Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.5, October 2025, pp. 2399~2409
e-ISSN: 2548-6861

Regional Clustering in Sumatera Based on Walfare Indicators Using Fuzzy C-Means

Rahma Dana Putri *, Devni Prima Sari [1]*
* Departement of Mathematics, Universitas Negeri Padang
devniprimasari@fmipa.unp.ac.id

Article Info

Article history:

Received 2025-07-07
Revised 2025-08-08
Accepted 2025-08-10

Keyword:

Welfare,
Sumatera Island,
Data Mining,
Clustering,
Fuzzy C-Means,

ABSTRACT

Welfare refers to a condition in which individuals have sufficient means to meet both physical and spiritual needs. In Indonesia, welfare is a national goal, yet Sumatra experiences the highest development disparity, contributing to unequal welfare distribution across regions. This study aims to cluster regions in Sumatra based on welfare indicators using the Fuzzy C-Means (FCM) method, analyze cluster characteristics, and provide policy recommendations for decision-makers. FCM is used because it accommodates uncertainty and allows each data point to belong to more than one cluster, making it suitable for welfare analysis. Cluster validity was tested using Partition Coefficient Index (PCI) and Silhouette Coefficient, both indicating that the optimal number of clusters is two. The results show that Cluster 1 consists of 62 regions with relatively higher welfare conditions, while Cluster 2 includes 92 regions with lower welfare characteristics. One notable member of Cluster 2 is Ogan Komering Ulu, with a high membership degree of 0.869. Recommended policies include improving access to clean water and healthcare, enhancing education, strengthening local economies, and delivering targeted social assistance to underdeveloped areas. For Cluster 1, sustainable development efforts should be maintained.

I. INTRODUCTION

Welfare is defined as the condition in which individuals in a country have sufficient access to basic needs, supported by social and economic policies [1]. The welfare of the people is one of the main goals of a country, including Indonesia. Issues related to welfare are still a major topic of discussion, especially on the island of Sumatra. Sumatra, Indonesia's third largest island located in the westernmost part of the country, is famous for its abundant natural resources such as petroleum, coal, natural gas, as well as agricultural and plantation products. Although these natural resources contribute significantly to the regional economy and the welfare of its people, the welfare gap across Sumatra remains high and is a serious concern. Ridho et al. (2022) in their study, highlighted that Sumatra has the highest development inequality among all major islands in Indonesia [2]. The economic growth rate of the provinces on the island of Sumatra fluctuated, especially a drastic decline in the provinces of Aceh and Riau [3].

The success of government programs in improving the welfare of its people is the determination of target identifiers, one of the government programs to improve people's welfare is through the Family Hope Program (PKH). This program not only provides assistance in the form of money but also in the form of health and education. However, according to Yulianti in her research explained that the effectiveness of this program target is still low in achieving the right target [4]. This inaccuracy in the targeting of aid further exacerbates inequality and leads to the ineffectiveness of government welfare improvement. This inequality causes difficulties in planning and allocating optimal resources and ineffective policies. To overcome this, an analytical method is needed that can objectively and thoroughly categorize districts/municipalities in Sumatra Island based on indicators of people's welfare. The Statistics Indonesia (Badan Pusat Statistik Indonesia) in 2024 examines the level of people's welfare through 8 fields, namely population, health and nutrition, education, employment, consumption levels and patterns, housing and environment, poverty and other social [5]. All of these fields certainly have a lot of data and variables that are certainly difficult to interpret directly. Taking into account the large and complex data, an analysis is needed that can group districts/cities on the island of Sumatra according to indicators of people's welfare. One of the analyses that can be used to group districts/cities in Sumatra Island is cluster analysis with the Fuzzy C-Means method.

Cluster analysis is one of the methods of data mining, the essence of clustering is that similar data will be grouped into one cluster and data that is not similar is placed in a different cluster [6]. Fuzzy c-means is a clustering algorithm that uses membership values to assign data to more than one cluster. In clustering, fuzzy c-means is the process of minimizing an objective function by updating cluster centers and membership degrees in an iterative process [7]. Bezdek (1981) proposed a fuzzy clustering validity index, namely the Partition coefficient Index [6]. This index is useful for evaluating the degree of membership.

Several previous studies have shown that the fuzzy c-means method is relevant for clustering regions based on certain indicators. Previous research that grouped districts / cities on Kalimantan Island based on indicators of people's welfare in 2020. This research PCI as a measure of validity and in determining the optimal number of clusters [8]. This research provides an overview of the level of people's welfare of Kalimantan Island seen from 18 welfare indicators and using fuzzy c-means method. Another study entitled "Application of the Fuzzy C-Means Clustering Method to Group regencies in East Java Based on People's Welfare Indicators" by Bahtiyar et al in 2024. This research combines fuzzy c-means with silhoutte to validate the cluster [9]. Silhoutte validation is an internal criteria cluster validation, built based on the average closeness of each value on a ratio scale [10].

There are many cluster validations, one of which is the Partition Entropy Index (PEI) according to research entitled "Use of the Fcm Method for Clustering Provinces in Indonesia Based on Environmental Health Indicators". This research was conducted using the fuzzy c-means method and the results obtained using PEI in determining the optimal number of clusters were 0.4633 with the optimal number of clusters as many as 2 clusters. A good PEI value is the value closest to 0 [11].

Research by Susetyo (2023) applied fuzzy c-means in the context of grouping land suitability zones. The use of Partition Coefficient Index (PCI) as validation is able to optimize the number of clusters and show the formation of clusters based on the similarity of land characteristics [12]. This research proves that the FCM method is flexible. Another research by Setiawan et al. combined the K-Means and Fuzzy C-Means methods in an effort to group hospitals in DKI Jakarta based on services and capacity. The validity index used is Partition Coefficient Index (PCI). The results show that combining the two methods can improve the accuracy of clustering [13].

This research employs a data mining approach using the Fuzzy C-Means (FCM) method to analyze the level of welfare in the Sumatra Island region. Welfare is not a binary variable such as "prosperous" or "not prosperous", but rather exists along a continuum without clearly defined boundaries. Therefore, a fuzzy or soft clustering approach is more appropriate for representing this condition compared to hard clustering methods, which are rigid and discrete in nature.

