Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 1 (2023) 120 - 128 ISSN Media Electronic: 2580-0760

Optimization Fuzzy Geographically Weighted Clustering with Gravitational Search Algorithm for Factors Analysis Associated with Stunting

Isran K. Hasan[1], Nurwan[2], Nur Falaq[3], Muhammad Rezky Friesta Payu[4]
[1,3,4] Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo
[2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo
[1]isran.hasan@ung.ac.id, [2]nurwan@ung.ac.id, [3]nurfalaq1811@gmail.com, [4]rezky@ung.ac.id

Abstract

Stunting is a significant threat to the quality of human resources in Indonesia because stunting does not only involve physical growth disorders but can also cause children to be vulnerable to disease and experience disorders of brain development and intelligence. Many factors cause stunting, not only malnutrition in pregnant women and toddlers. Grouping can be done to make it easier to see the characteristics of the factors causing stunting in Indonesia. The grouping is done based on the similarity of the characteristics of the factors causing stunting in each province. This study used Fuzzy Geographically Weighted Clustering (FGWC) with Gravitational Search Algorithm (GSA) to group and assess the best cluster using the Partition Coefficient validity index, Classification Entropy, Separation Index, Xie & Beni's Index, and IFV Index. Furthermore, a difference test was conducted to determine the dominant factor causing stunting in the formed cluster. The results showed that the FGWC-GSA gave the best clustering results on the fuzziness value of 2 with the number of clusters 2. Cluster 1 consisted of 16 provinces, and cluster 2 consisted of 18 provinces. Based on the T-test, the variables of infants who received exclusive breastfeeding had significant differences between clusters. Therefore, cluster 2 is a cluster that has dominant problems related to exclusive breastfeeding.

Keywords: fuzzy geographically weighted clustering (FGWC); gravitational search algorithm (GSA); FGWC-GSA; stunting

1. Introduction

Indonesia has a fairly high stunting rate compared to other middle-income countries [1]. Based on the 2021 Indonesian Nutrition Status Study (SSGI) results, Indonesia has a stunting rate of 24.4 percent. Compared with the standards set by the World Health Organization (WHO), Indonesia's stunting rate is still above the standards, namely 20 percent [2].

Stunting is a major threat to the quality of human resources in Indonesia because stunting is not only a matter of disrupting physical growth. Stunting can also result in the vulnerability of children to disease and impaired brain development and intelligence [3]. Clustering, in general, can make it easier to see the characteristics of the factors that cause stunting in Indonesia. So that government programs and policies to reduce stunting rates in Indonesia can be adapted to the characteristics and problems in each region [4].

Cluster analysis is a process of sorting a data set into separate cluster groups, each of which has something in common and aims to group the data into one cluster [5]. Cluster analysis is divided into hard and soft (fuzzy) clustering. Hard clustering is done by maximizing the homogeneity of the observed objects and making them a cluster. Meanwhile, fuzzy clustering is based on fuzzy logic, which assumes that an object can belong to more than one cluster at a certain degree [6]. In real-life applications, there are no clear boundaries between clusters, so fuzzy clustering is the only suitable method for data [7].

Research on the clustering of stunting data was conducted by Seifu Hagos Gebreyesus et al. [8], who conducted a cross-sectional study to evaluate the clustering of stunting factors in the southern Ethiopian area using logistic regression. In addition, Widya Sartika, Suryono, and Adi Wibowo [9] developed an information system to evaluate the factors that cause stunting in real-time using the k-means method. This research uses simple and unmodified clustering methods.

Accepted: 18-09-2022 | Received in revised: 13-01-2023 | Published: 03-02-2023

An interesting modification of clustering is Fuzzy Geographically Weighted Clustering (FGWC). FGWC applies environmental effects to the fuzzy clustering algorithm so that the results will be sensitive to geography [10]. However, FGWC has weaknesses in the initialization process. The initial cluster center selection is made randomly, so it can easily get stuck in the optimal local solution, affecting cluster quality [11]. One of the solutions to overcome FGWC's weaknesses is to use optimization methods to obtain the optimum global solution.

Several studies on FGWC using optimization methods include research conducted by Nasution et al. [12], who developed the FGWC using the Intelligent Firefly Algorithm (IFA) to overcome the weaknesses of the FGWC (IFA-FGWC) with a case study of social vulnerability in Indonesia. The results obtained from IFA-FGWC perform well as a grouping method in a comparative approach with 4 clusters. IFA-FGWC can distinguish social vulnerability characteristics between districts based on the Kruskal-Wallis Test. In addition, research conducted by Pamungkas and Pramana [11] integrated FGWC with the Gravitational Search Algorithm (GSA) optimization method (FGWC-GSA). In this study, the performance of the FGWC-GSA was compared to the standard FGWC and optimization methods of Particle Swarm Optimization, Artificial Bee Colony, and Simulated Annealing in the 2015 Central Java province educational profile case study using the Partition Coefficient (PC) validity index, Classification Entropy (CE), Separation Index (S), Xie Benni's Index, and IFV Index. The results show that the FGWC-GSA outperforms the standard FGWC, and all optimization methods on all validity indices and all clusters and also GSA are proven to overcome the weaknesses of the FGWC.

