# People's Welfare Clustering Using Fuzzy C-Means with Spatial Information

Ferra Yanuar, Ranti Dwiwanata Putri, Dodi Devianto, Aidinil Zetra, Arrival Rince Putri, Yudiantri Asdi

*Abstract*—**This article aimed to cluster districts and cities in West Sumatra Province based on various indicators of people's welfare using the fuzzy clustering technique. The approach involved both Fuzzy C-Means and Spatial Fuzzy C-Means algorithms. The former assigns data points to groups based on membership degrees, while the latter incorporates spatial information.The analysis revealed that two clusters were optimal. Fuzzy C-Means identified nine regions in Cluster 1 and ten in Cluster 2. Spatial Fuzzy C-Means, on the other hand, assigned eight regions to Cluster 1 and eleven to Cluster 2. Notably, Spatial Fuzzy C-Means outperformed Fuzzy C-Means in terms of Modified Partition Coefficient (MPC) validity index values.In conclusion, leveraging sophisticated algorithms like Spatial Fuzzy C-Means enhances the accuracy of clustering districts and cities in West Sumatra Province based on welfare indicators. This study provides valuable insights for researchers seeking to group regions effectively while considering the nuances of people's well-being.**

*Index Terms*—**Cluster Analysis, Fuzzy C-Means, Spatial Fuzzy C-Means, People's Welfare, Modified Partition Coefficient.**

## I. Introduction

Enhancing the populace's welfare is a fundamental objective of any country's development agenda [1]. Indonesia aspires to provide well-being for all its citizens, including West Sumatra - one of its provinces. To ensure that development programs hit the mark, it is crucial to categorize regions based on their inhabitants' welfare levels. Presently, issues concerning people's welfare still persist in West Sumatra Province. Notably, education, health and employment are not evenly distributed across districts/cities resulting in high poverty rates in some areas [2]. The factors influencing the residents' welfare vary among different districts/cities within West Sumatra Province; however, certain districts or cities may share similar indicators affecting them. As such, clustering can be employed based on these indicators with an aim to formulate policies geared towards enhancing people's welfare through targeted interventions [3].

Furthermore, the development of clustering methods due by considering the value of the degree of membership in the fuzzy set as a basis for weighting, named as fuzzy clustering. The fuzzy clustering method allows an object to become a member of one or more clusters, resulting in a more thorough clustering [4]. There is an object clustering method in fuzzy clustering, namely the Fuzzy C-Means (FCM) method. FCM clustering [5], [6], [7], [8] is an unsupervised technique that has been successfully applied to feature analysis, clustering, and classifier designs in fields such as astronomy, geology, medical imaging, target recognition, and area segmentation. An area can be represented in various feature spaces, and the FCM algorithm classifies the area by grouping similar data points in the feature space into clusters. This clustering is achieved by iteratively minimizing a cost function that is dependent on the distance of the pixels to the cluster centers in the feature domain.

However, the standard FCM does not consider any spatial information in modeling object, which makes it very sensitive to noise, outliers and other imaging object [9], [10]. Recently, many researchers have introduced the local spatial information derived from the object used into the standard FCM to improve the performance of data segmentation [11], [12]. Among them, researchers modified the FCM objective function by incorporating spatial neighborhood term and called the proposed algorithm as Spatial Fuzzy C-Means (sFCM) method [13], [14].

The objective of this study is to cluster the districts/cities in West Sumatra Province based on indicators of people's welfare using FCM and sFCM method. Thereafter, an evaluation will be carried out on the performance of both clusters derived from these methods. The best clustering of all districts/cities in West Sumatra Province is based on indicators of people's welfare. The subsequent sections are structured as follows: Section 2 provides data set used in this study and a comprehensive review of FCM and sFCM, while Section 3 delves into the analysis process and results of this study. Conclusions are presented in Section 4.

Manuscript received August 31, 2023; revised March 8, 2024. This research was funded by Andalas University Research (PGB) Grant No. 21/UN16.19/PT.01.03/IS-PDU-KRP2GB-Unand/2023.

Ferra Yanuar is a lecturer at the Department of Mathematics & Data Sciences, Faculty of Mathematics and Natural Sciences, Andalas University, Kampus Limau Manis, Padang, Indonesia, 25163. (corresponding author to provide e-mail: ferrayanuar@sci.unand.ac.id).

Ranti Dwiwanata Putri is an undergraduate student at the Department of Mathematics & Data Sciences, Faculty of Mathematics and Natural Sciences, Andalas University, Kampus Limau Manis, Padang, Indonesia, 25163. (e-mail: rantidwiwanataputri@gmail.com).

Dodi Devianto is a lecturer at the Department of Mathematics & Data Sciences, Faculty of Mathematics and Natural Sciences, Andalas University, Kampus Limau Manis, Padang, Indonesia, 25163. (e-mail: ddevianto@sci.unand.ac.id).