Methods such as K-Means, Hierarchical Clustering, or DBSCAN assign each region to a single cluster with absolute certainty (0 or 1), without accommodating the uncertainty inherent in social data like welfare. In contrast, FCM allows a region to have varying degrees of membership across multiple clusters, effectively capturing the complex and gradual nature of welfare data. This makes FCM a more flexible and representative method for clustering regions based on welfare indicators, which are typically continuous and uncertain in nature. As a result, FCM provides a better reflection of social realities than other clustering techniques.

For these reasons, the Fuzzy C-Means method was chosen in this study, as it is one of the clustering techniques that utilize degrees of membership (fuzziness). This fuzziness is represented linguistically, with linguistic terms functioning as group labels that describe specific [14].

Based on this background, this research aims to group districts and cities on Sumatra Island based on welfare indicators. The study uses nine indicators: population density, life expectancy, average years of schooling, gross regional domestic product per capita at current prices, average monthly per capita expenditure, percentage of households with access to proper drinking water sources, percentage of households that own their homes, percentage of poor people, and the number of beneficiary families.

The grouping is carried out using the Fuzzy C-Means method, and the study also analyzes the characteristics of each resulting cluster. This research is expected to provide accurate and useful data for the government in formulating policies to improve welfare across the Sumatra region.

II. Method

A. Research Stages

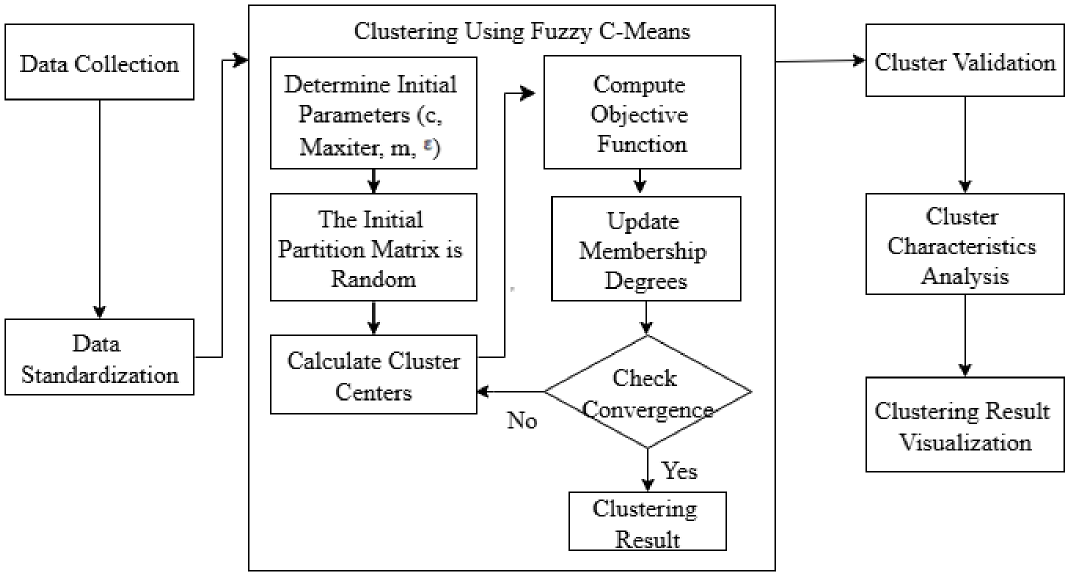

Figure 1 Research Stages

To achieve the research results in accordance with the objectives, this research is arranged through systematic stages starting from data collection, data standardization, clustering with Fuzzy C-Means (FCM), calculating cluster validity, analyzing the characteristics of each cluster and finally visualizing the cluster results into thematic maps. The data used in this research are indicators of people's welfare obtained from Badan Pusat Statistik Indonesia (Statistics Indonesia) for the year 2024. The stages of the research in Figure 1.

B. Indicators of People's Welfare

Welfare is a condition in which basic human needs are fulfilled, including access to adequate housing, clothing, food, education, and healthcare services. From an economic perspective, welfare can also be defined as a state in which individuals are able to maximize their utility within existing budget constraints, while achieving a balanced fulfillment of both physical and spiritual needs [15]. Furthermore, people's welfare refers to a country's capacity to ensure the fair and equitable fulfillment of its citizens' basic needs [16]. Thus, welfare can be understood as a comprehensive condition in which individuals and communities live in a decent, secure, and prosperous state.

According to the Statistics Indonesia [5], changes in the people's welfare are assessed across eight main dimensions: population, health and nutrition, education, employment, consumption levels and patterns, housing, poverty, and other social aspects. Each of these dimensions includes a range of indicators used to measure community welfare. The data used in this research were sourced from the central and provincial offices of Statistics Indonesia (BPS) for each province in Sumatera for the year 2024. However, in this research, the assessment of welfare is focused on nine selected indicators representative and relevant to the context of the study area, as follows:

1) Population Density (people/km²): Population density refers to the number of people per unit of land area and is used as one of the indicators of welfare by the Statistics Indonesia (BPS). The data in this research were obtained from BPS for each province on Sumatra Island in 2024. This indicator has a dual nature: on the one hand, high population density can improve the efficiency of public services such as education, healthcare, and transportation; on the other hand, without proper planning, overcrowding can lead to issues such as social inequality, the growth of slums, and limited access to basic services. Therefore, population density is considered a crucial aspect in assessing welfare. In this study, population density is represented as the independent variable X1.

2) Average Years of Schooling (Years)*:* Data on the average years of schooling were obtained from the Statistics Indonesia in 2024. Education is one of the essential pillars in improving the welfare of society. The Average Years of Schooling is an indicator used to measure the number of years individuals aged 15 and over spend in formal education [5]. A higher average indicates a better-educated population, which in turn opens up broader employment opportunities and provides economic benefits in the form of increased returns. This indicator not only reflects educational attainment but also contributes to the improvement of a region's Human Development Index (HDI). The United Nations Development Programme (UNDP, 2025) also considers the average years of schooling a key indicator for measuring the education dimension and its impact on overall welfare [17]. In this study, the average years of schooling is represented as the independent variable X2.