Based on the description above, researchers are interested in using the method proposed by Pamungkas and Pramana [11], namely FGWC with the GSA optimization method applied to stunting data in Indonesia by adding a comparison with FGWC without GSA and a difference test between clusters on each variable based on Nasution's research. et al. [12] and look at the results of various validity indices. From this study, it is hoped that the FGWC-GSA method can classify provinces in Indonesia according to the factors that cause stunting. By difference tests, the dominant factors that cause stunting in each cluster formed can be seen.

2. Research Methods

2.1 Data

The data used in this study is secondary data obtained from the results of the 2021 Indonesian Nutritional Status Study (SSGI), Publication of the Central Bureau of Statistics 2021, and the website of the Central Bureau of Statistics of the Republic of Indonesia (www.bps.go.id). There are nine variables used in this study, namely Toddlers who Get Exclusive Breastfeeding (X1), Toddlers who Get Early Initiation of Breastfeeding (IMD) (X2), Toddlers who Get Complete Basic Immunization (X3), Toddlers who are given MPASI (X4), Weight Low Birth Weight (LBW) (X5), Households with Access to Proper Sanitation (X6), Households with Access to Adequate Drinking Water (X7), Average Per Capita Expenditure in a Month (X8), and Average Per Capita Calorie Consumption a day (X9).

2.2 Research Process Steps

The stages in this study began with dataset preparation. The data normalization was done, setting clustering parameters and GSA parameters, determining geographical modification parameters, clustering with the Fuzzy Geographically Weighted Clustering – Gravitational Search Algorithm (FGWC-GSA), output clustering results, evaluation of clustering results with validity index, and difference tests were performed between clusters. Figure 1 is the flowchart.

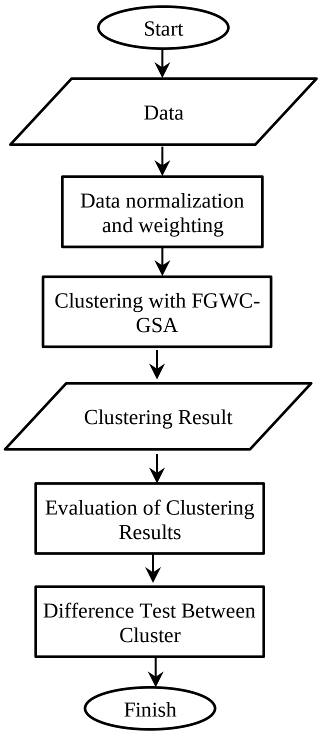

Figure 1. Research Stages

2.3 Normalization

In data analysis, differences in data ranges can affect the analysis results, where data with high values will have a greater influence than data with lower values. Therefore, it is necessary to normalize data [13]. In this study, min-max normalization was used for the data normalization process, and min-max normalization was carried out by Equation (1) [14].

DOI: https://doi.org/10.29207/resti.v7i1.4508

$$x_i' = \frac{x_i - min_{x_k}}{max_{x_k} - min_{x_k}}(new\_max_{x_k} - new\_min_{x_k}) + new\_min_{x_k} \quad (1)$$

Where x_i' is the normalized data value, x_i is the data value, min_{x_k} is the minimum value of the attribute, $max_{x_k}$is the maximum value of the attribute, new_min_{x_k} is the new minimum value of the attribute, and new_max_{x_k} is the new maximum value of the attribute.

2.4 Fuzzy Geographically Weighted Clustering (FGWC)

FGWC is an extension of Fuzzy C-Means, which is more geographically aware because it has added geographical effects in the form of distances between regions and populations, which affect the value of the cluster center. FGWC calculates the influence of an area on another area as the product of the area's population. The distance decay effect is implemented in the divider. The equation is used to determine the adjusted cluster membership for the FGWC algorithm calculated in each iteration of the fuzzy clustering algorithm (2) [10].

$$\mu_i' = \alpha \times \mu_i + \beta \times \frac{1}{A}\sum_j^n w_{ij} \times \mu_i \quad (2)$$

Where μ_i' is the new cluster membership of area i, μ_i is the old cluster membership before entering the spatial effect, A is a factor to scale the "sum" term to the range 0 to 1, and w_{ij} is the weight size. Parameters α and β are scaling variables that influence the membership proportion before and after weighting. The α and β are defined in Equation (3) [10].

$$\alpha + \beta = 1 \quad (3)$$

Weight size is calculated using the equation (4) [10].

$$w_{ij} = \frac{(m_i \times m_j)^b}{d_{ij}^a} \quad (4)$$

Where m_i is the total population of area i, m_j is the total population of area j, and d_{ij} is the distance between area $i$and area j. a and b set the effect of distance and population on weights and are determined by the user [11].

Furthermore, the objective function of FGWC that will be minimized is shown in Equation (5) [15].

$$J_{FGWC}(U, V: X) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \, |v_i - x_k|^2 \quad (5)$$

Where m is the fuzziness, u_{ik} is the element of the membership matrix, v_i is the center of the cluster, and x_k is the data.