Aidinil Zetra is a lecturer at Political Science Department, Faculty of Social and Political Sciences, Andalas University, Kampus Limau Manis, Padang, Indonesia, 25163. e-mail: aidinil@soc.unand.ac.id.

Arrival Rince Putri is a lecturer at the Department of Mathematics & Data Sciences, Faculty of Mathematics and Natural Sciences, Andalas University, Kampus Limau Manis, Padang, Indonesia, 25163. (e-mail: arrivalrinceputri@sci.unand.ac.id).

Yudiantri Asdi is a lecturer at the Department of Mathematics & Data Sciences, Faculty of Mathematics and Natural Sciences, Andalas University, Kampus Limau Manis, Padang, Indonesia, 25163. (e-mail: yudiantriasdi@sci.unand.ac.id).

## II. Materials and Methods

### A. *Data Set*

The data in this study were taken from the Central Bureau of Statistics (BPS) of West Sumatra Province regarding Social Affairs and Population. This secondary data is about indicators that affected the welfare of the people of districts/cities in West Sumatra Province in 2022. The indicators used in this study are based on previous researches. Those are the percentage of Human Development Index (HDI) ($X_1$), the percentage of the population with health complaints ($X_2$), the percentage of morbidity ($X_3$), the percentage of women who use birth control ($X_4$), the percentage of women who birth process assisted by health personnel ($X_5$), percentage of literacy rate ($X_6$), the average length of schooling ($X_7$), percentage of Net Enrollment Rate (NER) ($X_8$), percentage of Labor Force Participation Rate (LFPR) ($X_9$), Level Open Unemployment (LOU) ($X_{10}$), percentage of expenditure per capita per month ($X_{11}$), percentage of proper drinking water sources ($X_{12}$), percentage of proper sanitation ($X_{13}$), percentage of lighting sources ($X_{14}$), and percentage of poor people ($X_{15}$). The range of data are between 1 to 100, since all data are in percentage.

The descriptive analysis was then used to report the means and standard deviations for all fifteen variables [15], [16]. Bivariate association analysis was performed using the $\chi^2$ test. To compare means among variables is used Student's t-test for independent samples. This study applies cluster analysis to identify groups of districts/cities where share common welfare based on fifteen factors.

Cluster analysis (CA) is a statistical technique that helps reveal hidden structures by grouping entities or objects (e.g., individuals, products, locations) with similar characteristics into homogenous groups while maximizing heterogeneity across groups [17]. Entities or objects of interest are grouped together based on attributes that make them similar, with the final goal being to distinguish these entities or objects by clustering them into comparable groups and to separate them from differing groups. Conceptually, CA aims to identify cluster solutions that are relatively homogeneous within each group, leading to clusters that have high intra-class similarity, while maximizing heterogeneity between the groups, leading to low inter-class similarity across clusters. Geometrically, the objects within a cluster are close together, while the distance between clusters is further apart. CA is useful to identify groups when it is not clear which entity belongs to which group, and how many groups may best be used to cluster the entities; thus, CA helps to identify a latent structure within a dataset [18], [19]. Differences between clusters are measured using a distance system. The distance measure that will be used in this study is the Euclidean distance. The following formula defines the Euclidean distance [20]:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2}, \tag{1}$$

where $d_{ij}$ is the distance between the $i^{th}$ object and the $j^{th}$ object, $x_{ik}$ is the value of the $i^{th}$ object in the $k^{th}$variable, $x_{jk}$ is the value of the $j^{th}$ object in the $k^{th}$ variable, and $p$ is the number of observed variables. Then, data is grouped into some possible clusters using Principal Component Analysis (PCA) method. Each cluster is analyzed using FCM and sFCM algorithm. The optimum cluster then is determined based on the highest value of Modified Partition Coefficient (MPC).

### B. *Principal Component Analysis (PCA)*

PCA is a data compilation technique that aims to reduce the data dimensions by maintaining as much diversity as possible in the data set. PCA steps were as follows [21], [22], [23]:

1. Data standardization.

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{s^2{}_j}}, \tag{2}$$

where $Z_{ij}$ was the standard variable value for the *i*-th row data, *j*-th column, $x_{ij}$ was the *i*-th data of the *i*-th column, $\bar{x}_j$ was the average value of the *j*-th variable, and $\sqrt{s^2{}_j}$ was the standard deviation of the *j*-th variable.

2. Developing a correlation matrix, $R_{ij}$.

$$R_{ij} = \frac{s_{ij}}{\sqrt{s^2{}_i}\sqrt{s^2{}_j}}, \tag{3}$$

where $R_{ij}$ was the correlation between the *i*-th and *j*-th variables, $s_{ij}$ was the sample covariance of the *i*-th variable with the *j*-th variable, $s^2{}_i$ was the variance of the *i*-th variable, and $s^2{}_j$ was the variance the *j*-th variable.