3) Life Expectancy (Years): Life Expectancy is one of the main indicators in assessing the level of public health and people's welfare in general. Life Expectancy is defined as the average estimated life expectancy that can be achieved by a person since birth, based on mortality conditions that apply at a certain time [5]. In this context, the higher the Life Expectancy Rate, the better the quality of life and health services received by the community. Life expectancy data in this research was obtained from the Statistics Indoensia in 2024. Life expectancy does not only reflect access to health services, but is also closely related to other factors such as adequate nutrition, sanitation, and poverty alleviation programs. In addition to BPS, the World Health Organization (WHO, 2000) and Car (2009) also use Life Expectancy as a representative indicator in measuring the health level of a population[18][19]. In this study, the life expectancy is represented as the independent variable X3.

4) Gross Regional Domestic Product Per Capita at Current Prices (Rupiah): The data in this research were obtained from BPS for each province on Sumatra Island in 2024. Gross Regional Domestic Product per capita at current prices is one of the key indicators used to assess the economic condition of a region within a specific period. It is calculated by dividing the total value of goods and services produced in a region by its population, using prevailing market prices in the current year without adjusting for inflation [5]. This indicator provides an overview of the average economic income of the population in a given area. In this study, GRDP per capita at current prices is represented as the independent variable X4.

5) Monthly Per Capita Expenditure on Food and Non-Food (Rupiah): Household consumption expenditure is one of the key indicators for measuring the level of community welfare. The average monthly per capita expenditure, which includes both food and non-food items, reflects an individual's economic capacity and quality of life in meeting their basic needs. According to Statistics Indonesia (2024), this indicator measures the average monthly spending per person to meet consumption needs. Food expenditure includes staple items such as rice, vegetables, meat, and fish, while non-food expenditure includes needs such as education, housing, transportation, and others. Pape and Ali (2023) emphasize that household expenditure is an essential component in calculating poverty levels and assessing social inequality [20]. In this study, monthly per capita expenditure on food and non-food is represented as the independent variable X5.

6) The Percentage of Household with Acces to Improved Drinking Water Sources:Access to adequate drinking water is a fundamental human right and a critical component of a healthy and productive life [21]. This indicator measures the percentage of households that have access to sources of drinking water that meet health and safety standards, such as piped water, boreholes, protected wells, or springs, as defined by Statistics Indonesia (2024). In this study, access to adequate drinking water is represented as the independent variable X6.

7) Percentage of Households with Owned Housing (%): Owning a home is an important indicator in assessing household welfare and economic stability. According to Alcock et al. (2001), the provision of housing services is a form of government intervention aimed at improving public welfare [22]. Statistics Indonesia (2024) defines this indicator as the percentage of households that own their dwelling independently, rather than renting or living in another person's home. Home ownership reflects economic independence and a household's ability to meet basic shelter needs. It is therefore considered a significant indicator of community welfare and long-term economic security. In this study, the percentage of households with owned housing is represented as the independent variable X7.

8) Percentage of Poor Population (%)**:** Poverty is the antithesis of welfare, referring to a condition in which individuals are unable to meet their basic needs such as food, clothing, housing, education, and other essential services [22]. The percentage of poor population is used to measure the proportion of individuals living below the poverty line relative to the total population in a given area, as defined by Statistics Indonesia (2024). The poverty line is calculated based on the minimum expenditure required to fulfill both food and non-food needs. A higher PPM rate indicates a greater prevalence of poverty and, consequently, a lower level of welfare. In this study, the percentage of poor population is represented as the independent variable X8.

9) Number of Beneficiary Families of Food Social Assistance (Families): According to the World Bank (2015), social assistance programs positively impact community welfare by increasing household consumption and improving access to basic health and education services. These programs also foster social participation and enhance the self-confidence of underprivileged populations.The number of beneficiary families is commonly associated with government-led food assistance programs aimed at improving the well-being of vulnerable groups, such as low-income families, the elderly, persons with disabilities, and other at-risk communities [23].According to Statistics Indonesia (2024), food social assistance programs in Indonesia include initiatives such as the Family Hope Program (PKH) and the Non-Cash Food Assistance (BPNT). In this study, the number of beneficiary families of food social assistance is represented as the independent variable X9.

C. Fuzzy C-Means

FCM is one of the techniques in data mining that belongs to the unsupervised learning category, which is a machine learning process without direct supervision from humans [24][25]. In this method, unlabeled data will be grouped based on the similarity or closeness of the characteristics between the data. FCM is an iterative algorithm, the main purpose of this algorithm is to determine cluster centers that can be used to calculate the degree of membership of each data to each cluster. This process is carried out iteratively until a convergent condition is reached, namely when the change in the value of the cluster center between iterations is below a certain threshold [26]. Fuzzy c-means is a clustering method that groups objects based on the level of closeness measured using a distance measure, each object is not directly categorized strictly into a cluster, but has a degree of membership [24]. There are several types of distances that can be used in fuzzy c-means, including Euclidean distance, Square Euclidean, Manhattan, Canberra and others.

The distance used in this study is the Square Euclidean distance, this distance is the same as the euclidean distance, but does not use the square root [13][27].

$$\mathrm{d} = \sum_{\mathrm{j}}^{\mathrm{p}}\left(\mathrm{X_{ij}} - \mathrm{V_{kj}}\right)^2 \qquad (1)$$

The units of measure used can affect the results of data analysis, to avoid dependence on the choice of units, data should be normalized/standardized Standardization aims to equalize the scale of all variables, so that each has an equal contribution to the analysis [28]. Standardizing a measurement means subtracting the mean from the score and then dividing it by the standard deviation and is also referred to as the Z-score [29]. Thus, all variables will have a mean of zero and a standard deviation of one.

