Jurnal Informasi dan Teknologi

https://jidt.org/jidt

2024 Vol. 6 No. 3 Hal: 52-62 e-ISSN: 2714-9730

Determining Factors Affecting Mother's Decisions in Providing and Selecting Infant Formula Using Data Mining - Decision Support System Collaborative

Dwi Utari Iswavigra[1]✉, Bella Gusniar[2], Miranda Ika Vania[3]

[1,2,3]Universitas Sugeng Hartono

dwi.utari.iswavigra1997@gmail.com

Abstract

Indonesia is one of the countries with a population of around two hundred million people, ranking fourth after the United States in the list of the most populous countries in the world. According to data from the Central Statistics Agency (BPS), the birth rate is projected to reach 4.45 million people in 2022, an increase of 0.22% from the previous year, which was 4.45 million people. According to the 2007 Indonesian Demographic and Health Survey (IDHS), the rate of exclusive breastfeeding in Indonesia was only 32%. Factors such as maternal health, infant health, and formula milk promotions influence mothers' decisions to use formula milk. The availability of numerous formula brands complicates the decision-making process, with each brand offering different nutritional claims. This study employs the K-Medoids clustering algorithm to analyze factors affecting mothers' choices in formula feeding and the TOPSIS method to determine the most suitable formula for infants aged 0-6 months. The research involves clustering data from a questionnaire distributed to 100 mothers in the Solo Raya region into five categories: maternal health, maternal employment, formula promotion, infant health, and breastfeeding education. Results indicate that maternal health is the most influential factor, followed by infant health, maternal employment, and formula promotion. Lack of breastfeeding education does not significantly influence formula choice. The TOPSIS method, applied to evaluate 10 formula brands against six nutritional criteria, identifies Lactogen 1 as the best formula for infants aged 0-6 months with a highest value of 2.777104826. This data-driven approach provides a clear, systematic method for selecting an appropriate formula based on specific nutritional needs.

Keywords: *Infant Formula, Data Mining, Decision Support System, K-Medoids Clustering, TOPSIS.*

1. Introduction

Indonesia is one of the countries with a population of around two hundred million people, ranking fourth after the United States in the list of the most populous countries in the world. According to data from the Central Statistics Agency (BPS), the birth rate is projected to reach 4.45 million people in 2022, an increase of 0.22% from the previous year, which was 4.45 million people. According to the 2007 Indonesian Demographic and Health Survey (IDHS), the rate of exclusive breastfeeding in Indonesia was only 32%. Therefore, the Indonesian government has shown its commitment to supporting exclusive breastfeeding by enacting the Health Law No. 36 of 2009, Article 128, which emphasizes the right of babies to receive exclusive breastfeeding, except for medical reasons. Formula feeding for infants under 6 months is permitted under certain conditions, such as babies born weighing less than 2,500 grams, babies treated in incubators, babies with birth trauma, infections, and congenital abnormalities, or conditions in which the mother suffers from breast swelling, abscesses, malnutrition, HIV, etc. [1]. These specific conditions lead some mothers to opt for formula feeding. This choice is influenced by factors such as the mother's educational background, income, employment status, media influence, and healthcare access [2][3]. However, the variety of formula brands available makes it difficult for mothers to determine which formula provides the appropriate nutrition for their baby's needs.

Infant formula serves as a solution for mothers who are unable to provide breast milk; however, the wide range of infant formula brands available on the market makes the decision of choosing the right formula more complex. Each brand of infant formula makes different nutritional claims, ranging from iron and protein content to prebiotics, which are claimed to support infant development [4]. Data mining is a branch of computer science that involves the collection of large amounts of data to discover new patterns from these large datasets. One well-known technique in data mining is clustering, which involves grouping a set of data into clusters. One algorithm used in data mining is K-Medoids. The K-Medoids algorithm is a clustering method that uses objects as representatives (medoids) for the cluster center in each group. This study also uses the TOPSIS method. TOPSIS

Receipt: 15-09-2024 | Revision: 25-09-2024 | Publish: 01-10-2024 | doi: 10.60083/jidt.v6i3.593

is a multi-criteria decision-making method that evaluates each parameter based on the values of the positive ideal solution and negative ideal solution, as well as the distance between each alternative's values and the positive and negative ideal solution matrices [5]. The K-Medoids algorithm will help identify patterns in the factors influencing mothers' decisions, such as maternal health conditions, infant health, and formula milk promotions [6]. By clustering these factors into five influence clusters, the clustering results can provide clearer insights into how each factor contributes to mothers' decisions [7]. The TOPSIS process is then used to compare 10 brands of infant formula based on six nutritional criteria, such as iron, protein, and omega-6 & omega-3. This method provides a systematic and data-driven solution, allowing mothers to choose the formula that best meets their baby's specific needs [8].

The research utilizes the K-Medoids algorithm to identify factors that influence mothers in choosing formula milk for infants aged 0-6 months, and the TOPSIS method to select the most suitable formula based on nutritional needs. The study considers five variables: the mother's health, employment, formula promotion, baby's health, and lack of breastfeeding education. These variables are grouped into five clusters: not at all influential, not influential, somewhat influential, influential, and very influential using K-Medoids clustering. The TOPSIS method is then used to decide on the best formula from 10 alternatives based on six criteria: iron, protein, carbohydrates, omega 6 & 3, vitamin D & calcium, and prebiotics.

