Data-Driven Clustering of Stunting Prevention Services for Pregnant Women and Infants Using Fuzzy C-Means

Hanum Zalsabilah Idham✉
Universitas Negeri Makassar, Makassar, Indonesia
Ayu Safitri
Universitas Negeri Makassar, Makassar, Indonesia
Andi Akram Nur Risal
Universitas Gadjah Mada, Yogyakarta, Indonesia
Dewi Fatmarani Surianto
Universitas Negeri Makassar, Makassar, Indonesia
Firdaus
Universitas Negeri Makassar, Makassar, Indonesia

ABSTRACT

Purpose – This study addresses persistently high stunting rates in South Sulawesi, Indonesia, which remain above national targets despite declining trends. We developed a clustering model to overcome limitations of traditional methods in handling complex health data with overlapping characteristics, aiming to identify priority regions requiring targeted interventions.

Methods – Using 2,267 structured records from *Satu Data Indonesia* covering maternal and child health indicators, we implemented Fuzzy C-Means (FCM) algorithm with systematic preprocessing, optimal cluster determination via Elbow Method, and quality validation using Silhouette Coefficient.

Findings – Analysis revealed three distinct clusters for pregnant women (representing good, moderate, and low service coverage areas) and three corresponding clusters for infants. Validation showed Silhouette values ranging from 0.204 to 0.645, indicating variable cluster separation quality with Cluster 0 pregnant women achieving highest cohesion (0.638) and Cluster 2 infants showing strongest separation (0.645).

Research limitations – Data quality limitations affected cluster cohesion in some areas, particularly Cluster 1 infants (0.204 Silhouette value), constraining generalizability. The FCM approach accommodates real-world data complexity better than rigid clustering methods but requires high-quality input data.

Originality – This research contributes an adaptive framework for evidence-based stunting prevention through sophisticated data-driven segmentation. Findings offer immediate practical value for health policymakers in resource allocation and intervention planning, with potential adaptation to other regional contexts facing similar public health challenges.

Correspondence Author: ✉hanumzalzabilah@mail.com

To cite this article : Hanum Zalsabilah Idham, Ayu Safitri, Andi Akram Nur Risal, Dewi Fatmarani Surianto, & Firdaus. (2026). Data-Driven Clustering of Stunting Prevention Services for Pregnant Women and Infants Using Fuzzy C-Means. Artificial Intelligence in Lifelong and Life-Course Education, 1(2), 34–48.

INTRODUCTION

Stunting is a condition of growth in toddlers with a body that is too short compared to toddlers their age (Diana et al., 2024). Stunting can also be said to be a growth disorder in toddlers caused by malnutrition and repeated infections (Fadilah et al., 2022; Harlina et al., 2021). The health of infants and toddlers is very important to pay attention to because at this time their physical and mental growth and development is very fast (Rosyida et al., 2024). Infant and toddler health efforts include management and referral, nutrition, monitoring growth and development, immunization, rehabilitation and long-term care in chronic diseases, parenting and developmental stimulation (Rosyida et al., 2024).

The results of the Indonesian Nutrition Survey (SSGI) in 2022, the stunting rate in Indonesia is still high at 21.6%, despite a decrease from the previous year of 24.4% in 2021. In South Sulawesi Province, E-PPGBM measurement data shows a decrease in stunting prevalence from 11.41% in 2021 to 9.94% in 2022, while SSGI recorded a decrease from 34.7% to 31.1% in the same period. While this downward trend is positive, it still falls short of the 29.60% target set for 2022, with targets of 24.74% in 2023 and 14% in 2024 (Iskandar & Fatah, 2024).

Stunting in infants and toddlers is generally influenced by the nutritional status of the mother before and during pregnancy, as well as parenting and fulfillment of nutrients during pregnancy (Christiana et al., 2022; Marbun et al., 2024). besides that blood deficiency in pregnant women can also inhibit growth during postnatal and risk stunting (Hendrawati et al., 2024; Rahmawati et al., 2025). The consumption of nutrients obtained by toddlers since they were born also has a major effect on their growth and development, the non-implementation of IMD which triggers the non-achievement of exclusive breastfeeding in the first 6 months will affect the risk of stunting (Robbani et al., 2024).

Various studies have been conducted to optimize clustering methods in grouping regions, individuals, or facilities based on health indicators, including to support the handling of stunting and other health services. One of the studies conducted by Fadilah grouped districts/cities in Indonesia based on the level of stunting in toddlers using the K-Means algorithm. By applying the Elbow method, this study successfully found two optimal clusters that are expected to help the government in planning more appropriate health interventions. However, the K-Means algorithm has a weakness in the form of sensitivity to the number of initial clusters determined [2]. Another study by Ibanez and Wiriasto combined the Principal Component Analysis (PCA) method with K-Means to determine regional characteristics in Demak Regency. This study formed two optimal clusters that represented zones with different types. The use of PCA in this study helped in reducing the dimension of the data, thus improving the efficiency of the clustering process. However, the limitation of K-Means in selecting the number of clusters also affects the final result (Ibanez et al., 2024).

The study by Cytry et al. (2023) used the K-Means Clustering algorithm to identify villages at high risk of certain health problems. The villages were grouped into three clusters based on their level of risk. Although this method is effective, K-Means remains sensitive to the selection of the appropriate number of clusters. Research by Suraya & Wijayanto (2022)compared the effectiveness of several clustering methods such as Hierarchical Clustering, K-Means, and Fuzzy C-Means in classifying provinces in Indonesia based on human development indicators. This study found that the Average Linkage method produces clusters with better accuracy and efficiency than other methods, although it has high complexity and requires large computational resources.