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \qquad (2)$$

The steps in performing clustering using FCM are as follows[10][30] :

1. *Enter standardized data*: data in the form of a matrix of size $n \times p$. $\mathrm{i} = 1,2,\dots,\mathrm{n}$ is the number of data used and $\mathrm{j} = 1,2,\dots,\mathrm{p}$ is the number of variables or indicators used in the study.
2. *Determine the initial parameter value:* number of clusters ($c \geq 2$); Fuzziness ($m > 1$), the parameter that controls level of fuzziness. There is no universal optimal value of *m*, as it depends on the characteristics of the data; however, a value of m = 2 is commonly used in many studies[31]; Maximum Iteration, a higher MaxIter allows more chances for convergence; expected error (ε), the expected error is the difference in the objective function between iterations; initial objective function ($P_0 = 0$), initial iteration ($t = 1$)
3. *Generate random numbers:* μ_{ik}, $i = 1,2,\dots,n$; $k = 1,2,\dots,c$ to be members of the initial partition matrix of size i × k. With the sum of the values of each row in $\mu_{ik} = 1$
4. *Calculate the cluster center:* v_{kj} with *k = 1, 2, …, c* dan *j = 1, 2, …, p*

$$v_{kj} = \frac{\sum_{i=1}^{n}[(\mu_{ik})^m \times x_{ij}]}{\sum_{i=1}^{n}(\mu_{ik})^m} \qquad (3)$$

Description:
v_{kj} is the cluster center

5. *Calculate the objective function value:* the objective function value is calculated using equation

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left\{\left[\sum_{j=1}^{p}\left(x_{ij} - v_{kj}\right)^2\right](\mu_{ik})^m\right\} \qquad (4)$$

6. *Calculate the change in the value of the partition matrix:* the membership degree will be updated every time the iteration runs, the change is calculated using equation

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{p}(x_{ij}-v_{kj})^2\right]^{\frac{-1}{m-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{p}(x_{ij}-v_{kj})^2\right]^{\frac{-1}{m-1}}} \qquad (5)$$

7. *Check the stopping condition*: Repeat steps 4 to 6 until the difference in the objective function is ($|P_t - P_{t-1}| < \varepsilon$) or $t > MaxIter$.

D. Cluster Validation

1) Partition Coefficient Index: The Partition Coefficient Index (PCI) is one of the metrics used to evaluate the quality of fuzzy clustering results. This index is designed to assess the degree of membership of each data point, indicating how well the data has been clustered. Unlike other cluster validity indices that take into account geometric information or data distribution, PCI focuses solely on membership degrees, without considering the distance between data points [32]. The PCI value ranges between 0 and 1, where a higher value indicates more distinct and well-defined clusters. This index is calculated by summing the squares of membership degrees of all data points across each cluster, and then dividing the result by the total number of data points [33].

$$PCI = \frac{\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik})^2}{n} \qquad (6)$$

2) Silhouette Coefficient: The Silhouette Coefficient is a method used to determine the success of a clustering process. The following is the formula for the Silhouette Coefficient:

$$S_i = \frac{b_i - a_i}{max(b_i, a_i)}$$

S_i = Silhouette Score.
b_i = The average dissimilarity (distance) between data point i and all other points in the same cluster,

a_i = The lowest average dissimilarity of data point i to all points in any other cluster, of which i is not a member.

III. Result And Discussion

The method used in the process of clustering districts/cities in Sumatra Island is the Fuzzy C-Means (FCM) method. FCM is used to identify regional groups based on the similarity of welfare indicators in Sumatra Island in 2024. There are 154 districts/cities in Sumatra Island and there are 9 indicators used, namely population density, life expectancy, average years of schooling, gross regional domestic product per capita at current prices, average monthly per capita expenditure, percentage of households that have access to a proper drinking water source, percentage of households owning their own house, percentage of poor people, and number of beneficiary families. The data were processed with the assistance of Rstudio Software.

A. Data Standardization

The data in this study has a fairly large range or scale of data such as the number of beneficiaries and GRDP per capita, so it must be standardized first before performing cluster analysis using FCM. Standardization is done using the help of RStudio. For manual calculations, it can be calculated using Equation (6).

TABLE I
RAW DATA OF REGIONAL WELFARE INDIKATORS

No	Kabupaten/ Kota	X1	X2	…	X9
1	Kota Banda Aceh	1,745	13.10	…	35,670
2	Pidie	144	9.04	⋯	212,538
3	Kota Sabang	356	11.22	⋯	9,008
4	Aceh Utara	233	9,11	⋯	285,833
5	Kota Subulussalam	84	8.43	⋯	32,059
6	Nagan Raya	52	9.25	⋯	56,624
7	Bener Meriah	89	10.13	⋯	22,577
8	Aceh Tengah	50	10.03	⋯	64,051
9	Aceh Timur	81	848	⋯	157,848
…	…	…	…	⋱	…
154	Karimun	283	8.91	⋯	32.376

Manual standardization results can be done like the following formulation. Suppose the object value data standardization is carried out on population density data for the first data on Table I, namely Banda Aceh City. It is known that the raw data = 1.745. average =627.4462 and standard deviation = $1391{,}92$

$$Z_{11} = \frac{x_{11} - \bar{x}_1}{s_1} = \frac{1.745 - 627.4462}{1391{,}92} = 0{,}80289$$

The value of data standardization on population density in Banda Aceh City = 0.80289. The same method is applied to all data. The standardization data can be seen in Table II

TABLE III
STANDARDIZATION RESULT DATA

No.	X1	X2	X3	⋯	X8	X9
1	0.803	3.027	0.92	⋯	-0.629	-0.718
2	-0.347	-0.083	-1.14	⋯	2.014	1.695
3	-0.195	1.587	0.487	⋯	1.104	-1.082
4	-0.283	-0.029	-0.333	⋯	1.451	2.695
5	-0.39	-055	-2.396	⋯	1.513	-0.767
6	-0.413	0.078	-0.129	⋯	1.64	-0.432
7	-0.387	0.752	-0.11	⋯	1.921	-0.897
8	-0.415	0.675	-0.326	⋯	1.033	-0.331
9	-0.393	-0.512	-0.374	⋯	0.804	0.949
10	-0.411	0.729	-0.549	⋯	0.516	-0.367
…	…	…	…	⋱	…	…
154	-0.247	-0.183	0.624	⋯	-0.894	-0.763

B. Clustering Using Fuzzy C-Means

Clustering using the FCM method with the number of clusters c = 2 to c = 7. The number of clusters used aims to determine the highest PCI value of each cluster, which indicates the optimal number of clusters.