2. Research Methods

2.1 K-Medoids Clustering

The research study follows a structured framework with key steps: identifying the research problem, reviewing relevant literature, collecting data, conducting pre-processing, analyzing the data using K-Medoids Clustering, implementing findings through Rapid Miner, and thoroughly documenting all research activities [9]. The stages of the research workflow are depicted in Figure 1.

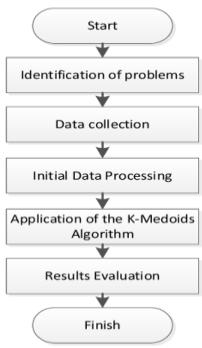

Figure 1. K-Medoids Method

Clustering, a method in Data Mining, involves grouping data or objects into clusters where the data within each cluster are similar. The key aspect of clustering is organizing patterns into appropriate groups to highlight similarities and differences, enabling valuable conclusions [10]. Clustering has several methods, one of which is K-Medoids. The K-Medoids algorithm or PAM (Partioning Around Medoids) is a data mining method represented by clusters, namely medoids [11].The K-Medoids method is designed to address the weaknesses of the K-Means method, which is sensitive to outlier data. The main goal of the K-Medoids method is to limit the proximity between cluster center points and the data points within the cluster [12]. The K-Medoids clustering algorithm is a variant of K-Means clustering that offers several advantages over K-Means. K-Medoids is more resistant to noise and outliers in the dataset, making it more suitable for datasets with high variability.[13].

The K-Medoids method involves several calculation stages that need to be completed, including [14][15]:

1. Determine the number of K for the cluster centers.
2. Assign each data point to the nearest cluster by measuring the Euclidean Distance using the formula in equation.
   $$JED = \sqrt{\sum n}\ (x\mathrm{i} - y\mathrm{i})2\ ;\ 1,2,3,\ \ldots\ \ldots\ ..\ n \quad (1)$$
3. Randomly select a data record from each cluster to become the new medoid.
4. Calculate the distance of objects within each cluster to the new potential medoid.
5. Calculate the total deviation by comparing the new total distance to the original total distance. If S < 0, find new data to create new K objects as medoids.
6. Repeat step 3 until the medoids no longer change, resulting in the final members of each cluster.

2.2 TOPSIS Method

After the K-Medoids clustering process, the TOPSIS method is applied to determine the infant formula that meets the needs of babies aged 0-6 months.

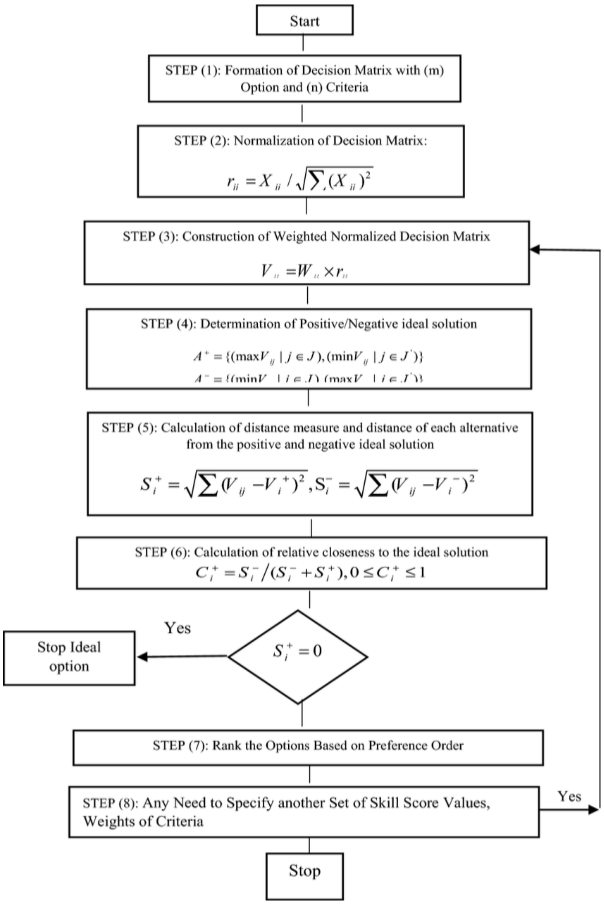

Figure 2. TOPSIS Method

The Decision Support System (DSS) is a computer-based method designed to assist in complex decision-making by providing relevant data, analyses, and models. Widely applied in fields like business, finance, and healthcare, DSS helps decision-makers develop strategies, plan actions, and manage resources more efficiently. [16]. TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) is a multi-criteria decision-making method used to select the best option among several alternatives based on multiple criteria. The method identifies the alternative that is closest to the ideal positive solution and farthest from the ideal negative solution, helping decision-makers evaluate and choose the optimal option [17][18]. The TOPSIS method starts by creating a decision matrix that outlines the values of each alternative based on various criteria. The matrix is then normalized to make the criteria comparable. Weights are assigned to each criterion according to their importance, and a weighted decision matrix is constructed. After determining the ideal positive and negative solutions, the method calculates the distance of each alternative from these solutions using Euclidean distance. The relative closeness of each alternative to the ideal solution is then measured, and the alternative with the highest closeness is selected as the optimal choice [19]. TOPSIS is a practical and straightforward method for multi-criteria decision-making, offering ease of implementation and interpretation, making it accessible even to those with limited technical expertise. It effectively handles trade-offs between criteria, helping decision-makers make balanced decisions. However, TOPSIS has limitations, including sensitivity to weight assignments and normalization, which can influence results. Additionally, determining ideal solutions for comparison may introduce subjectivity, potentially affecting the objectivity of the decision-making process [20]. This study utilizes 100 data points on factors influencing the choice of formula milk for infants aged 0-6 months. The data is categorized into five variables: maternal health condition, maternal employment, formula promotion factors, infant health condition, and lack of breastfeeding education. These variables are analyzed using the K-Medoids Clustering method and combined with the TOPSIS method to determine the best formula milk choice for mothers.