Furthermore, research by Mahardika & Abadi (2024) compared the K-Means and Fuzzy C-Means methods in mapping areas vulnerable to stunting in toddlers in Indonesia. This study shows that the Fuzzy C-Means method is more optimal in detecting vulnerable clusters and is very helpful in identifying areas that need more attention, although both methods are sensitive to outliers. Another study conducted used Fuzzy C-Means and K-Medoids algorithms to cluster districts/cities in Central Java based on infant and toddler health indicators. The validation results showed two optimal clusters, with Fuzzy C-Means chosen as the best method because it has a smaller standard deviation ratio (1.1062) (Fadilah et al., 2022).

Research by Ebrison et al. (2025) compared the K-Means and Fuzzy C-Means algorithms in grouping toddler data based on weight and height. Validation results using Silhouette Index show that Fuzzy C-Means is slightly superior with a value of 0.564 compared to K-Means of 0.563. This research shows the effectiveness of FCM in handling data that may have overlapping boundaries between clusters [16]. Meanwhile, research conducted by Farida et al. compared K-Means and Fuzzy C-Means methods in clustering health centers in Balikpapan City based on health service coverage. The evaluation result using Silhouette Coefficient shows that K-Means gives better result with the value of 0.2740 compared to Fuzzy C-Means of 0.2322. This study emphasizes the importance of clustering health facilities to support the readiness of the buffer zone of the National Capital City (Hayati et al., 2023). Furthermore, research by Nazilaturrahma et al. (2024) optimizes the clustering of routine immunization participants in Central Java Province using the Fuzzy C-Means algorithm. Validation using Silhouette Coefficient and Davies Bouldin Index showed the best results in 3 clusters with SC value of 0.5847 and DBI of 0.7785. This study successfully identified areas with high and low immunization coverage, and suggested further interventions in areas with less equitable services. Finally, research by Vina et al. applied the Fuzzy C-Means method to cluster blood donors at PMI Cirebon City based on age. The clustering results produced three clusters, namely young (17-34 years), middle-aged (35-50 years), and old (51-64 years). Validation was done through manual calculation and visualization with RapidMiner. Although there were data mismatches in some cases, the final results showed an accurate age classification and the dominance of the young cluster as the largest group of donors (Paulina et al., 2024).

Various studies have applied clustering methods such as K-Means and Hierarchical Clustering to cluster healthcare and stunting prevention data. Although these methods are quite popular, they have limitations in terms of flexibility, especially when faced with complex and not strictly separated data, such as maternal and infant health service data. Fuzzy C-Means is one of the more adaptive alternatives because it is able to accommodate multiple membership degrees in each data, making it more suitable to describe real conditions in the field. This research uses the Fuzzy C-Means method to cluster regions in South Sulawesi based on the coverage of stunting prevention services, with the aim of producing more accurate segmentation and supporting targeted policy interventions.

One of the main challenges is how to develop a clustering model that is able to handle variations in the characteristics of data on pregnant women and infants prone to stunting, given the different health service needs between groups. Currently, existing models are not able to handle these variations well because they tend to use methods that assume a fixed cluster structure. This research aims to improve the accuracy and effectiveness of clustering stunting prevention services for pregnant women and infants in South Sulawesi by developing a Fuzzy C-Means-based clustering approach. This method was chosen for its ability to provide flexibility with multiple membership degrees, where each data can be a member of more than one cluster with a probability value. The Fuzzy C-Means approach is suitable for handling health data that has complex characteristics and needs, such as data on pregnant women and infants who are vulnerable to the risk of stunting (Nisa et al., 2024).

This research is expected to contribute to the development of more adaptive health service clustering methods, especially in the context of stunting prevention in Indonesia. In addition, the findings are also expected to be adopted by relevant parties in designing more specific and effective data-based health interventions. The results of this study are also expected to be used as a basis for developing maternal and child health policies in South Sulawesi, with the long-term goal of reducing stunting rates in the area.

METHOD

Research Design

In this research, a method with a series of sequential processes is used, including the data collection stage, preprocessing stage, clustering stage, result stage, and finally the evaluation stage. These stages can be seen in Figure 1.
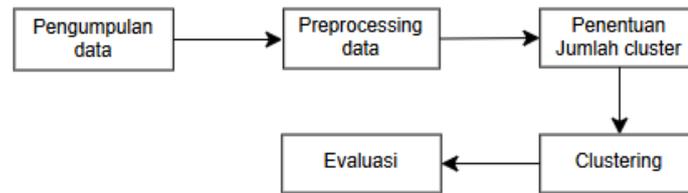
Figure 1. Research Stages

Data Collection

In this stage, the data used for clustering was obtained from Satu Data Indonesia with a total of 2,267 data that provided structured information related to the health conditions of pregnant women and children aged 0-59 months. The data collected includes information on the condition of pregnant women, such as the number of pregnant women with normal conditions, the number of pregnant women who conduct pregnancy/postpartum examinations, and the number of pregnant women who take blood supplement tablets during pregnancy. In addition, data on the condition of children aged 0-59 months is also included, such as the number of children with normal nutritional status, the number of children monitored for growth and development at posyandu or other health services, and the number of children who are malnourished or stunted. The collected attributes can be seen in Table 1.

