Implementation of the Fuzzy C-Means Method for Clustering Stunting Areas in West Sumatra

Dr. Susila Bahri[1*], Rahayu Ningsih[2], Dr. Ahmad Iqbal Baqi[3]

Department of Mathematics and Data Science, University of Andalas

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2024.904013

Received: 08 March 2024; Revised: 02 April 2024; Accepted: 05 April 2024; Published: 07 May 2024

ABSTRACT

Although the prevalence of stunting in Indonesia decreased from 27.6% to 21.6% from 2020 to 2022, this value still exceeds the target prevalence set by the WHO. West Sumatra is one of the provinces that contribute significantly to this prevalence. This research aims to classify 19 regencies and cities into stunting vulnerability groups (high, moderate, low) based on factors contributing to stunting reduction. By employing the FCM method, the study identified four areas with high stunting vulnerability, eight with moderate stunting, and seven with low stunting We performed the computation using Matlab software.

Keywords— Stunting vulnerability, FCM method, Prevalence, Stunting, Factors contributing to stunting reduction

INTRODUCTION

According to the Indonesian Nutrition Status Survey (SSGI), the prevalence rate of stunting in Indonesia decreased from 27.6% to 21.6% from 2020 to 2022 [1]. This figure remains above the target prevalence set by the World Health Organization (WHO). West Sumatra province has the highest stunting prevalence rate, reaching 25.2%, surpassing the WHO standard. Among the 19 districts in this province, West Pasaman district has the highest prevalence rate at 35.5%, while Sawahlunto city has the lowest stunting rate at 13.7% [2].

According to Databoks, in 2022, out of 19 districts/cities in West Sumatra, there are still 11 districts/cities that have a prevalence rate of stunting above 20%. However, their stunting prevalence rates exceed the WHO standard in other towns or districts. These rates still need to be considered low and have yet to reach the government's target of reducing stunting prevalence by 14% [2].

Clustering is a technique used to group data into several clusters or groups so that the data within one cluster have a high degree of similarity. In contrast, the data between clusters have a low degree of similarity. Fuzzy clustering is a clustering method that considers the degree of membership in a fuzzy set as the basis for weighting in clustering [3]. Fuzzy C-Means determine the presence of each data point in a cluster based on its degree of membership [4].

In 2021, Tri Mulyaningsih et al. studied stunted children in Indonesia. The researchers used a multilevel determinant model to analyze the determinants of stunting at the household, district/city, and provincial levels by analyzing data on stunted children using [5]. In contrast to that study, this research utilizes the fuzzy c-means method to group regions or districts/cities based on factors contributing to stunting reduction.

The Matlab software assists in classifying stunting levels in an area.

A. Factors Influencing the Reduction of Stunting

The factors or variables influencing the prevalence of stunted toddlers used in this study are as follows: The availability of community health centers or the ratio of community health centers per 100,000 population (X_1), The availability of active integrated health posts or the proportion of villages with sufficient integrated health posts(X_2), The availability of doctors or the ratio of doctors per 100,000 population(X_3), The availability of nutritionists or the ratio of nutritionists per 100,000 population(X_4), The availability of midwives or the ratio of midwives per 100,000 population(X_5), Percentage of infants with low birth weight (X_6), Adequacy of infant health services or the ratio of the number of infants aged 29 standards(X_7), and adequacy of toddler health services or the ratio of the number of children under five years old receiving health services according to standards(X_8). Stunting factors influence the area or location, dividing it intothree clusters: areas with high, moderate, and low levels of stunting [6].

B. Fuzzy C-Means Method

The Fuzzy C-Means (FCM) method is a clustering technique frequently employed in the data analysis and identification of patterns [7]. The Fuzzy C-Means Algorithm is:[6]