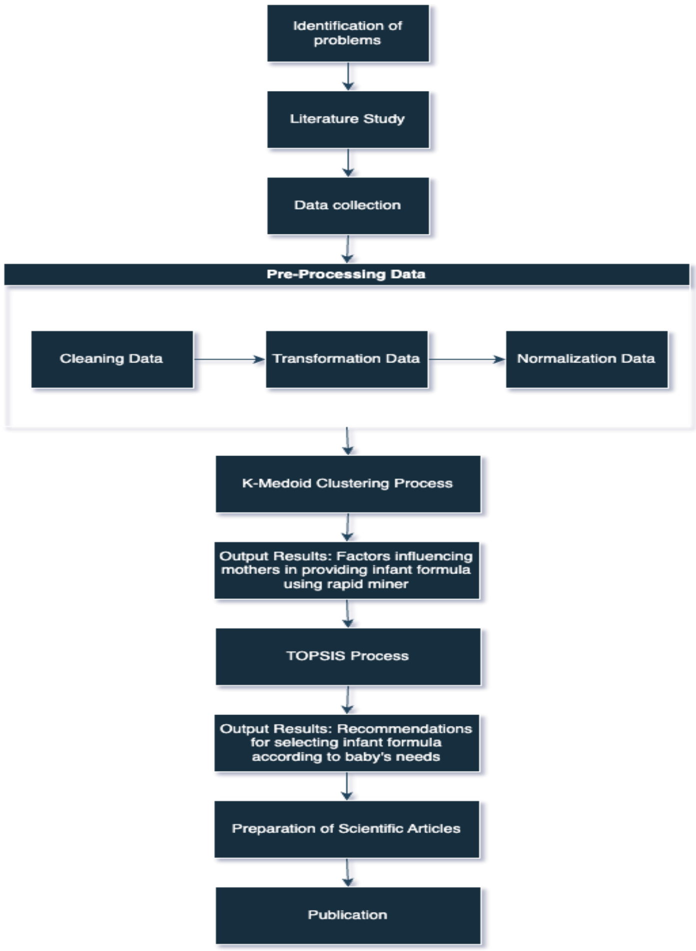

Figure 3. Research Workflow

3. Results and Discussion

3.1 K-Medoids Clustering

In the subsequent stage, the process involves several steps starting from data analysis to the completion of the method. The three key tasks to be carried out are as follows: Data Analysis, Data Preprocessing, and K-Medoids Clustering Calculation.

3.1.1 Data Analyze

Data analysis involves the systematic examination and interpretation of data to uncover valuable insights that support decision-making. It includes techniques for cleaning, organizing, and modeling data to identify patterns, trends, and relationships. The main goal of data analysis is to enhance understanding of the data, facilitating informed decisions and guiding strategic actions based on the findings [21]. The data used were obtained from the results of distributing questionnaires to 100 respondents in the Solo Raya area.

3.1.2 Preprocessing Data

Data preprocessing involves preparing and refining raw data before analysis. This includes tasks such as cleaning, transforming, and normalizing data to ensure accuracy and completeness. The goal is to improve data quality, making it more suitable for analysis and modeling, which enhances the reliability and effectiveness of the results [22].

1. Data Cleaning
   Data cleaning is the process of identifying and correcting errors or inconsistencies in a dataset to ensure its accuracy and reliability. This includes removing duplicates, correcting errors, addressing missing data, and standardizing formats. The goal of data cleaning is to produce a clean and dependable dataset that supports accurate analysis and informed decision-making [23]

2. Transformation Data
   Data transformation involves changing data from its original format or structure into a format that better supports analysis or processing needs. This includes tasks such as aggregating, normalizing, or encoding data

to align with analytical models or systems. The objective is to enhance the data's usability and compatibility, making it more effective for generating insights and conducting analysis [24]. This study identifies five categories for determining the factors influencing mothers in providing formula milk: very uninfluential, uninfluential, moderately influential, influential, and very influential. These categories are converted into a scale of 1-5 to facilitate questionnaire completion by respondents and simplify data processing for researchers.

3. Data Normalization
   Data normalization involves adjusting data to a uniform scale or format to maintain consistency and comparability throughout a dataset. This process may include scaling values to a particular range or standardizing data formats to remove variations. The purpose of data normalization is to improve the precision of analyses by ensuring that various variables are measured on a comparable scale, which allows for more accurate comparisons and interpretations [25].