Table 1. Dataset

No	Attribute Name	Description
1	Name_Province	Name
2	Name_District	Name
3	Name_District	Name
4	Village	Name
5	Normal pregnant women	Number of normal pregnant women
6	Normal nutrition baby	Number of children with normal nutrition
7	Pregnant mothers check	Number of pregnant women Check pregnancy
8	Pregnant women supplement	Number of pregnant women taking Blood Addition Tablet (TTD)
9	Child growth and development	Number of children monitored for growth and development
10	Undernourished children	Number of children experiencing malnutrition

Preprocessing

At this stage, data preprocessing is carried out on the columns of the number of pregnant women in normal conditions, the number of pregnant women checking pregnancy / postpartum, the number of pregnant women taking blood supplement tablets, the number of children aged 0-59 months with normal nutritional status, the number of children aged 0-59 months who are monitored for growth and development, and the number of children aged 0-59 months who are malnourished / stunted. This stage is carried out to clean the data, eliminate errors and duplications contained in the data that has been collected, and equalize the data format so that it is easily implemented into the clustering algorithm. The data format used includes numeric numbers in each column, ensuring there are no empty values, and normalizing the data into the range [0, 1] to equalize the scale across attributes. The data before normalization can be seen in Table 2.

Table 2. Data before normalization

Mother_Pregnant_ Normal	Baby_Nutrition Normal	Pregnant Mother Check	Pregnant Mother Supplements	Child_Growth Growth and Development	Child_Nutrition Lack
9	9	9	9	26	12
6	41	6	6	60	17
10	88	0	0	2	2
16	13	0	0	0	0
......	......	......	......	......	......
3	49	3	3	69	28

Determination of the Number of Clusters

At this stage, the Elbow Method is used to determine the optimal number of clusters. Elbow Method is a technique used to determine the optimal number of clusters in clustering analysis, especially in algorithms such as K-Means (Kustanto et al., 2024). The formula used for Elbow calculation is defined by the following Equation 1 (Juanita & Cahyono, 2024).

$$SSE = \sum_{k=1}^{k} \sum_{x1 \in sk} ||X_1 - C_k||_2^2 \tag{1}$$

This method calculates the distortion value or Fuzzy C-Means objective function which indicates how well the data in the cluster is grouped based on its proximity to the centroid (S.Intam et al., 2024). The "Elbow" point on the distortion graph is considered the optimal number of clusters because it provides a balance between good data separation and model complexity that is not excessive (Djun et al., 2024).

Clustering

The clustering stage is carried out to label the data using the Fuzzy C-Means clustering method. In the FCM algorithm, distance calculations are used to measure the extent of similarity between data points to determine the probability that the data point belongs to a particular cluster (S.Intam et al., 2024). At the beginning of the clustering stage, the number of clusters is initialized with 2. After the clustering process is complete, the next step is to identify clusters that represent groups of pregnant women or children aged 0-59 months with good health conditions, such as pregnant women who routinely conduct examinations and children with normal nutritional status, as well as clusters that represent groups that require more attention, such as pregnant women who do not routinely conduct examinations or children who are malnourished or stunted.

Evaluation

The evaluation stage is an important step in the clustering process to assess the extent to which the clustering results can correctly describe the patterns in the data. In this study, the evaluation is carried out using the Silhouette Coefficient to evaluate the quality of separation between clusters, with values ranging from -1 to 1. values close to 1 indicate that the clusters are well formed and mutually separated, values close to 0 indicate overlap between clusters, and negative values indicate that the data may be in the lowest cluster. The formula used for calculating the Silhouette Score is defined by the following Equation 2 (Risal et al., 2024):

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \tag{2}$$

RESULTS AND DISCUSSION

RESULT

In this study, the data that has been collected is 2,267 data with several attributes that have different range values. To overcome this, normalization is carried out attributes so that attributes that have different ranges can be compared directly without causing bias. The normalization results can be seen in Table 3.

Table 3. Data after normalization

Mother_Pregnant_Normal	Baby_Nutrition Normal	Pregnant Mother Check	Pregnant Mother Supplements	Child_Growth Growth and Development	Child_Nutrition Less
0.065693431	0.014423077	0.063380282	0.063380282	0.040688576	0.018779343
0.04379562	0.065705128	0.042253521	0.042253521	0.093896714	0.026604069
0.072992701	0.141025641	0	0	0.00312989	0.00312989
0.116788321	0.020833333	0	0	0	0
…….	…….	……	………	…..	……
0.058394161	0.033653846	0.056338028	0.056338028	0.053208138	0.007824726

To be able to group data with Fuzzy C-Means, it is necessary to determine the number of clusters required. Optimization of determining the number of clusters can use the elbow method. The point that shows the optimal cluster in this method will form an elbow in a graph. The results of determining K with the elbow method can be seen in Figure 2.
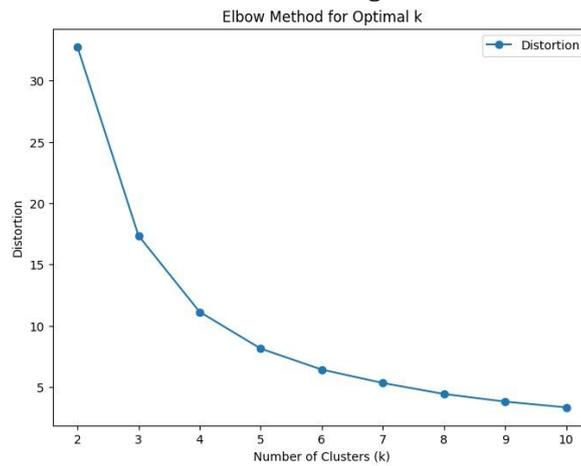

Figure 2. Plot of Determination of Optimal Number of Clusters Elbow Method in Pregnant Women
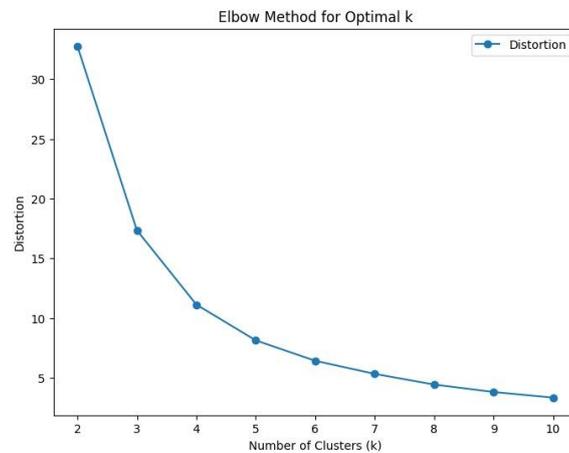