1. Insert the data on factors influencing stunting in West Sumatra into the following Table I

TABLE I: The Factors Influencing Stunting West Sumatera In 2022

No	Regency/ City	X1	X2	X3	X4	X5	X6	X7	X8
1	Pariaman City	7,237	145,78	24,814	19,64	160,26	5,53	93,2	74,91
2	Solok Regency	4,776	155,34	9,5518	8,798	117,89	4,39	87,5	94,08
3	Payakumbuh City	5,582	115,12	14,652	8,373	78,842	2,72	87,4	68,78
4	Sijunjung Regency	5,41	133,57	11,651	14,56	138,15	6,66	72,2	62,07
5	Tanah Datar Regency	6,113	162,65	10,099	8,239	81,589	4,73	81	78,42
6	Pasaman Barat Regen	4,52	107,12	15,594	9,944	184,42	2,96	70,9	69,48
7	Pasaman Regency	5,205	141,5	11,71	13,01	192,89	2,64	82,5	83,36
8	Lima Puluh Kota Rege	5,665	143,93	14,934	11,07	125,14	3,14	72,4	61,02
9	Bukittinggi City	5,723	112,01	13,899	8,993	48,238	2,77	89,5	75,81
10	Dharmasraya Regency	5,965	105,23	20,024	15,34	199,82	4,44	62,8	70,02
11	Pesisir Selatan Regenc	4,066	122,94	9,293	7,938	176,18	3,5	94,6	90,22
12	Sawahlunto City	9,034	140,03	28,609	18,07	191,23	5,82	75,8	69,48
13	Agam Regency	4,252	144,2	14,051	6,84	77,278	3,96	69,9	52,75
14	Solok City	5,274	100,2	29,005	14,5	127,88	6,34	94,5	73,11
15	Padang Panjang City	6,914	136,56	27,658	6,914	89,888	2,78	82,7	73,3
16	Solok Selatan Regenc	4,241	132,52	10,072	5,831	77,392	1,49	71,4	61,71
17	Padang Pariaman Reg	5,732	141,93	11,006	9,859	152,25	4,42	87,3	77,17
18	Padang City	2,611	13,491	10,771	5,766	45,259	2,35	80,3	64,64
19	Mentawai Regency	16,78	11,186	26,845	24,61	286,35	4,28	60,5	41,48

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue IV March 2024

As elements of a matrix with dimensions n x m, where n represents the 19 districts in West Sumatra, and m represents the eight factors (variables) contributing to the reduction of stunting, is presented in Fig 1.

$$X_{nxm} = X_{19x8} = \begin{bmatrix}
7,237461099 & 145,783145 & 24,81415234 & 19,64453727 & 160,2580672 & 5,53 & 93,21 & 74,91 \\
4,775921313 & 155,3431248 & 9,551842626 & 8,797749787 & 117,8898471 & 4,39 & 87,53 & 94,08 \\
5,581719867 & 115,1229723 & 14,65201465 & 8,372579801 & 78,84179313 & 2,72 & 87,37 & 68,78 \\
5,40952159 & 133,5735716 & 11,65127727 & 14,56409659 & 138,1508591 & 6,66 & 72,17 & 62,07 \\
6,112534416 & 162,6465679 & 10,09896991 & 8,238633344 & 81,58904634 & 4,73 & 80,96 & 78,42 \\
4,519988519 & 107,1237279 & 15,59396039 & 9,943974742 & 184,4155316 & 2,96 & 70,92 & 69,48 \\
5,204521428 & 141,4979263 & 11,71017321 & 13,01130357 & 192,8925754 & 2,64 & 82,5 & 83,36 \\
5,664628259 & 143,9330544 & 14,93401995 & 11,07177341 & 125,1367879 & 3,14 & 72,37 & 61,02 \\
5,723115664 & 112,0095494 & 13,89899518 & 8,993467472 & 48,23768917 & 2,77 & 89,47 & 75,81 \\
5,964731395 & 105,2349039 & 20,0244554 & 15,33788073 & 199,8185017 & 4,44 & 62,76 & 70,02 \\
4,065685997 & 122,9386004 & 9,292996565 & 7,9377679 & 176,1797266 & 3,5 & 94,64 & 90,22 \\
9,0343758 & 140,0328249 & 28,6088567 & 18,0687516 & 191,2276211 & 5,82 & 75,78 & 69,48 \\
4,252132999 & 144,2027713 & 14,05052643 & 6,840387868 & 77,27789538 & 3,96 & 69,91 & 52,75 \\
5,273566249 & 100,1977587 & 29,00461437 & 14,50230719 & 127,8839815 & 6,34 & 94,53 & 73,11 \\
6,914433881 & 136,5600691 & 27,65773552 & 6,914433881 & 89,88764045 & 2,78 & 82,65 & 73,3 \\
4,240679781 & 132,5212432 & 10,07161448 & 5,830934699 & 77,392406 & 1,49 & 71,37 & 61,71 \\
5,732248945 & 141,9304839 & 11,00591797 & 9,859468185 & 152,248532 & 4,42 & 87,29 & 77,17 \\
2,611122293 & 13,49079851 & 10,77087946 & 5,766228397 & 45,25945308 & 2,35 & 80,3 & 64,64 \\
16,77833581 & 11,18555721 & 26,8453373 & 24,60822586 & 286,3502645 & 4,28 & 60,49 & 41,48
\end{bmatrix}$$

Fig 1 Matrix the factors influencing stunting in West Sumatra in 2022

2. Establishing parameters, parameter descriptions, and parameter values used in the Fuzzy C-Means method. The parameters and their descriptions are the number of clusters (c=3), the power (w=2), the maximum iteration (maxiter=100), the slightest error (e=1e-5), and the initial function objective(P_0=0).