1. *Clustering using c = 2:* This stage carried out based on the FCM algorithm as follows:

The initial step before carrying out the clustering process using the fuzzy c-means method, first it is necessary to determine the initial parameter values, determining the initial parameters is very important because it affects the clustering results [24]. The initial parameter values are the number of clusters = 2, fuzziness = 2, MaxIter = 1000, expected error = 10^{-9}, initial objective function value = 0 and $t = 1$.

Next step Generate random numbers for the initial membership degree value, the value ranges from (0 - 1) with the number of each cluster row totaling 1. The initial membership degree with the number of clusters = 2, presented in Table III.

TABLE III
INITIAL MEMBERSHIP DEGREE C = 2

No.	*Cluster* 1	*Cluster* 2
1	0.72134	0.27866
2	0.66902	0.33098
3	0.41694	0.58306
4	0.80281	0.19719
5	0.67892	0.32108
6	0.17286	0.82714
7	0.51252	0.48748
8	0.71604	0.28396
9	0.45899	0.54101
…	…	…
154	0.96774	0.03226

After determining the random number for the initial membership degree, the next step is to determine the cluster center according to Equation (3). The following is an example of calculating the cluster center with c = 2 in iteration 1. The cluster center of iteration 1 is presented in Table IV.

TABLE IV
PUSAT CLUSTER AWAL C = 2

Indicators	*Cluster* 1	*Cluster* 2
X1	0.038	-0.0471
X2	-0.035	-0.0077
X3	-0.0433	0.0537
X4	-0.0653	0.0359
X5	0.0265	0.011
X6	-0.066	0.075
X7	0.0899	-0.0554
X8	0.03322	-0.07226
X9	-0.04751	0.05958

$$v_{11} = \frac{\sum_{i=1}^{154}[(\mu_{i1})^2 \times x_{i1}]}{\sum_{i=1}^{154}(\mu_{i1})^2} = \frac{[(0.72134)^2 \times 0.80289] + \ldots + [(0.96774)^2 \times (-0.24756)]}{(0.72134)^2 + \ldots + (-0.24746)^2} = 0.03797$$

$$\vdots$$

$$v_{29} = \frac{\sum_{i=1}^{154}[(\mu_{i2})^2 \times x_{i9}]}{\sum_{i=1}^{154}(\mu_{i2})^2} = \frac{[(0.27866)^2 \times (-0.71787)] + \ldots + [(0.03226)^2 \times 0.76281]}{(0.72134)^2 + \ldots + (-0.24746)^2} = 0.05958$$

To calculate the objective function value using Equation (5), the calculation is assisted by Rstudio. In iteration 1, the objective function value is 812.4197. The next step is to update the new membership degree or partition matrix using Equation (6). This partition matrix will be updated every iteration until the iteration stops. The latest membership degree is presented in Table V.

TABLE V
NEW MEMBERSHIP DEGREE C = 2

No.	*Cluster* 1	*Cluster* 2
1	0.4903	0.5097
2	0.506	0.4940
3	0.4979	0.5021
4	0.4978	0.5022
5	0.5280	0.472
6	0.5161	0.4839
7	0.53241	0.46759
8	0.54476	0.45524
9	0.51391	0.48609
10	0.51154	0.48846

The last in fuzzy c-means algorithm is to check the stop condition where the iteration process stops if it satisfies $(|P_t - P_{t-1}| < \varepsilon)$ or t > maxIter. This needs to be done to know the convergent condition. Check the first condition as follows:

$|P_1 - P_0| < \varepsilon$

$|812{,}41972 - 0| < 10^{-9}$

$812{,}419721 > 10^{-9}$

In the first iteration, the FCM algorithm has not met the stop condition. At c = 2 the iteration stops at the 45th iteration with an objective function of 682,0694

2. *Clustering Using c = 3 to c = 7:* Clustering for the number of clusters c = 3 to c = 7, the clustering process is carried out with the same steps as in clustering with c = 2, Iteration stops for c = 3 to c = 7, namely the 1000th iteration with consecutive objective function values of (452.2376 337.4892, 269.3552, 224.2831, 192.2375).

C. Determining the Optimal Cluster Using Partition Coefficient Index

To determine the optimal number of clusters, it is necessary to test the Partition Coefficient Index (PCI) value one by one cluster from cluster number 2 to cluster 7. Calculation of the PCI value using Equation (2), the PCI value is presented in Table VI. The greater the partition coefficient index value the better [34]. Based on the partition coefficient index value in Table VI, the optimal number of clusters is the number of clusters with the largest PCI value, namely cluster 2 with a PCI value of 0.55556. Therefore, the grouping of districts/cities in Sumatra Island based on welfare indicators uses a number of clusters of 2 clusters.

TABLE VII
PCI SCORE

Number of Cluster	PCI Score
2	0.55556
3	0.38413
4	0.29424
5	0.23588
6	0.19618
7	0.16769

The evaluation results using the Silhouette Coefficient also support this number of clusters, with the highest silhouette value obtained at two clusters, yielding a coefficient of 0.21. The silhouette coefficient values are presented in Table VII, which shows that the data points are not placed in the wrong clusters, as the silhouette values are above 0. A positive silhouette value indicates that the data are well matched to their assigned clusters.

Therefore, the optimal number of clusters in this study is two, based on the consistent evaluation results from both cluster validity methods.