Figure 3. Plot of Determination of the Optimal Number of Clusters by the Elbow Method on Infants

Based on Figure 2 and Figure 3 shows the process of finding the optimal K value in the clusters of pregnant women and infants formed, namely as many as 3 clusters. However, there are some significant differences as in Figure 2 shows a plot for pregnant women with a sharper decrease in distortion than k = 2 and k = 3 which indicates that the addition of clusters has a significant effect on the quality of data segmentation. Meanwhile, Figure 3 shows a slow decrease in the distortion of infants, with the decrease still visible up to k = 4 or k = 5 before reaching stability. Therefore, it can be concluded that the distribution of infant data is more spread out than pregnant women. After determining the number of clusters, the next step is to perform clustering using the Fuzzy C-Means algorithm. The following visualization of clusters using Fuzzy C-Means can be seen in Figure 4.
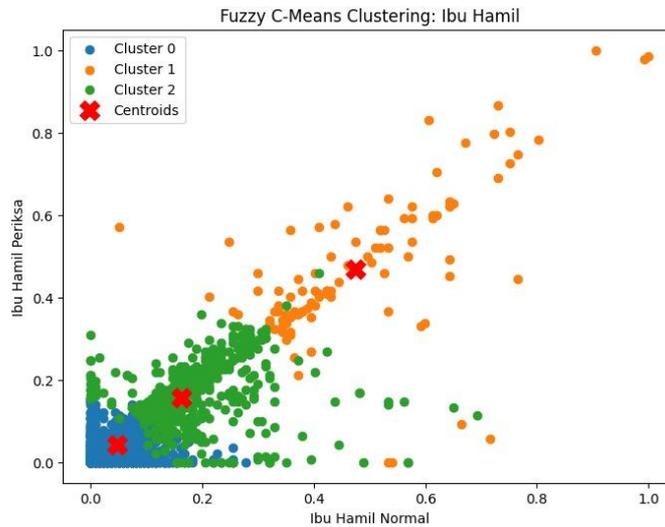

Figure 4. FCM Visualization for Pregnant Women Clustering

Based on the visualization in Figure 4, it is a combination of 3 clusters of pregnant women with centroid points. The total data in the first cluster (C0) is 696. To represent C0, the top 5 data points are taken as samples. In the first cluster, there are areas with prominent values, such as Gantarang Sub-district, Bialo Village which has a normal pregnant woman value of 0.1898. Pregnant women who regularly check their pregnancy is 0.1831 and pregnant women who receive supplements is 0.2113. In addition, Pasimasunggu Sub-district, Kembang Ragi Village also shows a fairly good value, namely 0.1387 for normal pregnant women, 0.1479 for pregnant women who regularly check pregnancy, and 0.1479 for pregnant women who get supplements. These areas represent the main characteristics in the first cluster, where some sub-districts show better performance of pregnant women's health services than other areas, the results of which can be seen in Table 4.

Table 4. Cluster 0 Pregnant Women

Province Name	District Name	Sub-district Name	Village Name	Normal Pregnant Mother	Pregnant Mother Check-up	Pregnant Mother Supplements	Cluster
South Sulawesi	Selayar Islands	Pasimasunggu	Yeast Flower	0,13868	0,14788	0,14788	0
South Sulawesi	Selayar Islands	Pasimarannu	Batu Bingkung	0,16058	0,15492	0,15492	0
South Sulawesi	Selayar Islands	Taka Bonerate	Rajuni	0,16788	0,14084	0,14084	0
South Sulawesi	Selayar Islands	Gantarang	Bialo	0,18978	0,183098	0,21126	0

...	...	...	...	....	....	....	...
South Sulawesi	North Toraja	Awan Karua	Rante Buntu Karua	0,08759	0,11971	0,11971	0

Furthermore, the total data in the second cluster (C1) is 101. To represent C1, the top 5 data were taken as samples. In this cluster, there are several areas that stand out with the highest values. One of them is Gantarang Sub-district, Gattareng Village, which has a value of 0.5255 for normal pregnant women, 0.5634 for pregnant women who regularly check their pregnancy, and 0.5634 for pregnant women who receive supplements. In addition, Paenre Lompoe Village in the same sub-district has a normal pregnant woman value of 0.3796, pregnant women who regularly check their pregnancy by 0.4155, and pregnant women who receive supplements by 0.4155. Other areas, such as Bontomasila Village, also have prominent numbers with a value of pregnant women who receive supplements of 0.4225. The cluster results can be seen in Table 5, which shows that the areas in this group have fairly good coverage of maternal health services, with a focus on pregnant women who regularly receive health attention.