3. For the initial iteration t=1, generate real random numbers u_{ik} where i = 1,2,…,19 and the number of clusters k = 1,2,3 as elements of the initial partition matrix U_0. Matrix U_0 represents the initial membership degrees of each district for each cluster.

$$U_0 = U_{19x8} = \begin{bmatrix}
0,26 & 0,4 & 0,34 \\
0,12 & 0,64 & 0,24 \\
0,53 & 0,12 & 0,35 \\
0,02 & 0,11 & 0,87 \\
0,4 & 0,53 & 0,07 \\
0,87 & 0,1 & 0,03 \\
0,32 & 0,53 & 0,15 \\
0,44 & 0,35 & 0,21 \\
0,2 & 0,29 & 0,51 \\
0,18 & 0,71 & 0,11 \\
0,2 & 0,29 & 0,51 \\
0,68 & 0,26 & 0,06 \\
0,26 & 0,08 & 0,66 \\
0,12 & 0,2 & 0,68 \\
0,55 & 0,33 & 0,12 \\
0,65 & 0,24 & 0,11 \\
0,13 & 0,43 & 0,44 \\
0,69 & 0,11 & 0,2 \\
0,34 & 0,62 & 0,04
\end{bmatrix}$$

4. Calculate the centroid values of each cluster (V) in the first iteration. The formula to compute the cluster centroids is

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue IV March 2024

$$v_{kj} = \frac{\sum_{i=1}^{n}(u_{ik}{}^{w} x_{ij})}{\sum_{i=1}^{n}(u_{ik})^{w}}$$

As an example of calculating the center of the first cluster for variable X_1:

$$v_{11} = \frac{\sum_{i=1}^{n}(u_{i1}{}^{2} x_{i1})}{\sum_{i=1}^{n}(u_{i1})^{2}}$$

$$v_{11} = \frac{(0{,}26)^2(7{,}237461099) + \cdots + (0{,}34)^2(16{,}77833581)}{(0{,}26)^2 + \cdots + (0{,}34)^2}$$

Using the same method, the center values of clusters 1, 2, 3 for each variable x1,…,x8 will be obtained as shown in Table II.

TABLE II: Center Values for First Iteration

Center	X1	X2	X3	X4	X5	X6	X7	X8
1	5,7325	109,378	17,0068	10,2746	128,807	3,2756	76,7129	67,859
2	7,2206	118,523	16,8703	13,4215	163,623	4,1084	77,2978	72,799
3	7,2309	118,691	16,8942	13,4405	163,855	4,1142	77,4074	72,902

5. Calculate the objective function for the first iteration (P_1). Using the equation, you can compute the accurate function value for the first iteration.

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}(x_{ij} - v_{kj})^2\right](u_{ik})^w\right)$$

The calculation of the objective function in the first iteration (P_1)

$$P_t = \sum_{i=1}^{19}\sum_{k=1}^{3}\left(\left[\sum_{j=1}^{8}(x_{ij} - v_{kj})^2\right](u_{ik})^2\right) = 54334{,}09847$$

6. Calculate the new values of the Uik matrix for the first iteration based on the formula in the equation.

$$u_{ik} = \frac{\left[\sum_{j=i}^{m}(x_{ij} - v_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}(x_{ij} - v_{kj})^2\right]^{\frac{-1}{w-1}}}$$

The membership degree value of object 1 in the first iteration is as follows:

$$u_{11} = \frac{\left[\sum_{j=i}^{8}(x_{1j} - v_{1j})^2\right]^{\frac{-1}{2-1}}}{\sum_{k=1}^{3}\left[\sum_{j=1}^{8}(x_{1j} - v_{1j})^2\right]^{\frac{-1}{2-1}}} = 0{,}165879894$$

Applying the same method, we obtain new membership degrees and present them in Table III.