Table 2. Data Normalization

No	Factor	Quation	Very Uninfluential	Uninfluential	Moderately Influential	Influential	Very Influential
1	Mother's Health Condition	Q1	0,421052632	0,083333333	0,238095238	0,55	0,833333333
2		Q2	0,210526316	0,125	0,142857143	0,55	0,972222222
3		Q3	0,315789474	0	0,19047619	0,5	1
...	....	....	....	....	....	....	....
5	Lack of Breastfeeding Education	Q25	0,578947368	0,458333333	0,619047619	0,75	0,166666667

3.1.3 K-Medoids Clustering

K-Medoids clustering is a method for partitioning a dataset into distinct clusters based on data point similarity. Unlike K-Means, which uses averages to define cluster centers, K-Medoids employs actual data points known as medoids. This approach reduces the impact of outliers and can be more robust in certain situations. The primary goal is to group similar data points, with each cluster represented by a real data point from the dataset [26].

1. Choosing the value of k, the number of clusters
   The results of the previous data normalization were divided into 5 clusters.
2. Randomly determine the initial medoid from n data.
   Select the first medoid from all clusters randomly from the normalized data presented in Table 4.

Table 3. First Medoid

Very Uninfluential	Uninfluential	Moderately Influential	Influential	Very Influential
0,421052632	0,083333333	0,238095238	0,55	0,833333333
0,210526316	0,125	0,142857143	0,55	0,972222222
0,315789474	0	0,19047619	0,5	1
0,526315789	0,583333333	0,333333333	0,6	0,361111111
0,368421053	0,291666667	0	0,4	0,944444444

3. Computes per-object proximity to a temporary medoid
   To calculate the proximity per object, the Euclidean Distance equation is used in Equation 1. The following is an example of the calculation on the first data.
   a. The first data goes to the center of cluster 1.
      $JED(X_1C_1) = \sqrt{(0{,}42 - 0{,}42)^2 + (0{,}083 - 0{,}083)^2 + (0{,}23 - 0{,}23)^2 + (0{,}55 - 0{,}55)^2 + (0{,}83 - 0{,}83)^2} = 0$

      Do a similar process for the 2nd data to the last data, so that the distance of each object to the temporary medoids will be produced as in Table 5.

Table 4. Distance of Object to Initial Medoid

C1	C2	C3	C4	C5	Euclidean Distance	Cluster
0	0,27279637	0,22487803	0,70402021	0,37109013	0	1
0,27279637	0	0,17956757	0,84972372	0,31046464	0	2
....	....	....	....	....	....	....
0,89169777	1,07820328	1,10450869	0,40042025	1,08754647	0,40042025	4
Total Proximity Distance Min					11,24	

4. Find the total proximity between clusters by marking the minimum distance of objects to Medoids.

Based on the minimum distance of the object to the initial/temporary medoids in table 4, the total proximity can be calculated by adding up all the minimum distances in table 5, so that the proximity value is 11,24.

5. Perform a new iteration of medoids by randomly re-selecting the new medoids presented in Table 6.

Table 5. Second Medoid

Very Uninfluential	Uninfluential	Moderately Influential	Influential	Very Influential
0,315789474	0,75	0,095238095	0,3	0,666666667
0,789473684	0,833333333	0	0,95	0,055555556
0,263157895	0,541666667	0,571428571	0,25	0,583333333
0,052631579	0,75	0,19047619	0,5	0,638888889
0,315789474	0,083333333	0,428571429	0,4	0,861111111

Repeat steps 3.3.3. to 3.3.4., so that the total closeness is 11,08

6. Find the total deviation value (S)

Next, calculate the total deviation (S) by calculating the difference between the number of new medoid proximity and the number of old medoid proximity. If S>0, then the process is stopped. If S<0, find a new medoid until you find a total of S>0.

S = 11,08 - 11,24 = -0,15

From the S value above, a negative result is obtained which is marked with an S value <0. Therefore, determine new medoids randomly and carry out the process and steps 3.3.3. to 3.3.4. The iteration continues until it reaches the 4th iteration, where in the 4th iteration, the S value > 0 with an S = 1.59, so the iteration is stopped.

S = 10,75 – 9,15 = 1,59

Table 6. Cluster Value

No	Factor	Quation	Very Uninfluential	Uninfluential	Moderately Influential	Influential	Very Influential	Cluster
1		Q1	0,421052632	0,083333333	0,238095238	0,55	0,833333333	2
2	Mother's Health Condition	Q2	0,210526316	0,125	0,142857143	0,55	0,972222222	2
3		Q3	0,315789474	0	0,19047619	0,5	1	2
16	....	....	....	....	....	....	....	....
25	Lack of Breastfeeding Education	Q25	0,578947368	0,458333333	0,619047619	0,75	0,166666667	4

Based on the calculation results of the K-Medoids method, it is known that the results in cluster 1 are 2 data, cluster 2 is 10 data, cluster 3 is 6, cluster 4 is 3, cluster 5 is 4 data. Based on the cluster results above, it is known that the mother's health condition factor dominates as a factor that influences mothers in deciding to use formula milk with a total of 4 categories (Q1, Q2, Q3 and Q5) among the 10 categories in cluster 2. The 2nd position in cluster 2 is influenced by the baby's health factor with a total of 3 questions fulfilled (Q1, Q2 and Q3). The 3rd position in cluster 2 is influenced by the mother's job factor with a total of 2 questions fulfilled (Q2 and Q3). The 4th position in cluster 2 is influenced by the promotion factor of formula milk with a total of 1 question fulfilled (Q3. Meanwhile, factor 5 "lack of breastfeeding education" is not included in cluster 2, where this category has no influence on the process of mothers deciding to give formula milk.