Table 5. Cluster 1 Pregnant Women

Province Name	District Name	Sub-district Name	Village Name	Normal Pregnant Mother	Pregnant Mother Check-up	Pregnant Mother Supplements	Cluster
South Sulawesi	Bulukumba	Gantarang	Gattareng	0,52554	0,56338	0,56338	1
South Sulawesi	Bulukumba	Gantarang	Paenre Lompoe	0,37956	0,41549	0,41549	1
South Sulawesi	Bulukumba	Gantarang	Bukit Harapan	0,39416	0,26760	0,33098	1
South Sulawesi	Bulukumba	Gantarang	Bontomasila	0,37226	0,21126	0,42253	1
......	....	....	......	....	....	....	....
South Sulawesi	North Toraja	Tondon	Tondon	0,49635	0,5	0,5	1

Based on Table 6, the total data in the third cluster (C2) is 1468. To represent C2, the top 10 data were taken as samples. In this cluster, there are several areas that stand out with the highest values, especially in the indicator of maternal health. One of the areas with the highest value in this cluster is Bontoharu Sub-district, Kahukahu Village, which has a normal pregnant woman value of 0.1168. In addition, Bonotoboru Village in the same sub-district has a normal pregnant mother value of 0.0657, while Bontolebang Village has a value of 0.0438. The areas in this cluster tend to have low mean scores on the pregnant women's health indicators. The results of this third cluster indicate that most of the areas in this group need more attention to improve the coverage of maternal health services.

Table 6. Cluster 2 Pregnant Women

Province Name	District Name	Sub-district Name	Village Name	Normal Pregnant Mother	Pregnant Mother Check-up	Pregnant Mother Supplements	Cluster
South Sulawesi	Selayar Islands	Bontoharu	Bontoborusu	0,06569	0,06338	0,06338	2
South Sulawesi	Selayar Islands	Bontoharu	Bontolebang	0,04379	0,04225	0,04225	2

South Sulawesi	Selayar Islands	Bontoharu	Bontotangnga	0,07299	0	0	2
South Sulawesi	Selayar Islands	Bontoharu	Kahukahu	0,11678	0	0	2
......	....	....	......	....	....	....	....
South Sulawesi	North Toraja	Awan Rante Karua	Londong Biang	0,02189	0,02112	0,02112	2

Thus, the Fuzzy C-Means method successfully clusters regions based on pregnant women's health indicators. The first group shows areas with good coverage of maternal health services, characterized by a high proportion of pregnant women who routinely go for antenatal check-ups and receive supplements. The second group shows areas with moderate health service coverage, with a focus on increasing the proportion of normal pregnant women. The third group shows areas with low health service coverage in general, thus requiring more attention to improve access and quality of health services for pregnant women in the area.

The clustering process was carried out on infant child data which also produced 3 clusters as the optimal result using the elbow method. The following cluster visualization using Fuzzy C-Means is shown in Figure 5.
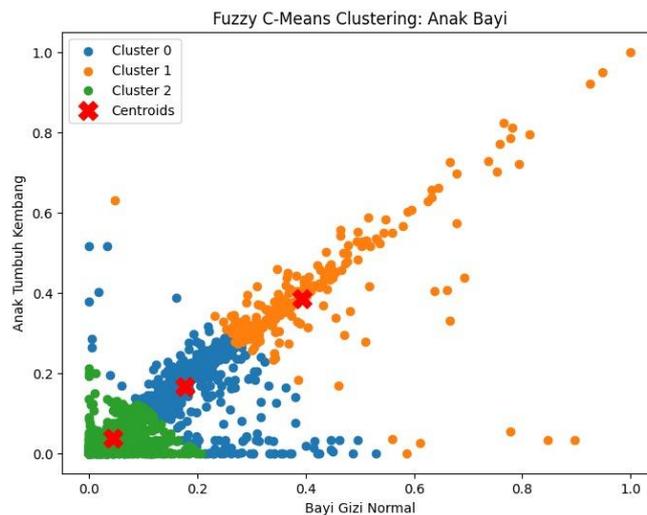

Figure 5. FCM Visualization for Infant Child Clustering

The total data in the first cluster (C0) amounted to 1,235. to represent C0, the top 5 data points were taken as samples. In the first cluster (C0), there are areas with prominent values, such as Bontoharu Sub-district, Bontotangnga Village, which has a value of infants with normal nutrition of 0.1410, children with good growth and development and children with poor nutrition both have a value of 0.0031. Bontohaaru Sub-district, Bontolebang Village, also showed good values of 0.0657 for normally-nourished infants, 0.0938 for children with good growth and development, and 0.0266 for malnourished children. These areas represent the main characteristics of the first cluster, with some sub-districts performing better than others in terms of child health services. The results can be seen in Table 7.

Table 7. Cluster 0 Infant Children

Province Name	District Name	Subdistrict Name	Village Name	Normal Nutrition Infants	Children with Growth and Development	Undernourished Children	Cluster
South Sulawesi	Selayar Islands	Bontoharu	Bontoborusu	0,01442	0,04068	0,01877	0

South Sulawesi	Selayar Islands	Bontoharu	Bontolebang	0,06570	0,09389	0,02660	0
South Sulawesi	Selayar Islands	Bontoharu	Bontotangnga	0,14102	0,00312	0,00312	0
South Sulawesi	Selayar Islands	Bontoharu	Kahukahu	0,02083	0	0	0
......	....	....	......	....	....	....	....
South Sulawesi	North Toraja	Awan Rante Karua	Londong Biang	0,07852	0,10798	0,04381	0

Furthermore, in Table 8 the total data in the second cluster (C1) is 231. To represent this cluster, the top 5 data are taken as samples. One of the prominent areas is Mandai Sub-district, Bonto Matene Village, with values of normal nutrition infants, developmental children, and undernourished children of 1.0000 each. In addition, Damai village in Tanralili sub-district recorded 0.9487 normal nutrition for infants, 0.9499 for children with good growth and development, and 0.0485 for malnourished children. Other areas, such as Maradekaya village in Bajeng sub-district, had a normal infant nutrition score of 0.9247, a good growth and development score of 0.9202, and a malnourished child score of 0.0156. These results indicate that the areas in this cluster have fairly good coverage of child health services, especially in the aspects of normal nutrition and child growth and development.