TABLE III: The Degree of the New Partition Matrix

Regency/ City	cluster 1	cluster 2	cluster 3
Pariaman City	0,16588	0,4148	0,41932
Solok Regency	0,39698	0,30162	0,3014
Payakumbuh City	0,58101	0,21006	0,20893
Sijunjung Regency	0,40132	0,30081	0,29787
Tanah Datar Regency	0,45599	0,27234	0,27167
Pasaman Barat Regency	0,09162	0,45299	0,45539
Pasaman Regency	0,12398	0,43458	0,44144
Lima Puluh Kota Regency	0,47376	0,26387	0,26237
Bukittinggi City	0,50171	0,24963	0,24866
Dharmasraya Regency	0,13942	0,4292	0,43138
Pesisir Selatan Regency	0,11586	0,43853	0,44561
Sawahlunto City	0,1214	0,43619	0,4424
Agam Regency	0,50879	0,24614	0,24507
Solok City	0,6309	0,18537	0,18373
Padang Panjang City	0,54961	0,2257	0,22469
Solok Selatan Regency	0,5436	0,22877	0,22763
Padang Pariaman Regency	0,18563	0,40595	0,40843
Padang City	0,43704	0,28199	0,28097
Mentawai Regency	0,28222	0,35881	0,35896

a. Check the stopping condition for t=1 (initial iteration)

The clustering process concludes when the absolute difference between P_t and P_{t-1} is less than 1e-5. As $|P_1 - P_0| = |54334.0985 - 0| = 54334.0985 > 1e\text{-}5$, the fourth step is reiterated until the value of $|P_t - P_{t-1}| < 1e{-}5$.

In this study, using the Matlab program, the iteration process stops at the 88th iteration with the value: $|P_{88} - P_{87}| = |26581, 903000500 - 26581, 903009057| = 0, 000008557 < 1e{-}5$. With the final membership degree values at the 88th iteration In Table IV, the last membership values (U_{akhir}) for the 88th iteration will be displayed as follows:

TABLE IV: The Degree of the Iteration 88

Regency/ City	cluster 1	cluster 2	cluster 3
Pariaman City	0,0513	0,8604	0,08883
Solok Regency	0,3418	0,558	0,1002
Payakumbuh City	0,9619	0,028	0,0101
Sijunjung Regency	0,0975	0,8288	0,0737
Tanah Datar Regency	0,7392	0,1997	0,0611

Pasaman Barat Regency	0,0248	0,1399	0,8352
Pasaman Regency	0,0728	0,5009	0,4263
Lima Puluh Kota Regency	0,2552	0,6556	0,0892
Bukittinggi City	0,8603	0,0956	0,0442
Dharmasraya Regency	0,0056	0,0238	0,9705
Pesisir Selatan Regency	0,0733	0,613	0,3137
Sawahlunto City	0,4765	0,455	0,0684
Agam Regency	0,8667	0,0986	0,0348
Solok City	0,2854	0,5403	0,1743
Padang Panjang City	0,8679	0,1035	0,0287
Solok Selatan Regency	0,0119	0,0338	0,9543
Padang Pariaman Regency	0,0202	0,9537	0,0261
Padang City	0,5081	0,2708	0,2212
Mentawai Regency	0,1607	0,2629	0,5764

U_{akhir} obtained the location of the cluster from each area, as shown in Table V.

TABLE V: Location of the Cluster of Each Data

Regency/ City	cluster 1	cluster 2	cluster 3
Pariaman City		*	
Solok Regency		*	
Payakumbuh City	*		
Sijunjung Regency		*	
Tanah Datar Regency	*		
Pasaman Barat Regency			*
Pasaman Regency		*	
Lima Puluh Kota Regency		*	
Bukittinggi City	*		
Dharmasraya Regency			*
Pesisir Selatan Regency		*	
Sawahlunto City	*		
Agam Regency	*		
Solok City		*	
Padang Panjang City	*		
Solok Selatan Regency			*
Padang Pariaman Regency		*	
Padang City	*		
Mentawai Regency			*

Observation of the clustering results is possible in Table VI

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue IV March 2024

TABLE VI: Result Clustering

Regency/ City	cluster	Total
Payakumbuh City, Tanah Datar Regency, Agam City, Bukittinggi City, Sawahlunto City, Padang City, Padang Panjang City.	1	7
Pariaman City, Solok Regency, Sijunjung Regency, Pasaman Regency, Lima Puluh Kota Regency, Pesisir Selatan Regency, Solok City, Padang pariaman Regency.	2	8
Pasaman Barat Regency, Dharmasraya Regency, Solok Selatan Regency, Mentawai Regency	3	4

Cluster characterization of the variables (factors) contributes to reducing stunting. The tool is employed to determine the level of stunting in a region. Subsequently, the cluster centroid values will be presented in Table VII.