TABLE VIII
SILHOUETTE SCORE

Cluster	Silhouette Score
2	0.21
3	0.111
4	0.062
5	0.016
6	0.019
7	0.026

D. Optimal Cluster Members

After the iteration of the FCM algorithm stops, the final membership degree is obtained which shows how much each data (district/city) is included in each cluster. The final membership degree in the optimal number of clusters can be seen in Table VIII. Each data will be put into the cluster with the highest degree of membership [30]. The highest membership degree shows the tendency of data to enter the cluster which shows the proximity of the data to the center of the cluster.

TABLE VIIII
FINAL MEMBERSHIP VALUE C = 2

No	District/City	Degree Final Membership		Cluster
		C1	C2	
1	Kota Banda Aceh	0.6535	0.3465	1
2	Pidie	0.345	0.655	2
3	Kota Sabang	0.6386	0.3614	1
4	Aceh Utara	0.3632	0.6368	2
5	Kota Subulussalam	0.3492	0.6508	2
6	Nagan Raya	0.4867	0.5132	2
7	Bener Meriah	0.4377	0.5623	2
8	Aceh Tengah	0.3353	0.6647	2
9	Aceh Timur	0.1883	0.8117	2
10	Aceh Tenggara	0.6535	0.3465	1
…	…	…	…	...
154	Karimun	0.7057	0.2943	1

E. Determining Cluster Characteristics

After clustering the region using fuzzy c-means and obtaining the clustering results. The next step is to analyze the characteristics of the 2 clusters formed. The characteristics of these 2 clusters can be determined based on the centroid value or average center of the cluster against each indicator used. The cluster center values are presented in Table IX.

TABLE IX
FINAL CLUSTER CENTER VALUE C = 2

Indicators	Cluster 1	Cluster 2
X1	0.2311	-0.1944
X2	0.4305	-0.347
X3	0.2425	-0.1828
X4	0.2581	-0.2393
X5	0.4467	-0.3651
X6	0.3239	-0.2646
X7	-0.4014	0.3297
X8	-0.3771	0.3133
X9	-0.1573	0.1168

Based on Table VII, In Fuzzy C-Means (FCM), the average of each variable within a cluster is calculated as the cluster center (centroid). based on the centriod, it can be seen that cluster 1 reflects a district/city area with a higher level of welfare than cluster 2. This can be seen from welfare indicators such as Population Density (X1), Life Expectancy (X2) and Average Years of Schooling (X3) which have higher values than cluster 2. This shows that cluster 1 has better access to health and education. In addition, GRDP per capita at current prices (X4) and Average Monthly Expenditure per Capita (X5) are also higher than cluster 2. These two indicators show the economic strength and higher purchasing power of the community, so this region is more advanced in the economic field. Percentage of Households with Access to Adequate Drinking Water (X6) also shows more adequate infrastructure and identifies that adequate water has an impact on health in the region. Beneficiaries (X9) indicates that the region is not dependent on government assistance.

In contrast, cluster 2 shows the opposite characteristics, reflecting a lower level of welfare, characterized by weak economy, health and education but high percentage of poor people and dependency on government assistance. The percentage of Owned Households (X7) in cluster 2 is higher than cluster 1 which means that cluster 2 has more owned housing. However, this could be because there are more owned houses in the village than in the City but with poor sanitation and water quality. Thus, the difference in characteristics between these two clusters emphasizes the welfare gap between regions, where cluster 1 excels in economic, health and education aspects, while cluster 1 still faces multidimensional challenges such as health, education and poverty.

F. Suggestions for Policymakers

Cluster 2 is defined as a group of regions with lower levels of welfare compared to those in Cluster 1. This cluster comprises 92 regional members. The province with the largest number of regions classified into Cluster 2 is South Sumatra Province, indicating a concentration of lower-welfare areas in this region. Notably, one of the districts with the highest membership degree in Cluster 2 is Ogan Komering Ulu Regency, also located in South Sumatra, with a membership value of 0.869. This high degree of membership signifies that Ogan Komering Ulu is highly representative of the general characteristics within Cluster 2. In contrast, the province with the most regions categorized under Cluster 1 is Riau Islands Province, where only one region, Lingga Regency, falls into Cluster 2. This indicates that the majority of regions in the Riau Islands exhibit relatively higher welfare levels.

Based on the specific characteristics of each cluster, several policy considerations are recommended for regions within Cluster 2, particularly for Ogan Kemiring Ulu in South Sumatra Province:

1) *Improvement of basic infrastructure:* Particularly in terms of access to safe drinking water and healthcare services, to enhance the overall quality of life.

2) *Development of the education sector:* Recognizing that a high rate of home ownership does not necessarily reflect welfare if not accompanied by proper access to clean water.

3) *Strengthening of economic empowerment programs and poverty reduction efforts:* Through more targeted and effective social assistance programs.

Meanwhile, regions classified in Cluster 1 have demonstrated more sustainable development conditions. Therefore, policy efforts should focus on strengthening and maintaining the quality of public services to ensure that the achieved development is sustained and does not regress.