The objective function $J_{FGWC}(U, V; X)$ will be minimized by optimizing it through parameters U and V. To find the optimum value of u_ik and v_i. The Lagrange multiplier λ_k is used with the constraint $\sum_{i=1}^{c} u_{ik} = 1$. Furthermore, the Lagrange function for FGWC is derived from each parameter and then equalized to zero to obtain the optimum value of u_{ik} and v_i, resulting in two objective function formulations in Equation (6) and Equation (7).

$$J_{FGWC}(V; X) = \sum_{i=1}^{c}\sum_{k=1}^{n} \frac{|v_i - x_k|^2}{\left(\sum_{j=1}^{c}\left(\left|\frac{v_i - x_k}{v_j - x_k}\right|\right)^{\frac{2}{m-1}}\right)} \quad (6)$$

$$J_{FGWC}(U; X) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \left|\frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m} - x_k\right|^2 \quad (7)$$

These two formulation processes are commonly known as Alternating Optimization (AO), in this case, used to optimize the FGWC model through extreme conditions of the objective function (J_{FGWC}). The two functions of this formulation are referred to as FGWC-V and FGWC-U, where u_{ik} in FGWC-U is a modification of geographic cluster membership [16].

2.5 Gravitational Search Algorithm (GSA)

GSA is a population-based algorithm inspired by Newton's laws of motion and gravity developed by Rashedi, H. Nezambadi-pour, and S. Saryazdi in 2009. GSA aims to enhance the exploration and exploitation of population-based algorithms to achieve optimal solutions [11].

Agents in GSA are considered objects whose performance is measured by their mass [11]. Each of these objects is attracted to each other by the force of gravity, resulting in a global movement of all objects toward the object with a heavier mass. The object's position corresponds to the problem's solution, and the fitness function determines its gravitational and inertial masses. The heavy mass presents the optimal solution [17].

GSA starts by generating an initial population of N agents at random. The agent position is defined in Equation (8).

$$X_i = (x_i^1, \dots, x_i^d, \dots, x_i^n) \quad i = 1,2, \dots, N \quad (8)$$

Where x_i^d is the position of agent i in dimension d, and n is the dimension of the search space.

At time t, a force acts on mass i from mass j. This force is defined in Equation (9).

$$F_{ij}^d = G(t)\frac{M_{pi}(t) \times M_{aj}(t)}{R_{ij} + \varepsilon}\left(x_j^d(t) - x_i^d(t)\right) \quad (9)$$

Where $G(t)$ is the gravitational constant at time t, M_{pi} is the passive gravitational mass of agent i, M_{aj} is the active gravitational mass of agent j, $R_{ij}(t)$ is the euclidean distance between agent i and agent j, and ε is a small constant. Meanwhile, the Euclidean distance of agent i and agent j is defined by $R_{ij}\ (t)$in Equation (10).

$$R_{ij}(t) = \left\|X_i(t), X_j(t)\right\|_2 \quad (10)$$

The net force that acts on mass i in dimension d in time t is defined in Equation (11).

$$F_i^d(t) = \sum_{j\in Kbest, j\neq 1}^{N} rand_j F_{ij}^d(t) \quad (11)$$

where $rand_j$ is a random number in the interval [0,1], $Kbest$ is the first set of K agents with the best fitness value and the largest mass.

The gravitational constant G is initialized at the start of the search and is reduced over time to control search accuracy. The gravitational constant G is defined in Equation (12).

$$G(t) = G(G_0, t) \quad (12)$$

where G is a function of the initial value (G_0) and time (t).

Equation (13),(14),(15) is used to update the gravitational and inertial mass.

$$M_{ai} = M_{pi} = M_{ii} = M_i, \quad i = 1,2, \dots N \quad (13)$$

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \quad (14)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)} \quad (15)$$

Where $fit_i\ (t)$ is the fitness value of agent i at time t, best(t), and worst(t) is determined by the fitness value. For $best(t)$ and $worst(t)$, minimization problems are defined in Equation (16) and Equation (17).

$$best(t) = \min_{j\in\{1,\dots N\}} fit_j(t) \quad (16)$$

$$worst(t) = \max_{j\in\{1,\dots N\}} fit_j(t) \quad (17)$$

The acceleration of agent i at time t in dimension d is defined in Equation (18).

$$a_i^d = \frac{F_i^d(t)}{M_{ii}(t)} \quad (18)$$

where F_i^d is the net force acting on mass i and M_{ii} is the inertial mass of agent i. Calculation of the speed and position of the agent used Equation (19) and Equation (20).

$$v_i^d(t+1) = rand_i \times v_i^d + a_i^d(t) \quad (19)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (20)$$

where $rand_i$ is the uniform random variable in the interval [0,1], $v_i^d(t)$ is the velocity of agent i at time t in dimension d, and $a_i^d(t)$ is the acceleration of agent i at time t in dimension d.

Repeat the steps until the stopping criteria are met. The stopping criterion is fulfilled if it has reached the maximum iteration or the change in fitness value is smaller than the specified threshold value.

2.5 FGWC-GSA

Based on the previous explanation, we describe the steps for processing data using the FGWC-GSA in Figure 2. The clustering parameters used in this study are threshold = 10^{-6} based on research [6], then based on research [15], this study used the maximum iteration $= 100$. This study also used several fuzziness values (m) and the number of clusters, namely $m = 2$ and $m = 2.5$, with many clusters of 2 to 5. The GSA parameters used in this study were based on previous research by [6]. Namely, the initial gravity constant (G) $= 1$ and Vmax $= 0.7$. Then in this study, the number of agents is 34 (according to the number of provinces). Furthermore, the geographic modification parameters (equations (2) and (4)) used are based on research [15], namely α=0.5, β=0.5, a = 1, and b = 1.