TABLE VII: Cluster Centroid

Center	X1	X2	X3	X4	X5	X6	X7	X8
1	5,3117	125,597	19,3187	14,6299	199,674	3,0117	84,0856	68,122
2	5,9305	133,74	15,8973	13,0988	150,18	4,7894	80,3997	74,0492
3	7,0281	98,7981	15,0462	7,6905	77,9703	3,9661	69,3883	67,8853

Observe the cluster centroid values for each variable in Table VII. three levels of stunting in the region are high, medium, and low. Therefore, for cluster 1, the variables with the highest centroid values are X_3, X_4, X_5, and X_7, while the variables with medium centroid values are X_2 and X_8. Subsequently, the variables with low centroid values are X_1 and X_6 for the other clusters. Similarly obtained as Table VIII.

TABLE VIII: Cluster Characteristics

	cluster 1	cluster 2	cluster 3
high	X3, X4, X5, X7	X2, X6, X8	X1
medium	X2, X8	X1, X3, X4, X5, X7	X6
low	X1, X6		X2, X3, X4, X5, X7,X8

CONCLUSIONS

i. Analyzing the clustering results of regions with factors influencing the reduction of stunting in West Sumatra using the Fuzzy C-Means method reveals that Kota Payakumbuh, Tanah Datar District, Agam City, Bukittinggi City, Sawahlunto City, Padang City, and Padang Panjang City, classified into cluster 1, display a low level of stunting vulnerabilityThe low level of stunting vulnerability in these areas is attributed to four high factors contributing to stunting reduction: the availability of doctors, nutritionists, and midwives per 100,000 population, along with the adequacy of infant health services or the ratio of the number of infants aged 29 days to 11 months receiving health services according to standards. These factors are more abundant compared to other clusters
ii. Regions in cluster 2, namely Pariaman City, Solok District, Sijunjung District, Pasaman District, Lima Puluh Kota District, Pesisir Selatan District, Solok City, and Padang Pariaman District, exhibit amoderate level of stunting vulnerability due to the availability of active integrated health posts andhigh adequacy of health services for toddlers, or a high ratio of the number of children under

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume IX Issue IV March 2024

five years old receiving health services according to standards.

iii. Meanwhile, cluster 3, comprising the regions Pasaman Barat District, Dharmasraya District, Solok Selatan District, and Mentawai Islands District, have a high level of stunting vulnerability because these areas have one high factor contributing to stunting reduction, namely the availability of community health centers) or the ratio of community health centers per 100,000 population.

REFERENCES

1. Annur, Cindy Mutia. (2023). Daftar Prevalensi Balita Stunting di Indonesiapada 2022. https://databoks.katadata.co.id/datapublish/2023/02/02/daftar-prevalensi-balita;stunting-di-indonesia-pada-2022-provinsi-mana-teratas.
2. Annur, Cindy Mutia. (2023). Ini Wilayah denganPrevalensi Stunting Tertinggi di Sumatra Barat pada 2022**.** https://databoks.katadata.co.id/datapublish/2023/02/16/ini-wilayah-dengan-prevalensi-stunting-tertinggi-di-sumatra-barat-pada-2022.
3. Nurmin, D., Hayati, M. N., &Goejantoro, R. (2022). Penerapan Metode Fuzzy C-Means Pada PengelompokanKabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Kesejahteraan Rakyat Tahun 2020. EKSPONENSIAL, 13(2), 189-196.
4. Wijayanti, W., HG, I. R., & YANUAR, F. (2021). Penggunaan Metode Fuzzy C-Means Untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kesehatan Lingkungan. Jurnal Matematika UNAND, 10(1), 129-136.
5. Mulyaningsih, T., Mohanty, I., Widyaningsih, V., Gebremedhin, T. A., Miranti, R., &Wiyono, V. H. (2021). Beyond personal factors: Multilevel determinants of childhood stunting in Indonesia. PloS one, 16(11), e0260265.
6. Christyanti, R. D., Sulaiman, D., Utomo, A. P., & Ayyub, M. (2022). Implementation of Fuzzy C-Means in Clustering Stunting Prone Areas. International Journal of Natural Science and Engineering, 6(3)
7. Setiawan, K. E., Kurniawan, A., Chowanda, A., & Suhartono, D. (2023). Clustering models for hospitals in Jakarta using fuzzy c-means and k-means. Procedia Computer Science, 216, 356-363.