3.2 RapidMiner Studio Application Implementation

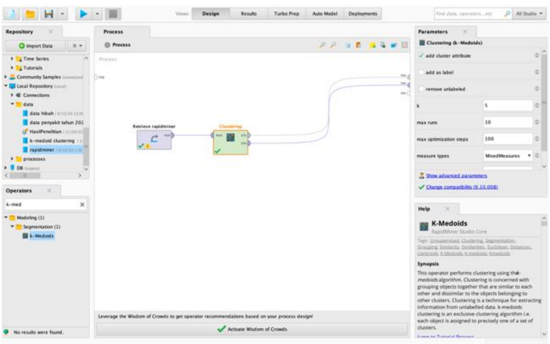

Figure 4. Data Connector and K- Medoids Clustering

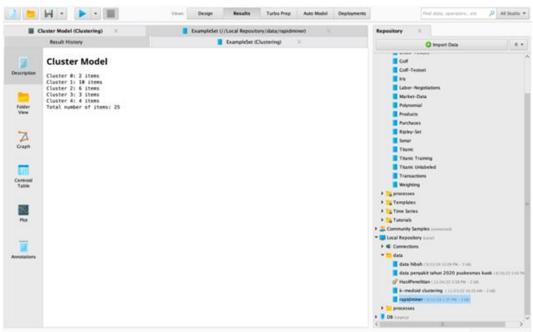

Figure 5. RapidMiner Text Cluster Values

3.3 TOPSIS Method

The following are the steps of the TOPSIS method:

A. Data Analysis

1. Criteria Data

There are 6 criteria used to assess expedition services, which are listed in Table 8.

Table 7. Criteria Data

Code	Criteria
C1	Iron
C2	Protein
C3	Carbohydrate
C4	Omega 6 & Omega 3
C5	Vitamin D and Calsium
C6	Prebiotik

2. Alternative Data

There are 10 brands of formula milk that are used as alternatives in research on formula milk selection based on predetermined criteria as shown in Table 9.

Table 8. Alternative Data

Code	Criteria
A1	Similac Advance
A2	S-26 Procal Gold
A3	Frisolac Gold 1
A4	Bebelac 1
A5	Morinaga BMT Platinum
A6	Lactogen 1
A7	Nutrilon Royal 1
A8	Nutricia Aptamil Profutura 1
A9	Similac Total Comfort
A10	SGM Eksplor Advance+ Soya

3. Criteria Weighting Data

The following are the weighting values of the criteria table and the criteria weights shown in Table 10 to Table 15.

Table 9. Weighting Criteria C1 (Iron)

Iron	Weight
< 1 mg / 100 kcal	1
1 - 1.2 mg / 100 kcal	2
1.2 - 1.5 mg / 100 kcal	3
> 1.5 mg / 100 kcal	4

Table 10. Weighting Criteria C2 (Protein)

Protein	Weight
No protein	1
Plant based protein	2
Cow's milk protein (whey or casein)	3
Combination of Whey protein and Casein	4

Table 11. Weighting Criteria C3 (Carbohydrate)

Carbohydrate	Weight
No carbohydrates	1
Carbohydrates other than lactose	2
Lactose as the main carbohydrate	3
Lactose and other sources of carbohydrates	4

Table 12. Weighting Criteria C4 (Omega 6 & Omega 3)

Omega 6 & Omega 3	Weight
No DHA or ARA	1
DHA only	2
ARA only	3
Combination of DHA and ARA	4

Table 13. Weighting Criteria C5 (Vitamin D and Calsium)

Vitamin D and Calsium	Weight
< 50 IU / 100 kcal, Kalsium <= 80 mg	1
50 - 100 IU / 100 kcal, Kalsium 50 - 120 mg	2
101 - 150 IU / 100 kcal, Kalsium 121 - 150 mg	3
> 70 IU / 100 kcal, Kalsium > 70 mg	4

Table 14. Weighting Criteria C6 (Prebiotik)

Prebiotik	Weight

No prebiotics	1
Contains GOS only	2
Contains FOS only	3
Combination of GOS and FOS	4
Combination of GOS and FOS and others	5

B. TOPSIS Method Calculation

In the TOPSIS method, the first process carried out is to determine the standard weight value. Table 16 shows the standard weight value used.

Table 15. Standard Weight Value

Information	**Value**
Not Important	1
Less Important	2
Quite Important	3
Important	4
Very Important	5

From what is shown in Table 16, it is known that the preference weights based on the criteria are as in Table 17.

Table 16. Criteria Preference Weight Standard

Criteria	C1	C2	C3	C4	C5	C6
Weight	4	5	4	5	5	4

The preference weight values for the criteria were derived from an interview with nutrition expert Mrs. Himatunnisa Mahmudah, a lecturer. After establishing the weight values and preference weights, the next step involves assessing the suitability of alternatives against the criteria. This assessment is based on observing the composition of each alternative.