2.6. Cluster Validity Index

Measurements commonly used to measure the performance of clustering algorithms are Partition Coefficient (PC), Classification Entropy (CE), Separation Index (SI), Xie and Beni's Index (XB), and IFV Index (IFV) . These measurements will be used as an index of validity in this study. The five validity indices are used to be able to see/assess well the performance of the clustering algorithm to obtain the ideal number of clusters in the clustering. As for the PC and IFV values, the highest indicates an optimal cluster, while CE, SI, and XB indicate the opposite. The lowest value indicates the better the cluster is formed [18].

2.7. Difference Test of clusters

A difference test between clusters for each variable was carried out to see the dominant factors causing stunting in each cluster that was formed. In the difference tests, if the decision to reject H_0 is obtained, there is sufficient evidence to state that the variable has different characteristics between clusters. To find out the dominant problem in each cluster is done by looking at the average variable in each cluster, if this variable increases the stunting rate, then this variable becomes the dominant variable in the cluster with the highest average. If this variable reduces the stunting rate, then this variable becomes the dominant variable in the cluster with the lowest average [12]. If the optimal number of clusters is 2 clusters, then the T-test will be used as a difference test. If the optimal number of clusters is greater than 2, then the one-way ANOVA test will be used as a difference test. In addition, if the variable data used does not meet the assumptions for a difference test with a parametric approach, a difference test will be carried out with a non-parametric approach. If the optimal number of clusters is 2 clusters, then the Mann-Whitney test will be used as a difference test. If the number of optimal clusters exceeds 2, the Kruskall-Wallis test will be used as a difference test.

3. Results and Discussions

3.1. Data Normalization

The research data used consists of several variables with different or varied units. So it is necessary to normalize the data before conducting data analysis.

Data normalization results are shown in Table 1.

Table 1. Normalization Result

Province	X1	X2	X3	...	X9
Aceh	0,1209	0,0000	0,0000	...	0,4538
Bali	0,4257	0,9536	0,9970	...	0,4915
Banten	0,0806	0,2781	0,2874	...	0,5830
⋮	⋮	⋮	⋮	⋱	⋮
North Sumatera	0,1134	0,3709	0,4726	...	0,4217

3.2 Weighting Data

The FGWC algorithm in Figure 2 applies population and distance effects for each iteration. The population used is data on the total population of Indonesia for 2021 obtained from the website of the Central Bureau of Statistics of the Republic of Indonesia.

Meanwhile, the distance matrix was made from the 2019 Indonesian administrative boundary map using a shapefile map format that was processed with R software. The weighting results can be seen in table 2.

Table 2. Distance Matrix Data

Province	Aceh	Bali	Banten	...	Sumatra Utara
Aceh	0,0	2456,03	1562.08	...	327,51
Bali	2456,03	0,0	1018.26	...	2132,88
Banten	1562,08	1018,26	0,0	...	1236,67
⋮	⋮	⋮	⋮	⋱	⋮
North Sumatera	327,51	2132,88	1236,67	...	0,0

3.3 FGWC and FGWC-GSA Performance Evaluation

To see the performance of GSA in optimizing the initial cluster centers, the following shows a comparison between the FGWC and the FGWC-GSA in grouping provinces based on the causes of stunting, which was done with fuzziness $m = 2$ and $m = 2.5$ with several clusters of 2 to 5.

To evaluate the two clustering algorithms, the validity index Partition Coefficient (PC), Classification Entropy (CE), Separation Index (SI), Xie and Beni's Index (XB), and IFV Index were used. The results of the evaluation of the FGWC and FGWC-GSA, which are presented in Table 3.