Table 8. Cluster 1 Infant Children

Province Name	District Name	Subdistrict Name	Village Name	Normal Nutrition Infants	Children with Growth and Development	Undernourished Children	Cluster
South Sulawesi	Bulukumba	Gantarang	Palambarae	0,36698	0,44913	0,10172	1
South Sulawesi	Bulukumba	Gantarang	Paenre Lompoe	0,37179	0,41001	0,04538	1
South Sulawesi	Bulukumba	Gantarang	Benteng Gattareng	0,30448	0,26447	0,04851	1
South Sulawesi	Bulukumba	Gantarang	Taccorong	0,28846	0,29107	0,00938	1
......	....	....	......	....	....	....	....
South Sulawesi	North Toraja	Tondon	Tondon	0,53205	0,53364	0,00625	1

In Table 9, the total data in the third cluster (C2) is 799. To represent this cluster, the top 5 data were taken as samples. One of the areas with the highest score is Bontonompo Sub-district, Panciro Village, with a normal nutrition infant score of 0.8477. In addition, Bontokassi Village in Bontomarannu Sub-district recorded a normal nutritional infant value of 0.8278, while Bajeng Village in Bajeng Sub-district had a value of 0.8132. The areas in this cluster tend to have fairly good child health indicator values in some locations, but still need more attention to improve health service coverage evenly.

Table 9. Cluster 2 Infant Children

Province Name	District Name	Sub-district Name	Village Name	Normal Nutrition Infants	Children with Growth and Development	Undernourished Children	Cluster
South Sulawesi	Selayar Islands	Bontomatene	Tanete	0,13301	0,14553	0,14553	2
South Sulawesi	Selayar Islands	Bontosikuyu	Patikarya	0,16506	0,16118	0	2

South Sulawesi	Selayar Islands	Pasimarannu	Batu Bingkung	0,19230	0,18779	0,18779	2
South Sulawesi	Selayar Islands	Pasimarannu	Lambego	0,11378	0,12676	0,01564	2
……	………	………	……	………	………	………	…….
South Sulawesi	North Toraja	Awan Rante Karua	Batu Lotong	0,20673	0,20187	0,20187	2

The FCM method successfully clusters regions based on infant child health indicators. The first group shows areas with excellent child health service coverage, characterized by a high proportion of normally nourished infants and children with optimal growth and development. The second group shows areas with moderate coverage of child health services, with some areas still needing to increase the proportion of children with normal nutrition and reduce the number of undernourished children. The third group shows areas with low coverage of child health services in general, requiring more attention to improve access to and quality of health services, especially on nutrition and child development in these areas.

Next, conduct an evaluation stage to measure the results or quality of the clusters and models formed. The technique used is the Sillhouette Coefficient as shown in Table 10 which shows the cluster results of pregnant women.

Table 10. Silhouette Coefficient of Pregnant Women Cluster Results

Cluster	Silhouette Value	Description
Cluster 0	0.63807	High
Cluster 1	0.33828	Medium
Cluster 2	0.30211	Low

Cluster 0 has a Silhouette Coefficient value of 0.63807, indicating that the data in this cluster has excellent cohesion and is clearly separated from other clusters. This shows that Cluster 0 has the highest clustering quality compared to other clusters. Furthermore, cluster 1 has a Silhouette Coefficient value of 0.33828, reflecting that this cluster has moderate separation. Although the data in this cluster is well-defined, there is some possibility of overlap with other clusters. Cluster 2 has a Silhouette Coefficient value of 0.30211, which is the lowest value among the other clusters. This indicates that the data in cluster 2 is less cohesive and has some overlap with other clusters. In addition to the pregnant mother clusters, the infant child results can be seen in Table 11.

Table 11. Sillhouette Coefficient of Infant Child Cluster Results

Cluster	Silhouette Value	Description
Cluster 0	0.29236	Medium
Cluster 1	0.20400	Low
Cluster 2	0.64537	High

Cluster 0 has a Silhouette Coefficient value of 0.29236, which indicates that this cluster has a moderate quality of separation. The data in cluster 0 is well-defined, but there is some overlap with other clusters, especially in areas with low to moderate child health indicator values. Furthermore, cluster 1 has a Silhouette Coefficient value of 0.20400, which is the lowest value among the other clusters. This indicates that the data in cluster 1 lacks cohesion and has a significant possibility of overlap with other clusters. Although some regions in cluster 1 show high indicators, the wide spread of the data and the variation between regions cause the quality of separation of this cluster to be lower. Cluster 2 has a Silhouette Coefficient value of 0.64537, which is the highest value among the other clusters. This indicates that the data in Cluster 2 is highly cohesive and has a clear separation

from other clusters. However, even though this cluster has good separation, there are still challenges in equitable distribution of health services in various locations.

DISCUSSION

The results showed that the Fuzzy C-Means (FCM) method was able to cluster regions in South Sulawesi based on the quality of stunting prevention services for pregnant women and infants. The formation of three clusters in each category provides a fairly clear picture of the distribution of regions based on health indicators. For the category of pregnant women, clusters with high values describe areas that have good service coverage, characterized by high rates of antenatal check-ups and consumption of blood supplement tablets. Meanwhile, clusters with the lowest values indicate areas that require special attention due to the low quality of health services received by pregnant women. A similar pattern is seen in the infant category, where the best clusters indicate areas where children are well-nourished and have optimal development, while clusters with low scores indicate areas with challenges in nutrition and monitoring child growth and development.

These results are in line with previous studies, such as Mahardika & Abadi (2024), which show that the FCM method is superior in handling complex and diverse data such as health data, compared to conventional clustering methods such as K-Means. FCM provides more flexibility because it can accommodate multiple memberships, making it suitable for cases that have variations in characteristics between regions. In addition, this finding also strengthens the results of Suraya & Wijayanto (2022) research, which states that non-hierarchical methods such as FCM are more effective in distinguishing patterns based on human development indicators, including health services.

