMASATI (Bulletin of Multi-Disciplinary Science and Applied Technology), 1*(3), 111–114.

Rahman, M. M., Islam, M. R., & Hasan, M. M. (2022). Application of machine learning algorithms for predicting child malnutrition using anthropometric indicators. BMC Public Health, 22(1), 1543. https://doi.org/10.1186/s12889-022-13945-7

Romadhoni, A. R., Romadhoni, A., Taqiyyudin, D., Abid, A., & Bangsa, U. P. (2024). Application of the K-means algorithm in infants and toddlers in Semarang City: Insights and problem-solving plans. *[Journal name not provided], 3*(1), 32–41.*

Sundari, S., Buaton, R., & Saragih, R. (2021). Clustering of Trans Binjai bus user service satisfaction using the cluster data mining method: Case study of the Binjai City Transportation Agency. In *Proceedings of SENATIKA 2021* (Vol. 5, pp. 86–100).

World Health Organization. (2021). WHO child growth standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age. WHO Press.