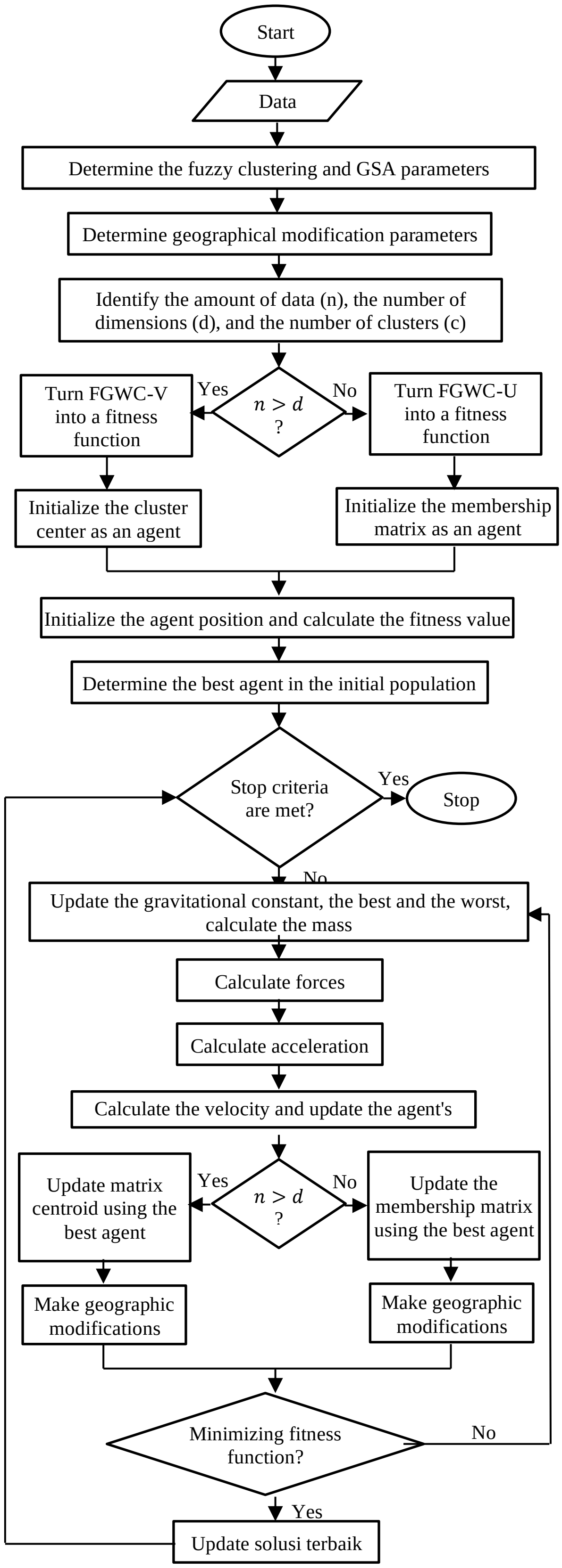

Figure 2. FGWC-GSA Algorithm

Table 3. Comparison of FGWC and FGWC-GSA with various validity indices

m	Cluster	PC		IFV		CE		SI		XB	
		FGWC	FGWC-GSA	FGWC	FGWC-GSA	FGWC	FGWC-GSA	FGWC	FGWC-GSA	FGWC	FGWC-GSA
2	2	0.5000	0.5269	0.00014	0.11544	0.6931	0.6656	7092.657	9.6196	1.738	2.467
	3	0.3334	0.3351	0.00121	0.08189	1.0986	1.0960	67120.88	79.5852	1.161	1.481
	4	0.2500	0.2519	0.00048	0.13374	1.3863	1.3827	127503.7	301.3976	0.869	1.271
	5	0.2000	0.2018	0.00023	0.17645	1.6094	1.6050	297132804	60.1331	0.695	0.920
2,5	2	0.5000	0.5012	0.00002	0.02985	0.6931	0.6919	55663.63	33.5191	1.226	1.406
	3	0.3333	0.3334	0.00004	0.00747	1.0986	1.0985	2238097	1077.1980	0.667	0.688
	4	0.2500	0.2501	0.00002	0.01534	1.3863	1.3861	3088662	3240.4040	0.433	0.466
	5	0.2000	0.2001	0.00002	0.01893	1.6094	1.6093	1564090209	1842.5730	0.310	0.328

Based on the results of the performance evaluation of the FGWC and FGWC-GSA in Table 3, it can be seen that the FGWC-GSA gave the highest PC and IFV index values for all the clusters in the two fuzziness values used. It means that the cluster formed by the FGWC-GSA is more optimal than the FGWC.

Furthermore, the same can be seen in the smaller CE and SI validity index values for all the clusters in the two fuzziness values. However, on the validity index XB, the smallest value is given by FGWC. So, according to the XB index FGWC gives the best results.

Therefore, it was found that FGWC-GSA provided better clustering quality than FGWC, where FGWC-GSA outperformed FGWC on four validity indices, PC, IFV, CE, and SI, in all fuzziness values and all numbers of clusters.

3.4 FGWC-GSA Best Cluster

The FGWC-GSA algorithm was carried out to evaluate the best clusters with several different fuzziness values, namely fuzziness m = 2 and m = 2.5 with a number of clusters of 2 to 5 clusters.

Furthermore, an evaluation was carried out with the Partition Coefficient (PC) validity index, Classification Entropy (CE), Separation Index (SI), Xie and Beni's Index (XB), and IFV Index.

Table 4 shows that the validity indices of PC, CE, and SI show optimal values for cluster 2 with a fuzziness value of 2. In contrast, the IFV validity index shows optimal values for cluster 4 with a fuzziness value of 2, and the XB validity index shows optimal values for the number of clusters 5 with a fuzziness value of 2.5.

Table 4 shows that three of the five validity indexes used, namely the PC, CE, and SI validity indices, show the best cluster quality in the number of clusters 2 with a fuzziness value of 2. So that the number of clusters 2 with a fuzziness value of 2 is the best cluster to be used in forming Province clusters based on factors that cause stunting.

Table 4. FGWC-GSA validity index

m	cluster	PC	IFV	CE	SI	XB
2	2	**0.527**	0.115	**0.666**	**9.620**	2.467
	3	0.335	0.082	1.096	79.585	1.481
	4	0.252	**0.134**	1.383	301.398	1.271
	5	0.202	0.176	1.605	60.133	0.920
2,5	2	0.501	0.030	0.692	33.519	1.406
	3	0.333	0.007	1.098	1077.19	0.688
	4	0.250	0.015	1.386	3240.40	0.466
	5	0.200	0.019	1.609	1842.57	**0.328**

The membership value of the cluster formation results is presented in Table 5.