Evaluation of clustering quality using the Silhouette Coefficient shows variations in the level of cohesion and separation between clusters. In the pregnant women category, there are clusters with the highest Silhouette value of 0.638, indicating that the regions within the group are highly defined and clearly separated from other clusters. However, there are also clusters with lower values, such as 0.302, indicating overlapping data between clusters. A similar condition was also found in the infant category, where one cluster obtained a Silhouette value of 0.645 and another cluster only 0.204. This indicates that although FCM is quite effective, data complexity and the spread of service quality in the field remain challenges that affect the segmentation results.

Overall, the results of this study make a significant contribution to the mapping of stunting-prone areas with a more flexible and adaptive approach. Information from the clustering results can be utilized by local governments and health agencies in designing more targeted interventions, such as adding health facilities, increasing supplement distribution, and educating the community. However, this study has limitations related to data quality and completeness, so for further development it is recommended to integrate the FCM method with optimization or dimension reduction techniques such as PCA to improve the accuracy of clustering results.

CONCLUSION

This study uses clustering with the Fuzzy C-Means algorithm to cluster stunting prevention services for pregnant women and infants in South Sulawesi. The clustering results showed three distinct groups for each category. For pregnant women, cluster 0 in Gantarang sub-district, Bialo village represents areas with good health services characterized by a high number of pregnant women who regularly check their pregnancies and take supplements. Cluster 1 in Gantarang sub-district, Gattareng village, represents an area with moderate health services, while cluster 2 in Bontoharu sub-district, Kahukahu village, represents an area where health services need more attention, with low health service coverage.

In the infant category, the clustering results also form three groups: cluster 0 includes areas in Gantarang sub-district in Gattareng village with good child health services, with normally nourished

infants and children with optimal growth and development. Cluster 1 includes areas in Mandai sub-district in Bonto Matene village showing moderate service coverage, and cluster 2 in Bontonompo sub-district in Panciro village reflecting areas with poor health services, particularly in nutrition and growth and development. Evaluation using the silhouette coefficient showed variation in clustering quality, with some clusters having good cohesion and being clearly separated, while others showed overlap between data. This research can be used to design more targeted health interventions and support policy development to reduce stunting in South Sulawesi.

ACKNOWLEDGMENT

AUTHOR CONTRIBUTION STATEMENT

HZI led the study design and data acquisition. AS and AANR conducted preprocessing and clustering analysis. DFS and F contributed to validation, interpretation, and policy relevance discussion. HZI drafted the manuscript. All authors participated in revisions and approved the final version.

AI DISCLOSURE STATEMENT

AI tools were used solely to enhance linguistic clarity and coherence. No AI systems were involved in data processing, clustering analysis, or interpretation of findings. The authors take full responsibility for all scientific content.

REFERENCES

Christiana, I., Nazmi, A. N., & Anisa, F. H. (2022). Hubungan pola asuh ibu dengan kejadian stunting pada balita di desa kertosari wilayah kerja puskesmas kertosari banyuwangi. *Jurnal Ilmiah Keperawatan (Scientific Journal of Nursing)*, *8*(2), 397–409. https://doi.org/10.33023/jikep.v8i2.1161

Cytry, D. M., Defit, S., & Nurcahyo, G. (2023). Penerapan Metode K-Means dalam Klasterisasi Status Desa terhadap Keluarga Beresiko Stunting. *Jurnal KomtekInfo*, *10*(3), 122–127. https://doi.org/10.35134/komtekinfo.v10i3.423

Diana, S. N., Firmani, U., Rahim, A. R., Widiharti, W., & Sukaris, S. (2024). SOSIALISASI PEMANFAATAN TANAMAN OBAT KELUARGA UNTUK PENCEGAHAN STUNTING. *DedikasiMU : Journal of Community Service*, *6*(1), 105. https://doi.org/10.30587/dedikasimu.v6i1.7497

Djun, S. F., Gunadi, I. G. A., & Sariyasa, S. (2024). Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means Clustering pada Model Data RFM. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *5*(4), 354–364. https://doi.org/10.35746/jtim.v5i4.434

Ebrison, M. A. H., Baenudin, M., & Ridwan, M. (2025). Analisis Clustering Data Balita dengan Algoritma K-Means dan Fuzzy C-Means: Sebuah Studi Komparatif Menggunakan Silhouette Index. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, *13*(2), 266–271. https://doi.org/10.26418/justin.v13i2.82529

Fadilah, A., Pangestu, M. N., Lumbanbatu, S., & Defiyanti, S. (2022). Pengelompokan kabupaten/kota di indonesia berdasarkan faktor penyebab stunting pada balita menggunakan algoritma K-MEANS. *JIKO (Jurnal Informatika Dan Komputer)*, *6*(2), 223. https://doi.org/10.26798/jiko.v6i2.581

Harlina, H., Hidayanty, H., & Nur, M. I. (2021). Studi Fakor Resiko Kejadian Stunting Pada Balita Di Wilayah Dataran Tinggi Dan Dataran Rendah. *Jurnal Ilmiah Kesehatan Sandi Husada*, *10*(2), 501–510. https://doi.org/10.35816/jiskh.v10i2.634

Hayati, F. N., Silfiani, M., & Nurlaily, D. (2023). Perbandingan pengelompokkan pusat kesehatan masyarakat di kota balikpapan menggunakan metode k-means dan fuzzy C-MEANS. *VARIANCE: Journal of Statistics and Its Applications*, *5*(1), 55–66. https://doi.org/10.30598/variancevol5iss1page55-66