G. Visualization of Cluster Result

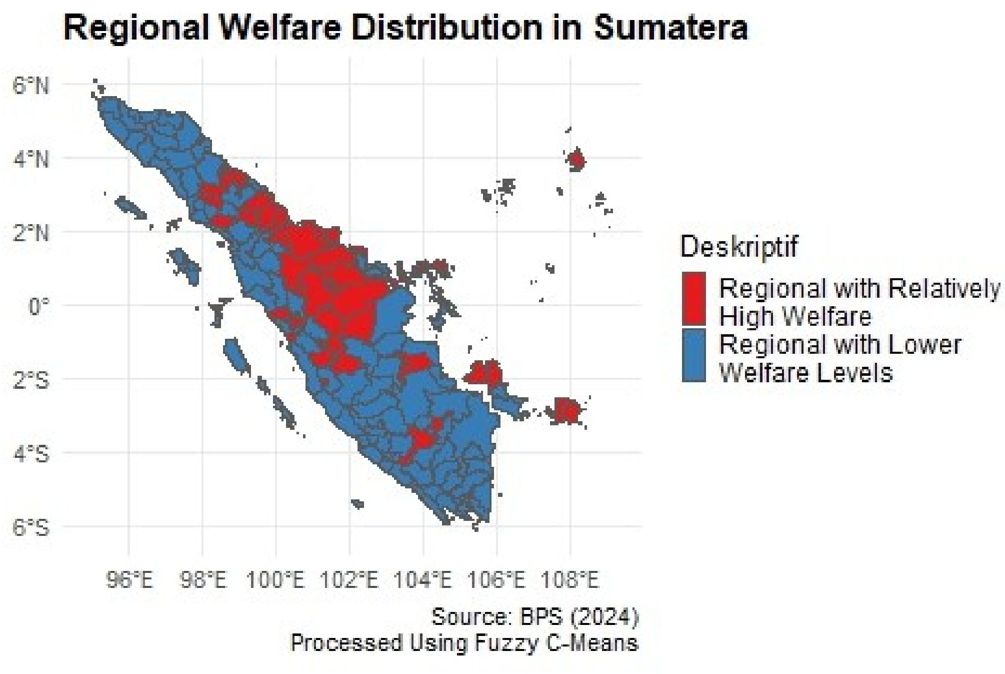

Figure 2 Visualization of Cluster Result

The clusters formed using fuzzy c-means are visualized in the thematic map presented in Figure 1. Figure 1 presents the visualization of the results of clustering regions in Sumatra Island based on welfare indicators using the FCM method. The clustering results produce 2 clusters, the regency/city area that belongs to cluster 1 is shown in red and presents a higher level of welfare compared to cluster 2 which is shown in the blue regency/city area. Ogan Komering Ulu with 86.9% more dominant to cluster 2 is an area of South Sumatra Province with the highest level of low welfare.

IV. CONCLUSION

Based on the results of the analysis using the Fuzzy C-Means (FCM) method and validity testing with the Partition Coefficient Index (PCI) and silhouette coefficient on the grouping of regencies/cities in Sumatra Island in 2024 based on indicators of people's welfare, two optimal clusters were obtained that reflect different levels of welfare. Cluster 1 consists of 62 regencies/cities that show higher welfare levels, characterized by high life expectancy, average years of schooling, GRDP per capita at current prices, per capita expenditure, and access to safe drinking water. The low percentage of poor people and the number of beneficiary families in this cluster illustrate a good level of economic independence. Meanwhile, cluster 2, which consists of 92 kabupaten/cities, has lower welfare characteristics, reflected in weak education, health, and economic indicators, as well as high dependence on government assistance. Although the percentage of owned houses is higher in this cluster, this is more indicative of rural conditions than welfare indicators.

In responding to the different welfare levels in Sumatra Island, the government can improve welfare in cluster 2 such as equal distribution of educational facilities, adding health facilities and health insurance and improving economic infrastructure. Based on these results, the author suggests that future research can use more welfare indicator variables so that the socio-economic conditions of the people on the island of Sumatra can be analyzed more thoroughly. In addition, it is recommended to conduct optimal cluster tests using other approaches and compare clustering results with different methods.

REFERENCES

[1] Irwan, W. Sanusi, and K. Nurani, "Penggunaan Analisis Biplot dalam Pengelompokan Kabupaten/Kota diProvinsi Sulawesi Selatan Berdasarkan Indikator KesejahteraanMasyarakat," *J. Math. Comput. Stat.*, vol. 6, no. 2, pp. 177–187, 2023.

[2] M. Ridho, Z. Zulgani, and E. Achmad, "Analisis ketimpangan pembangunan wilayah antar provinsi di Pulau Sumatera," *e-Jurnal Perspekt. Ekon. dan Pembang. Drh.*, vol. 11, no. 1, pp. 47–58, 2022, doi: 10.22437/pdpd.v11i1.14005.

[3] A. Febriyanda, "Analisis Ketimpangan Pembangunan Ekonomi di Pulau Sumatera," Universitas Islam Negeri Ar-Raniry, 2022. [Online]. Available: https://repository.ar-raniry.ac.id/20754/1/Andri Febriyanda_180604012_FEBI_IE_082363618625.pdf

[4] R. Yulianti, "Efektivitas Program Keluarga Harapan (Pkh) Terhadap Kesejahteraan Masyarakat (Studi Kasus di Desa Bojongmalang Kecamatan Cimaragas Kabupaten Ciamis)," vol. 4, no. 2, pp. 114–121, 2020.

[5] Badan Pusat Statistik Indonesia, "Indikator Kesejahteraan Rakyat 2024," 2024. [Online]. Available: https://www.bps.go.id/id/publication/2024/11/06/3ef10d3d82ed93f616ba9113/indikator-kesejahteraan-rakyat-2024.html

[6] H. Y. Wang, J. S. Wang, and G. Wang, "Multi-fuzzy clustering validity index ensemble: A Dempster-Shafer theory-based parallel and series fusion," *Egypt. Informatics J.*, vol. 24, no. 4, p. 100417, 2023, doi: 10.1016/j.eij.2023.100417.

[7] N. Zamri *et al.*, "Development of Fuzzy C-Means with Fuzzy Chebyshev for genomic clustering solutions addressing cancer issues," *Procedia Comput. Sci.*, vol. 237, no. 2021, pp. 937–944, 2024, doi: 10.1016/j.procs.2024.05.182.

[8] D. Nurmin, M. N. Hayati, and R. Goejantoro, "Penerapan Metode Fuzzy C-Means Pada Pengelompokan Kabupaten / Kota di Pulau Kalimantan Berdasarkan Indikator Kesejahteraan Rakyat Tahun 2020 Application of the Fuzzy C-Means Method in the Grouping of Regencies / Cities in Kalimantan Island Based on People," *J. EKSPONENSIAL*, vol. 13, pp. 189–196, doi: http://dx.doi.org/10.30872/eksponensial.v13i2.1068.

[9] R. Bahtiyar, S. Lestanti, and S. Nur Budiman, "Penerapan Metode Fuzzy C-Means Cluster Ing Untuk Pengelompokan Kabupaten Di Jawa Timur Berdasarkan Indikator Kesejahteraan Rakyat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 5, pp. 3701–3710, 2024, doi: 10.36040/jati.v7i5.7812.