Table 5. Membership Value

No	Province	Membership Value		Cluster
		1	2	
1	Aceh	0.3122	0.6878	2
2	Bali	0.6340	0.3660	1
3	Banten	0.2376	0.7624	2
4	Bengkulu	0.4602	0.5398	2
5	DI Yogyakarta	0.5114	0.4886	1
6	DKI Jakarta	0.4193	0.5807	2
7	Gorontalo	0.5318	0.4682	1
8	Jambi	0.6975	0.3025	1
9	West Java	0.2884	0.7116	2
10	Central Java	0.3744	0.6256	2
11	East Java	0.6308	0.3692	1
12	West Kalimantan	0.3779	0.6221	2
13	South Kalimantan	0.3672	0.6328	2
14	Central Kalimantan	0.4598	0.5402	2
15	East Kalimantan	0.3522	0.6478	2
16	North Kalimantan	0.5066	0.4934	1
17	Bangka Belitung Islands	0.4950	0.5050	2
18	Riau Islands	0.3858	0.6142	2
19	Lampung	0.5735	0.4265	1
20	Maluku	0.4290	0.5710	2
21	North Maluku	0.5259	0.4741	1
22	Nusa Tenggara Barat	0.5428	0.4572	1
23	Nusa Tenggara Timur	0.5759	0.4241	1
24	Papua	0.5085	0.4915	1
25	West Papua	0.5024	0.4976	1
26	Riau	0.5600	0.4400	1
27	West Sulawesi	0.5369	0.4631	1
28	South Sulawesi	0.6606	0.3394	1

DOI: https://doi.org/10.29207/resti.v7i1.4508

No	Province	Membership Value		Cluster
		1	2	
29	Central Sulawesi	0.5269	0.4731	1
30	South east Sulawesi	0.4259	0.5741	2
31	North Sulawesi	0.4527	0.5473	2
32	West Sumatera	0.3077	0.6923	2
33	South Sumatera	0.2889	0.7111	2
34	North Sumatera	0.4711	0.5289	2

Furthermore, the results of clustering based on membership values in Table 5 can be seen in Table 6.

Table 6. Provincial Clustering Results.

Cluster 1	Cluster 2
Bali	Aceh
DI Yogyakarta	Banten
Gorontalo	Bengkulu
Jambi	DKI Jakarta
East Java	West Java
North Kalimantan	Central Java
Lampung	West Kalimantan
North Maluku	Kalimantan Selatan
Nusa Tenggara Barat	Central Kalimantan
Nusa Tenggara Timur	East Kalimantan
Papua	Bangka Belitung Islands
West Papua	Riau Islands
Riau	Maluku
West Sulawesi	South east Sulawesi
South Sulawesi	North Sulawesi
Central Sulawesi	West Sumatera
	South Sumatera
	North Sumatera

Based on Table 6, it is known that cluster 1 consists of 16 provinces, and cluster 2 consists of 18 provinces. Where cluster 2 has more members than cluster 1. The following figure 3 is a mapping of clustering results.

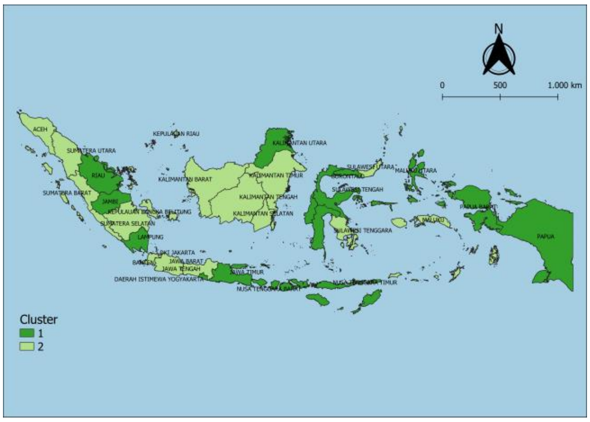

Figure 3. Mapping of Provincial Clustering Results

3.5 Testing of Assumptions

The assumption test performed includes the data normality test and variance homogeneity test. Data normality testing was carried out using the Shapiro-Wilk test.

The results of the data normality test are shown in Table 7.

Based on Table 7, the data normality test on variables X1, X2, X3, X4, X5, and X7 obtained p-value > alpha value (0.05). So accepting H_0, means that the variables X1, X2, X3, X4, X5, and X7 meet the data normality assumption. In the variables X6, X8, and X9, p-values are less than the alpha value (0.05). Then reject H_0, meaning that the variables X6, X8, and X9 do not meet the data normality assumption.

Table 7. Normality Test Results

Variable	Cluster	Significance Shapiro Wilk P-Value	Keputusan
X1	Cluster 1	0,928	H_0 is
	Cluster 2	0,911	Accepted
X2	Cluster 1	0,191	H_0 is
	Cluster 2	0,219	Accepted
X3	Cluster 1	0,192	H_0 is
	Cluster 2	0,249	Accepted
X4	Cluster 1	0,407	H_0 is
	Cluster 2	0,914	Accepted
X5	Cluster 1	0,364	H_0 is
	Cluster 2	0,618	Accepted
X6	Cluster 1	0,001	H_0 is
	Cluster 2	0,312	Rejected
X7	Cluster 1	0,898	H_0 is
	Cluster 2	0,322	Accepted
X8	Cluster 1	0,669	H_0 is
	Cluster 2	0,003	Rejected
X9	Cluster 1	0,018	H_0 is
	Cluster 2	0,471	Rejected

Testing the homogeneity of the variance of the data was carried out using the F test. The results of testing the homogeneity of the variance of the data are shown in Table 8.