Table 17. Suitability of Alternatives to Criteria

No		C1	C2	C3	C4	C5	C6
1	A1	4	4	3	4	2	2
2	A2	1	4	3	4	2	4
3	A3	1	4	3	4	1	4
...	...	...	...	...	...	...	...
10	A10	2	2	3	4	2	3

1. Determine the normalized decision matrix with the following formula:
   a. C1 (Iron)

$C_1 = \sqrt{4^2 + 1^2 + 1^2 + 2^2 + 3^2 + 2^2 + 2^2 + 2^2 + 4^2 + 2^2} = 7{,}93$

$r_{1.1} = 4/7{,}93 = 0{,}50$
$r_{2.1} = 1/7{,}93 = 0{,}12$
$r_{3.1} = 1/7{,}93 = 0{,}12$
$r_{4.1} = 2/7{,}93 = 0{,}25$
$r_{5.1} = 3/7{,}93 = 0{,}37$
$r_{6.1} = 2/7{,}93 = 0{,}25$
$r_{7.1} = 2/7{,}93 = 0{,}25$
$r_{8.1} = 2/7{,}93 = 0{,}25$
$r_{9.1} = 4/7{,}93 = 0{,}50$
$r_{10.1} = 2/7{,}93 = 0{,}25$

Perform the same calculations for criteria 3 to 6, resulting in the following R values.

Table 18. R values

R =

0,50395263	0,3549426	0,32539569	0,31622777	0,34299717	0,1833397
0,12598816	0,3549426	0,32539569	0,31622777	0,34299717	0,3666794
0,12598816	0,3549426	0,32539569	0,31622777	0,17149859	0,3666794
0,25197632	0,3549426	0,32539569	0,31622777	0,34299717	0,3666794
0,37796447	0,3549426	0,32539569	0,31622777	0,34299717	0,1833397
0,25197632	0,26620695	0,32539569	0,31622777	0,17149859	0,1833397
0,25197632	0,3549426	0,32539569	0,31622777	0,34299717	0,3666794
0,25197632	0,26620695	0,32539569	0,31622777	0,34299717	0,45834925
0,50395263	0,26620695	0,21693046	0,31622777	0,34299717	0,27500955
0,25197632	0,1774713	0,32539569	0,31622777	0,34299717	0,27500955

2. Determine the weighted normalized decision matrix based on the following formula:
   a. C1 (Iron)

$y_{1.1} = 4 * 0{,}50 = 2{,}01$
$y_{2.1} = 4 * 0{,}12 = 0{,}50$

$y_{3.1}$ = 4 * 0,12 = 0,50
$y_{4.1}$ = 4 * 0,25 = 1,007
$y_{5.1}$ = 4 * 0,37 = 1,511
$y_{6.1}$ = 4 * 0,25 = 1,007
$y_{7.1}$ = 4 * 0,25 = 1,007
$y_{8.1}$ = 4 * 0,25 = 1,007
$y_{9.1}$ = 4 * 0,50 = 2,01
$y_{10.1}$ = 4 * 0,25 = 1,007

Perform the same calculation for criteria 2 to 6, resulting in a weighted normalized decision matrix.

3. Calculate the distance between each alternative with the positive ideal solution and the negative ideal solution based on the formula:
   a. Before calculating the distance between the alternatives and the ideal positive solution and the ideal negative solution, we first determine the values of the ideal positive solution and the ideal negative solution.

Calculating the Ideal Positive Solution

y_1^+ = Max
{2,015810523;0,503952631;0,503952631;1,007905261;1,511857892;1,007905261;1,007905261;1,007905261;2,015810523;1,007905261} = 2,015810523
y_2^+ = 1,774713019
y_3^+ = 1,301582747
y_4^+ = 1,58113883
y_5^+ = 1,714985851
y_6^+ = 1,833396994

Calculating the Ideal Negative Solution

y_1^+ = Min
{2,015810523;0,503952631;0,503952631;1,007905261;1,511857892;1,007905261;1,007905261;1,007905261;2,015810523;1,007905261} = 0,503952631
y_2^+ = 0,887356509
y_3^+ = 0,867721831
y_4^+ = 1,58113883
y_5^+ = 0,857492926
y_6^+ = 0,733358798

   b. Calculating the distance to the ideal positive solution (D_i^+)

$D_1^+ = \sqrt{}$ (0,503952631 − 2,015810523)^2 + (0,887356509 − 1,774713019)^2 + (0,867721831 − 1,301582747)^2 + (1,58113883 − 1,58113883)^2 + (0,857492926 − 1,714985851)^2 + (0,733358798 − 0,733358798)^2 = 1,100038196

Perform the same calculation for D_2^+ to D_{10}^+, resulting in the distance of each alternative from the ideal positive solution as follows.