[10] H. Rusnedy, G. W. Nurcahyo, and Sumijan, "Identifikasi Tingkat Pemakaian Obat Menggunakan Metode Fuzzy C-Means," *J. Inf. dan Teknol.*, vol. 3, no. 4, pp. 196–204, 2021, doi: https://doi.org/10.37034/jidt.v3i4.152.

[11] W. Wijayanti, I. Rahmi Hg, and F. Yanuar, "Penggunaan Metode Fuzzy C-Means Untuk Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Kesehatan Lingkungan," *J. Mat. UNAND*, vol. 10, no. 1, p. 129, 2021, doi: 10.25077/jmu.10.1.129-136.2021.

[12] Y. A. Susetyo, "Model Clustering Kesesuaian Lahan menggunakan Kombinasi Algoritma Fuzzy C-Means dan Partition Coefficient Index," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1395–1404, 2023, doi: https://doi.org/10.30865/mib.v7i3.6097.

[13] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, "Clustering models for hospitals in Jakarta using fuzzy c-means and k-means," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 356–363, 2022, doi: 10.1016/j.procs.2022.12.146.

[14] I. Wahyuni, *Logika Fuzzy Tahani*. Komojoyo Press, 2021. [Online]. Available: http://digilib.uinkhas.ac.id/18226/1/Buku Logika Fuzzy Tahani turnitin.pdf

[15] M. P. Todaro and S. C. Smith, *Economic Development*, 9th ed. Pearson, 2006.

[16] B. Prayitno, A. S. Fenat, and P. Mahditia, *Kesejahteraan Rakyat atas Papan: Akselerasi Pemenuhan Kebutuhan Papan*. 2012.

[17] United Nations Development Programme, *Human Development Report 2025*. in Human Development Report. United Nations, 2025. doi: 10.18356/9789211542639.

[18] O. W. Health, "Health Systems: Improving Performance," 2000. [Online]. Available: https://www.who.int/publications/i/item/924156198X

[19] D. Carr, *Encyclopedia of the Life Course and Human Development*, 3rd ed. Cengage Learning, 2009.

[20] U. J. Pape and R. Ali, "Pathways Towards Economic Security - Indonesia Poverty Assessment : Ringkasan Eksekutif (Bahasa (Indonesian))," World Bank, 2023. [Online]. Available: http://documents.worldbank.org/curated/en/099050223030020132

[21] A. Sen, *Development as freedom*. Oxford University Press, 1999. [Online]. Available: https://dokumen.pub/welfare-and-wellbeing-richard-titmusss-contribution-to-social-policy-9781847425355.html

[22] P. Alcock, H. Glennerster, A. Oakley, and A. Sinfield, *Welfare and Wellbeing: Richard Titmuss's Contribution to social Policy*. Policy Press, 2001.

[23] World Bank, "World Development Report 2015: Mind, Society, and Behavior," The World Bank, Dec. 2015. doi: 10.1596/978-1-4648-0342-0.

[24] J. Han, P. Jian, and H. Tong, *Data Mining: Concepts and Techniques*, 4th ed. Cambridge: Morgan Kaufmann, 2022.

[25] F. Nelli, *Python Data Analytics: Data Analysis and science Using Pandas, Matplotlib, and the Python programming language*. New York: Apress, 2015.

[26] N. G. A. Dasriani, M. Mayadi, and A. Anggrawan, "Klasterisasi Lokasi Promosi PMB Dengan Fuzzy C-means Masa Pandemi Covid 19," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 327–336, 2022, doi: 10.30812/matrik.v21i2.1832.

[27] E. Z. Khulaidah and N. Irsalinda, "FCM using squared euclidean distance for e-commerce classification in Indonesia," *J. Phys. Conf. Ser.*, vol. 1613, no. 1, 2020, doi: 10.1088/1742-6596/1613/1/012071.

[28] N. Aini, A. Lestari, M. N. Hayati, F. Deny, and T. Amijaya, "Analisis cluster pada data kategorik dan numerik dengan pendekatan Cluster Ensemble (Studi kasus : puskesmas di Provinsi Kalimantan Timur kondisi Desember 2017)," *J. EKSPONENSIAL Vol. 11*, vol. 11, no. 2, pp. 117–126, 2020.

[29] G. Shmueli, P. C. Bruce, and N. R. Patel, *Data Mining for Business Analytics: Concepts, techniques, and Applications with XLMiner*, 3rd ed. Canada: New Jersey, 2016.

[30] R. Nur'aini, T. Widiharih, and B. A. Saputra, "Pengelompokan Kabupaten/K0ta di Provinsi Jawa Tengah Berdasarkan Indikator Kesehatan Bayi dan Balita Menggunakan Algoritma Fuzzy C-Means dan K-Medoids," *J. Gaussian*, vol. 13, no. 1, pp. 189–198, 2024, doi: https://doi.org/10.14710/j.gauss.13.1.189-198.

[31] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The Fuzzy c-Means Clustering Algorithm," *Comput. Geosci.*, vol. 10, no. 2–3, pp. 191–203, 1984, doi: 00983004849000207.

[32] H. S. Firdaus, A. L. Nugraha, B. Sasmito, and M. Awaluddin, "Perbandingan Metode Fuzzy C-Means Dan K-Means Untuk Pemetaan Daerah Rawan Kriminalitas Di Kota Semarang," *Elipsoida J. Geod. dan Geomatika*, vol. 4, no. 01, pp. 58–64, 2021, doi: 10.14710/elipsoida.2021.9219.

[33] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1981. doi: 10.1007/978-1-4757-0450-1.

[34] F. W. Nugraha, S. Fauziati, and A. E. Permanasari, "Combination of Fuzzy C-Means and Simple Additive Weighting Using Partition Coefficient Index," *Proceeding - 2020 3rd Int. Conf. Vocat. Educ. Electr. Eng. Strength. Framew. Soc. 5.0 through Innov. Educ. Electr. Eng. Informatics Eng. ICVEE 2020*, 2020, doi: 10.1109/ICVEE50212.2020.9243282.