Hendrawati, S., Mardiah, W., & Febri, R. A. (2024). Pemenuhan vitamin d pada ibu hamil untuk mencegah stunting: sebuah narrative review. *Jurnal Ilmu Keperawatan Dan Kebidanan*, *15*(1), 50–67. https://doi.org/10.26751/jikk.v15i1.2178

Ibanez, G. F., Wiriasto, G. W., & Rosmaliati. (2024). Kombinasi Principal Component Analysis dengan Algoritma K-Means untuk Klasterisasi Data Stunting. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *5*(1), 131–141. https://doi.org/10.30865/klik.v5i1.1977

Juanita, S., & Cahyono, R. D. (2024). K-Means Clustering With Comparison of Elbow and Silhouette Methods for Medicines Clustering Based on User Reviews. *Jurnal Teknik Informatika (JUTIF)*, *5*(1), 283–289. https://doi.org/10.52436/1.jutif.2024.5.1.1349

Kustanto, Y., Arumi, R., Sasongko, D., Ully Artha, E., & Prabowo, N. A. (2024). Implementasi K-Modes Clustering Untuk Pengelompokan Data Bermain Game Pada Mahasiswa Ditinjau Dari Durasi Belajarnya. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(5), 2495–2505. https://doi.org/10.30865/klik.v4i5.1619

Mahardika, B. W., & Abadi, A. M. (2024). IMPLEMENTATION OF K-MEANS AND FUZZY C-MEANS CLUSTERING FOR MAPPING TODDLER STUNTING CASES IN GUNUNGKIDUL DISTRICT. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *18*(4), 2231–2246. https://doi.org/10.30598/barekengvol18iss4pp2231-2246

Marbun, A. H., Badi'ah, A., & Badi'ah. (2024). Pengaruh Pola Asuh Pola Pemberian Makan Dan Status Gizi Ibu Saat Hamilterhadap Kejadian Stunting Pada Balita Usia 12-59. *Jurnal Media Informatika (JUMIN)*, *6*(2), 760–768.

Nazilaturrahma, F., Sudarno, S., & Tarno, T. (2024). Optimasi pengklasteran menggunakan fuzzy c-means pada peserta imunisasi rutin di provinsi jawa tengah. *Jurnal Gaussian*, *13*(2), 363–372. https://doi.org/10.14710/j.gauss.13.2.363-372

Nisa, H. A., Salma, A., Vionanda, D., & Mukhti, T. O. (2024). Impelementation of Subtractive Fuzzy C-Means Method in Clustering Provinces in Indonesia Based on Factors Causing Stunting in Toddlers. *UNP Journal of Statistics and Data Science*, *2*(2), 165–172. https://doi.org/10.24036/ujsds/vol2-iss2/164

Paulina, V., Asfi, M., & Sokibi, P. (2024). Klasterisasi pendonor berdasarkan usia menggunakan metode fuzzy c-means (studi kasus: pmi kota cirebon). *Jurnal Jaringan Sistem Informasi Robotik (JSR)*, *8*(2), 222–229. http://ojsamik.amikmitragama.ac.id

Rahmawati, R., Kirana, R., Laili, F. J., & Isnaniah, I. (2025). Hubungan Kejadian Anemia Pada Ibu Hamil Dengan Kejadian Stunting di Wilayah Kerja Puskesmas Pekapuran Raya. *Jurnal Penelitian Multidisiplin Bangsa*, *1*(8), 1136–1143. https://doi.org/10.59837/jpnmb.v1i8.209

Risal, A. A. N., Andayani, D. D., Suherman, M. I., & Kaswar, A. B. (2024). Utilizing the K-Means Clustering Algorithm for Analyzing Student Achievement Assessment at SMK Negeri 1 Gowa. *Journal of Embedded Systems, Security and Intelligent Systems*, *05*(1), 60–67. https://doi.org/10.59562/jessi.v5i1.2178

Robbani, M. A., Firmansyah, G., Widodo, A. M., & Tjahjono, B. (2024). Clustering of Child Stunting Data in Tangerang Regency Using Comparison of K-Means, Hierarchical Clustering and DBSCAN Methods. *Asian Journal of Social and Humanities*, *2*(12), 3105–3115. https://doi.org/10.59888/ajosh.v2i12.422

Rosyida, I. A., Arisandra, M. L., Noviyanti, D. A., Aprilian, R., Cahyono, C. B., & Abidin, K. U. (2024). Pemantauan Status Gizi Balita Dan Pentingnya Pemberian Pmt Pada Balita Desa Durikedungjero, Ngimbang, Lamongan. *Jurnal Pengabdian Masyarakat : BAKTI KITA*, *5*(1), 24–33. https://doi.org/10.52166/baktikita.v5i1.5475

S.Intam, R. N. J., Wulandari, Risal, A. A. N., & Surianto, D. F. (2024). Klasifikasi Mahasiswa Berprestasi Menggunakan Fuzzy C-Means Dan Naive Bayes. *Jurnal Ilmiah Informatika Global*, *15*(1), 9–16. https://doi.org/10.36982/jiig.v15i1.3666

Suraya, G. R., & Wijayanto, A. W. (2022). Comparison of Hierarchical Clustering, K-Means, K-Medoids, and Fuzzy C-Means Methods in Grouping Provinces in Indonesia according to the Special Index for Handling Stunting. *Indonesian Journal of Statistics and Its Applications*, *6*(2), 180–201. https://doi.org/10.29244/ijsa.v6i2p180-201

Iskandar, M. S., & Fatah, Z. (2024). Implementasi Metode Algoritma K-Means Clustering Untuk Menentukan Penerima Program Indonesia Pintar (PIP). *Gudang Jurnal Multidisiplin Ilmu*, *2*(10), 1–8. https://doi.org/10.59435/gjmi.v2i11.1027