Table 8. Variance Homogeneity Test Results

Variable	Significance F-Test P-Value	Keputusan
X1	0,177	H_0 is Accepted
X2	0,312	H_0 is Accepted
X3	0,509	H_0 is Accepted
X4	0,205	H_0 is Accepted
X5	0,076	H_0 is Accepted
X7	0,795	H_0 is Accepted

Based on Table 8, the homogeneity test of variance of the data on the variables X1, X2, X3, X4, X5, and X7 obtained a p-value > alpha value (0.05). So accepting H_0 means that the variables X1, X2, X3, X4, X5, and X7 meet the assumption of homogeneity of variance.

3.6 Difference Test Between Clusters Using the T Test

Based on the results of the assumption test that has been carried out, it is found that the variables X1, X2, X3, X4, X5, and X7 meet the assumptions of data normality and homogeneity of data variance. Then the difference tests on the variables X1, X2, X3, X4, X5, and X7 were carried out using the T-test. The results of the difference tests between clusters using the T-test are shown in Table 9.

Based on Table 9, the difference tests between clusters on variable X1 obtained a p-value of 0.0113 <alpha value (0.05). Then reject H_0, meaning that the variable

X1 has a significant difference or has different characteristics between clusters.

Table 9. T-test Result

Variable	Significance F-Test P-Value	Keputusan
X1	0,0113	H_0 is Rejected
X2	0,4188	H_0 is Accepted
X3	0,2339	H_0 is Accepted
X4	0,0592	H_0 is Accepted
X5	0,5938	H_0 is Accepted
X7	0,2867	H_0 is Accepted

Difference tests between clusters on variables X2, X3, X4, X5 and X7 obtained p-value > alpha value (0.05). Then accept H_0, which means that the variables X2, X3, X4, X5, and X7 have no significant differences between clusters.

Based on the T-test, it is known that the variable X1 (Toddlers Getting Exclusive Breastfeeding) is a variable that has significant differences between clusters. Because the variable Toddlers Getting Exclusive Breastfeeding reduces stunting rates, the cluster with the lowest average in the variable Toddlers Getting Exclusive Breastfeeding has dominant problems related to this variable. So the variable Toddlers Getting Exclusive Breastfeeding becomes the differentiating variable between the formed clusters.

3.7 Difference test Between Clusters Using the Mann-Whitney Test

Based on the results of the assumption test that has been carried out, it is found that the variables X6, X8, and X9 do not meet the data normality assumption. Then the tests on variables X6, X8, and X9 were carried out using the Mann-Whitney test. The results of the difference tests between clusters using the Mann-Whitney test are shown in Table 10.

Table 10. Mann-Whitney Test Result

Variable	Significance Difference test P-Value	Keputusan
X6	0,2695	H_0 is Accepted
X8	0,2373	H_0 is Accepted
X9	0,0878	H_0 is Accepted

Based on Table 10, the difference test between clusters on variables X6, X8, and X9 obtained p-value > alpha value (0.05). Then accept H_0, meaning that the variables X6, X8, and X9 do not significantly differ between clusters.

Based on the difference test using the T-test, the variable Toddlers Getting Exclusive Breastfeeding has a significant difference between clusters. From the grouping results, the average value of the variable under five getting exclusive breastfeeding in cluster 1 is 0.45504, while the average variable under five getting exclusive breastfeeding in cluster 2 is 0.38776. It is known that cluster 2 has a low average on the variable Toddlers Getting Exclusive Breastfeeding, so cluster 2 is a cluster that has a dominant problem related to the variable Toddlers Getting Exclusive Breastfeeding, namely exclusive breastfeeding is still low for toddlers in cluster 2 when compared to cluster 1.

4. Conclusion

The results of the analysis and discussion that have been carried out can be concluded that in the grouping of provinces in Indonesia based on the causes of stunting, the FGWC-GSA performs well on a fuzziness value of 2 with some clusters of 2 based on the validity index Partition Coefficient (PC) and Classification Entropy (CE).

The results of grouping provinces in Indonesia based on the causes of stunting with the FGWC-GSA formed 2 clusters where cluster 1 consisted of 16 provinces, namely Bali, DI Yogyakarta, Gorontalo, Jambi, East Java, North Kalimantan, Lampung, North Maluku, West Nusa Tenggara, Nusa Tenggara East, Papua, West Papua, Riau, West Sulawesi, South Sulawesi, Central Sulawesi. Moreover, cluster 2 consists of 18 provinces, namely Aceh, Banten, Bengkulu, DKI Jakarta, West Java, Central Java, West Kalimantan, South Kalimantan, Central Kalimantan, East Kalimantan, Bangka Belitung Islands, Riau Islands, Maluku, Southeast Sulawesi, North Sulawesi, West Sumatra, South Sumatra, North Sumatra.

The variable of Toddlers Getting Exclusive Breastfeeding indicated significant differences between clusters based on the difference test results. So that the variable Toddlers Getting Exclusive Breastfeeding is the variable that differentiates between the clusters formed. It was found that cluster 2 had dominant problems related to exclusive breastfeeding.

The suggestion that the writer can give for further research is that further research can be carried out regarding the parameters used in the GSA to maximize its performance of the GSA.