Table 19. Distance to the ideal positive solution

D1+	1,1000382
D2+	1,55568894
D3+	1,77636206
....	....
D10+	1,530062

   c. Calculating the distance to the ideal negative solution (D_i^-)

Table 20. Distance to the ideal negative solution

D1-	1,99916114
D2-	1,49958198
D3-	1,23022437
....	....
D10-	1,14540449

4. Continued by searching for preference values for each alternative (Vi) using the formula:

$vi = Di^- / Di^- + Di^+$;
V_1 = 1,99916114 / 1,99916114 + 1,1000382 = 2,1000382
V_2 = 1,49958198 / 1,49958198 + 1,55568894 = 2,55568894
V_3 = 1,23022437 / 1,23022437 + 1,77636206 = 2,77636206
V_4 = 1,58199696 / 1,58199696 + 1,07253289 = 2,07253289
V_5 = 1,65130373 / 1,65130373 + 1,20998028 = 2,20998028
V_6 = 0,79940849 / 0,79940849 + 1,77710483 = 2,77710483
V_7 = 1,58199696 / 1,58199696 + 1,07253289 = 2,07253289
V_8 = 1,6076169 / 1,6076169 + 1,10123722 = 2,10123722
V_9 = 1,83093216 / 1,83093216 + 0,96067727 = 1,96067727
V_{10} = 1,14540449 / 1,14540449 + 1,530062 = 2,530062

A larger Vi value indicates that the alternative Ai is more preferred. V6, represented by A6, was selected as the best choice for infant formula suitable for babies aged 0-6 months, with a value of 2.77710483. Below is the table showing preference values and the ranking of alternatives for selecting the infant formula used, as shown in Table 22.

Table 21. Preference Values and Ranking of Alternatives

Kode	Merk Sufor	Preferensi	Ranking
A1	Similac Advance	2,100038196	7
A2	S-26 Procal Gold	2,555688937	3
A3	Frisolac Gold 1	2,776362065	2
A4	Bebelac 1	2,072532889	8
A5	Morinaga BMT Platinum	2,209980284	5
A6	Lactogen 1	2,777104826	1
A7	Nutrilon Royal 1	2,072532889	8
A8	Nutricia Aptamil Profutura 1	2,101237218	6
A9	Similac Total Comfort	1,960677268	9
A10	SGM Eksplor Advance+ Soya	2,530061998	4

Based on Table 22, the preference calculations for selecting the best infant formula, analyzed using the TOPSIS method, show that the top 3 alternatives are Lactogen 1 with the highest value of 2.777104826,

7. Conclusion

The study analyzed factors influencing mothers in choosing formula milk by clustering data from a questionnaire distributed to 100 mothers in the Solo Raya region. It focused on five categories: mother's health, occupation, formula milk promotion, baby's health, and breastfeeding education. Using the K-Medoids clustering algorithm implemented in RapidMiner, results indicated that the mother's health was the dominant factor, influencing four out of ten categories in cluster 2. The baby's health ranked second, followed by the mother's occupation and formula milk promotion. Lack of breastfeeding education was not included in cluster 2, indicating it had no influence. Additionally, the TOPSIS method was used to identify the most suitable infant formula for babies aged 0-6 months, with Lactogen 1 being the top choice, achieving the highest score of 2.777104826.

References

[1] M. A. Biro, G. A. Sutherland, J. S. Yelland, P. Hardy, and S. J. Brown, “In-Hospital Formula Supplementation of Breastfed Babies: A Population-Based Survey,” Birth, vol. 38, no. 4, pp. 302–310, Sep. 2011, doi: 10.1111/j.1523-536x.2011.00485.x.

[2] Y. Huang, J. Labiner-Wolfe, H. Huang, C. J. Choiniere, and S. B. Fein, “Association of Health Profession and Direct-to-Consumer Marketing with Infant Formula Choice and Switching,” Birth, vol. 40, no. 1, pp. 24–31, Mar. 2013, doi: 10.1111/birt.12025.

[3] K. Parry, E. Taylor, P. Hall-Dardess, M. Walker, and M. Labbok, “Understanding Women’s Interpretations of Infant Formula Advertising,” Birth, vol. 40, no. 2, pp. 115–124, Jun. 2013, doi: 10.1111/birt.12044.

[4] “AOAC Official Method 2012.13Labeled Fatty Acids Content in Milk Products, Infant Formula, and Adult/Pediatric Nutritional Formula,” Official Methods of Analysis of AOAC INTERNATIONAL, Jan. 2023, doi: 10.1093/9780197610145.003.4081.

[5] B. Lönnerdal, “Infant formula and infant nutrition: bioactive proteins of human milk and implications for composition of infant formulas,” The American Journal of Clinical Nutrition, vol. 99, no. 3, pp. 712S-717S, Mar. 2014, doi: 10.3945/ajcn.113.071993.

[6] W. N. Hamilton, “Factors Influencing Maternal Decision-Making on Infant Feeding Practices,” Infant Feeding - Breast versus Formula, Oct. 2020, doi: 10.5772/intechopen.91325.

[7] C.-Y. Fu, X.-J. Tang, L.-P. Pan, H.-Y. Jin, J.-F. Yao, and L.-Z. Wang, “Exclusive breastfeeding rate and related factors among mothers within maternal health WeChat groups in Jiaxing, Zhejiang province, China: a cross-sectional survey,” International Breastfeeding Journal, vol. 17, no. 1, Nov. 2022, doi: 10.1186/s13006-022-00521-5.

[8] M. B. Regasa, T. Refera Soreta, O. E. Femi, and P. C. Ramamurthy, “Development of Molecularly Imprinted Conducting Polymer Composite Film-Based Electrochemical Sensor for Melamine Detection in Infant Formula,” ACS Omega, vol. 5, no. 8, pp. 4090–4099, Feb. 2020, doi: 10.1021/acsomega.9b03747.