Acknowledgment

This research is funded by LPPM-UNG via PNBP-Universitas Negeri Gorontalo according to DIPA-UNG No. 023.17.2.677521/2021, under contract No. B/144/UN47.DI/PT.01.03/2022.

Reference

[1] Badan Pusat Statistik and Kementerian Kesehatan RI, *Laporan Pelaksanaan Integrasi SUSENAS Maret 2019 dan SSGBI Tahun 2019*. Jakarta: Badan Pusat Statistik, 2019.

[2] Kementerian Kesehatan Republik Indonesia, *Buku Saku Hasil Studi Status Gizi Indonesia (SSGI) Tingkat Nasional, Provinsi, dan Kabupaten/Kota Tahun 2021*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2021.

[3] Khairani, “Situasi Stunting di Indonesia,” *Jendela data dan Inf. Kesehat.*, vol. 208, no. 5, pp. 1–34, 2020, [Online]. Available: https://pusdatin.kemkes.go.id/download.php?file=download/pusdatin/buletin/buletin-Situasi-Stunting-di-Indonesia_opt.pdf.

[4] D. Satriawan and D. A. Styawan, “Pengelompokkan Provinsi di Indonesia Berdasarkan Faktor Penyebab Balita Stunting Menggunakan Analisis Cluster Hierarki,” *J. Stat. dan Apl.*, vol. 5, no. 1, pp. 61–70, 2021, doi: https://doi.org/10.21009/JSA.05106.

[5] S. P. Tamba, M. D. Batubara, W. Purba, M. Sihombing, V. M. Mulia Siregar, and J. Banjarnahor, “Book data grouping in libraries using the k-means clustering method,” *J. Phys. Conf. Ser.*, vol. 1230, no. 1, p. 012074, Jul. 2019, doi: 10.1088/1742-6596/1230/1/012074.

[6] R. E. Caraka *et al.*, “Micro, small, and medium enterprises’ business vulnerability cluster in indonesia: An analysis using optimized fuzzy geodemographic clustering,” *Sustain.*, vol. 13, no. 14, 2021, doi: 10.3390/su13147807.

[7] V. K. Malhotra, H. Kaur, and M. A. Alam, “An Analysis of Fuzzy Clustering Methods,” *Int. J. Comput. Appl.*, vol. 94, no. 19, pp. 9–12, 2014, doi: 10.5120/16497-6578.

[8] S. H. Gebreyesus, D. H. Mariam, T. Woldehanna, and B. Lindtjørn, “Local spatial clustering of stunting and wasting among children under the age of 5 years: implications for intervention strategies,” *Public Health Nutr.*, vol. 19, no. 8, pp. 1417–1427, Jun. 2016, doi: 10.1017/S1368980015003377.

[9] W. Sartika, S. Suryono, and A. Wibowo, “Information System for Evaluating Specific Interventions of Stunting Case Using K-means Clustering,” *E3S Web Conf.*, vol. 202, p. 13003, Nov. 2020, doi: 10.1051/e3sconf/202020213003.

[10] G. A. Mason and R. D. Jacobson, “Fuzzy Geographically Weighted Clustering,” *Proceeding 9th Int. Conf. Geocomputation*, pp. 1–7, 2007.

[11] I. H. Pamungkas and S. Pramana, “Improvement Method of Fuzzy Geographically Weighted Clustering using Gravitational Search Algorithm,” *J. Ilmu Komput. dan Inf.*, vol. 11, no. 1, pp. 10–16, Feb. 2018, doi: 10.21609/jiki.v11i1.580.

[12] B. I. Nasution, R. Kurniawan, T. H. Siagian, and A. Fudholi, “Revisiting social vulnerability analysis in Indonesia: An optimized spatial fuzzy clustering approach,” *Int. J. Disaster Risk Reduct.*, vol. 51, 2020, doi: 10.1016/j.ijdrr.2020.101801.

[13] V. Herlinda, D. Darwis, and Dartono, “Analisis Clustering untuk Recredesialing Fasilitas Kesehatan menggunakan Fuzzy C-Means,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, pp. 94–99, 2021.

[14] J. F. Hair JR, W. C. Black, B. J. Babin, and R. E. Anderso, *Multivariat Data Analisys*, Seventh. New York: Prantice Hall International, 2010.

[15] S. N. Abdussamad, S. Astutik, and A. Efendi, “Fuzziness Evaluation on Hybrid Context Based Clustering Methods with Fuzzy Geographically Weighted Clustering-Particle Swarm Optimization Algorithm,” *J. Phys. Conf. Ser.*, vol. 14, no. 1, pp. 10–15, 2021, doi: 10.1088/1742-6596/1811/1/012087.

[16] A. W. Wijayanto and A. Purwarianti, “Improvement of fuzzy geographically weighted clustering using particle swarm optimization,” in *2014 International Conference on Information Technology Systems and Innovation (ICITSI)*, Nov. 2014, pp. 7–12, doi: 10.1109/ICITSI.2014.7048229.

[17] T. Eldos and R. Al Qasim, “On The Performance of the Gravitational Search Algorithm,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 8, pp. 74–78, 2013, doi: 10.14569/ijacsa.2013.040811.

[18] B. Lazzerini, L. C. Jain, and D. Dumitrescu, *Fuzzy Sets & Their Application to Clustering & Training*. USA: CRC Press, 2000.