[9] S. M. F. D. Syed Mustapha, “Predictive Analysis of Students’ Learning Performance Using Data Mining Techniques: A Comparative Study of Feature Selection Methods,” Applied System Innovation, vol. 6, no. 5, p. 86, Sep. 2023, doi: 10.3390/asi6050086.

[10] D. U. Iswavigra, L. Endriani Zen, Okfalisa, and H. Hanim, “Marketing Strategy UMKM Dengan CRISP-DM Clustering & Promotion Mix Menggunakan Metode K-Medoids”, jidt, vol. 5, no. 1, pp. 45-54, Feb. 2023.

[11] E. T. Ena Tasia and M. Afdal, “Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir,” Indonesian Journal of Informatic Research and Software Engineering (IJIRSE), vol. 3, no. 1, pp. 65–73, Mar. 2023, doi: 10.57152/ijirse.v3i1.523.

[12] A. Leis et al., “K-medoids clustering of hospital admission characteristics to classify severity of influenza virus infection,” Dec. 2022, doi: 10.22541/au.167056743.33133902/v1.

[13] B. Mrugalska, “Clustering of Cities based on their Smart Performances: A Comparative Approach of Fuzzy C-Means,

K-Means, and K-Medoids_supp2-3333753.docx", doi: 10.1109/access.2023.3333753/mm1.
[14] L. Zahrotun, U. Linarti, B. H. T. Suandi As, H. Kurnia, and L. Y. Sabila, "Comparison of K-Medoids Method and Analytical Hierarchy Clustering on Students' Data Grouping," JOIV : International Journal on Informatics Visualization, vol. 7, no. 2, p. 446, May 2023, doi: 10.30630/joiv.7.2.1204.
[15] H. Amin, "CLUSTERING ANALYSIS OF ADMISSION OF NEW STUDENTS USING K-MEANS CLUSTERING AND K-MEDOIDS ALGORITHMS TO INCREASE CAMPUS MARKETING POTENTIAL," Tech-E, vol. 7, no. 1, pp. 43–56, Aug. 2023, doi: 10.31253/te.v7i1.2264.
[16] G. Clark, A. Vistoso Monreal, N. Veas, and G. E. Loeb, "Perspective Chapter: Highly Structured Data Collection and Predictive Diagnostic Assist Algorithms in Orofacial Pain and Oral Medicine," Decision Support Systems (DSS) and Tools [Working Title], Dec. 2023, doi: 10.5772/intechopen.1003849.
[17] P. Trivedi, J. Shah, R. Čep, L. Abualigah, and K. Kalita, "A Hybrid Best-Worst Method (BWM)—Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) Approach for Prioritizing Road Safety Improvements," IEEE Access, vol. 12, pp. 30054–30065, 2024, doi: 10.1109/access.2024.3368395.
[18] M. Zaman, F. Ghani, A. Khan, S. Abdullah, and F. Khan, "Complex Fermatean fuzzy extended TOPSIS method and its applications in decision making," Heliyon, vol. 9, no. 9, p. e19170, Sep. 2023, doi: 10.1016/j.heliyon.2023.e19170.
[19] S. Chakraborty, "TOPSIS and Modified TOPSIS: A comparative analysis," Decision Analytics Journal, vol. 2, p. 100021, Mar. 2022, doi: 10.1016/j.dajour.2021.100021.
[20] Fuzzy AHP and Fuzzy TOPSIS as an effective and powerful Multi-Criteria Decision-Making (MCDM) method for subjective judgements in selection process," International Research Journal of Modernization in Engineering Technology and Science, May 2023, doi: 10.56726/irjmets36629.
[21] N. Beridze and G. Janelidze, "Organizing Decision-Making Support System Based Multi-Dimensional Analysis of the Educational Process Data," Journal of ICT, Design, Engineering and Technological Science, vol. 4, no. 1, Jun. 2020, doi: 10.33150/jitdets-4.1.1.
[22] A. P. Joshi and B. V. Patel, "Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process," Oriental journal of computer science and technology, vol. 13, no. 0203, pp. 78–81, Jan. 2021, doi: 10.13005/ojcst13.0203.03.
[23] G. Gledec, M. Horvat, M. Mikuc, and B. Blašković, "A Comprehensive Dataset of Spelling Errors and Users' Corrections in Croatian Language," Data, vol. 8, no. 5, p. 89, May 2023, doi: 10.3390/data8050089.
[24] H. Yu, P. Sang, and T. Huan, "Adaptive Box–Cox Transformation: A Highly Flexible Feature-Specific Data Transformation to Improve Metabolomic Data Normality for Better Statistical Analysis," Analytical Chemistry, vol. 94, no. 23, pp. 8267–8276, Jun. 2022, doi: 10.1021/acs.analchem.2c00503.
[25] Y. Yao et al., "Normalization Approach by a Reference Material to Improve LC–MS-Based Metabolomic Data Comparability of Multibatch Samples," Analytical Chemistry, Dec. 2022, doi: 10.1021/acs.analchem.2c04188.
[26] L. Teixeira, I. Eleutério, M. Cazzolato, M. Gutierrez, A. Traina, and C. Traina-Jr., "KluSIM: Speeding up K-Medoids Clustering over Dimensional Data with Metric Access Method," Proceedings of the 26th International Conference on Enterprise Information Systems, pp. 73–84, 2024, doi: 10.5220/0012599900003690.