3. Calculating the eigen values.
The eigen values were obtained from the correlation matrix obtained. Eigen values can be calculated using the following formula

$$det\,(R - \lambda I) = 0, \tag{4}$$

where $R$ was the correlation matrix, $\lambda$ was the eigen value, and $I$ was the identity matrix. The eigen values obtained will be used to calculate the cumulative diversity and diversity values of each main component. The percentage of diversity explained by each component was:

$$diversity = \frac{\lambda_j}{\sum_{j=1}^{m} \lambda_j}, \tag{5}$$

where $\lambda_j$ was the eigenvalue of the *j*-th variable. If the number of components taken was as many as $h$ components, with $h \le j$, then the percentage of cumulative diversity explained was

$$Cumulative\ diversity = \frac{\lambda_1 + \lambda_2 + \lambda_{3+\ldots+}\lambda_h}{(\sum_{j=1}^{m} \lambda_j)} \times 100\%. \tag{6}$$

The main components that obtain an eigenvalue of more than 1 and obtain a cumulative percentage value of not less than 60% will be selected as the number of main components used.

### C. *Fuzzy C-Means (FCM) Method*

The algorithm of the FCM method was as follows [7], [8], [24]:

1. Preparing the data to be clustered $X_{ij}$ with size $n \times m$, where $n$ is the number of data sample, $m$ is the attribute of

each data, and $X_{ij}$ is the $i$-th sample data ($i = 1,2, ..., n$), the $j$-th attribute ($j = 1,2, ..., m$).

2. Determining the number of clusters ($c$), the value of the weighted power ($w > 1$), the smallest expected error ($\varepsilon$), the initial objective function ( $P_0 = 0$ ), and the initial iteration ( $t = 1$ ). The optimum weighting power value that is often used is 2.
3. Generating random numbers $\mu_{ik}, i = 1,2, \dots, n; k = 1,2, ..., c$; as an element of the initial martition matrix U. Elements of the partition matrix are membership values of fuzzy sets that satisfy the following conditions:
$\mu_{ik} \in [0,1]$, dimana $\sum_{i=1}^{n} \mu_{ik} = 1$.
4. Calculating the $k$-th ($V_{kj}$) cluster centre:
$$\boldsymbol{V_{kj}} = \frac{\sum_{i=1}^{n} (\mu_{ik})^{w} (X_{ij})}{\sum_{i=1}^{n} (\mu_{ik})^{w}}. \quad (7)$$
5. Calculating the objective function in the $t$-th ($P_t$)
$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left( \left[ \sum_{j=1}^{m} \left( X_{ij} - V_{kj} \right)^2 \right] (\mu_{ik})^{w} \right).$$
6. The degree of membership obtained can be seen from the tendency of regions to be in specific clusters. The most considerable degree of membership value indicates that the area is a cluster member. Table 3 shows that the Cluster 1 consists of nine regions, those are South Pesisir, Dharmasraya, Padang, Solok, Sawahlunto, Padang Panjang, Bukittinggi, Payakumbuh, and Pariaman. The Cluster 2 consists of ten regions: Mentawai Islands, Solok, Sijunjung, Tanah Datar, Padang Pariaman, Agam, Lima Puluh Kota, Pasaman, South Solok, and West Pasaman. Calculating the changes in the partition matrix elements $\mu^{*}_{ik}$.
$$\mu^{*}_{ik} = \frac{\left( \left[ \sum_{j=1}^{m} \left( X_{ij} - V_{kj} \right)^2 \right] \right)^{\frac{-1}{w-1}}}{\sum_{k=1}^{c} \left( \left[ \sum_{j=1}^{m} \left( X_{ij} - V_{kj} \right)^2 \right] \right)^{\frac{-1}{w-1}}}.$$
7. Checking the stop conditions, if $|P_t - P_{t-1}| < \varepsilon$ then stop, if not then $t = t + 1$ and repeat back to step 4.

*D. Spatial Fuzzy C-Means (sFCM) Method*

This method is a development of the FCM method by utilizing spatial object information in calculating the degree of membership value. The experimental results of Chuang et al. [5] stated that the best results were using the sFCM method with the non-spatial control parameter $p$ and the spatial control parameter $q$ having the same value. Based on the experimental results by Ali, El Abbassi, & Bouattane [25], they proved that the optimum number of clusters obtained from the FCM process was used to determine the $p$ and $q$ parameter values that would produce the best clusters in the sFCM method. The algorithm of the sFCM method was as follows [6]:

1. Preparing the data to be clustered, $X_{ij}$.
2. Determining the number of clusters ($c$), the value of the weighted power ($w > 1$), the smallest expected error ($\varepsilon$), the initial objective function ($P_0 = 0$), the initial iteration ($t = 1$), non-spatial control parameters ($p$), and spatial control parameter ($q$).
3. Generating random numbers, $\mu_{ik}$.
4. Calculating the $k$-th cluster center ($V_{kj}$).
5. Calculating the changes in the elements of the partition matrix $\mu^{*}_{ik}$.
6. Computing spatial functions ($h_{kj}$)
$$h_{kj} = \sum_{i \in NB(x_j)} \mu^{*}_{ik}. \quad (8)$$
7. Calculating the change in the elements of the partition matrix from $\mu^{*}_{ik}$ to $\mu'_{ik}$
$$\mu'_{ik} = \frac{\mu^{*p}_{ik} h^{q}_{kj}}{\sum_{k=1}^{c} (\mu^{*p}_{ik} h^{q}_{kj})}. \quad (9)$$
8. Calculating the objective function in the $t$-th ($P_t$)
$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left( \left[ \sum_{j=1}^{m} \left( X_{ij} - V_{kj} \right)^2 \right] \left( \mu'_{ik} \right)^{w} \right).$$
9. Checking the stop condition, if $|P_t - P_{t-1}| < \varepsilon$ then stop, if not then $t = t + 1$ and repeat back to step 4.

*E. Modified Partition Coefficient (MPC)*

MPC is a validity index method that involves a membership value which is an improved validity index of the Partition Coefficient (PC) method. The MPC value was defined by the following equation:
$$MPC = 1 - \frac{c}{c-1}(1 - PC), \quad (10)$$
as for the PC equation that was:
$$PC = \frac{1}{n} \sum_{k=1}^{c} \sum_{i=1}^{n} (\mu_{ik})^2. \quad (11)$$
Optimum cluster determination of the number of clusters produced is determined from the largest MPC value.

## III. RESULTS AND DISCUSSION

In this section, we shall endeavor to deploy the Fuzzy C-Means (FCM) and Spatial Fuzzy C-Means (sFCM) methods in order to cluster data pertaining to the welfare of individuals residing in districts and cities within West Sumatra Province for the year 2022. The primary objective of these techniques is to discern patterns and relationships embedded within the data that can facilitate a more comprehensive understanding of the current state of welfare prevalent throughout the region.

To achieve this end, we shall duly adhere to a series of requisite steps for each method. These include pre-processing the data, defining appropriate parameters, initializing membership values, computing cluster centers, updating membership values, and repeating these procedures until convergence is attained. By employing these advanced methodologies, our hope is that we will be able to gain invaluable insights into the overall welfare landscape characterizing West Sumatra Province while also identifying potential areas for improvement or intervention.

*A. Defining Multiple Clusters*

Number of clusters that will be used in this study can be seen from the results of the Principle Component Analysis (PCA). The result of PCA is presented in the Table 1.

TABLE 1.
PRINCIPLE COMPONENT ANALYSIS (PCA)

| Principle Component | Eigen Value | Diversity (%) | Cumulative Diversity (%) |
|---|---|---|---|
| **1** | **7.7280** | **51.5** | **51.5** |
| **2** | **2.3809** | **15.9** | **67.4** |
| **3** | **1.5124** | **10.1** | **77.5** |
| **4** | **1.1823** | **7.9** | **85.4** |
| 5 | 0.6460 | 4.3 | 89.7 |
| 6 | 0.5433 | 3.6 | 93.3 |
| 7 | 0.3246 | 2.2 | 95.4 |
| 8 | 0.2943 | 2.0 | 97.4 |
| 9 | 0.1578 | 1.1 | 98.5 |
| 10 | 0.0817 | 0.5 | 99.0 |
| 11 | 0.0667 | 0.4 | 99.5 |
| 12 | 0.0428 | 0.3 | 99.7 |
| 13 | 0.0310 | 0.2 | 99.9 |
| 14 | 0.0064 | 0.0 | 100.0 |
| 15 | 0.0018 | 0.0 | 100.0 |

TABLE 2.
MPC VALUE

| Clusterization Method | Number of Clusters | MPC Value |
|---|---|---|
| FCM | **2** | **0.380095412** |
| | 3 | 0.317232435 |
| | 4 | 0.302489159 |
| sFCM | **2** | **0.454651895** |
| | 3 | 0.388283183 |
| | 4 | 0.352730278 |

TABLE 3.
CLUSTERING MEMBERSHIP DEGREE VALUE USING FCM METHOD

| Districs / City | Cluster 1 | Cluster 2 |
|---|---|---|
| Kepulauan Mentawai | 0.42234 | **0.57766** |
| South Pesisir | **0.50332** | 0.49668 |
| Solok | 0.08705 | **0.91295** |
| Sijunjung | 0.33866 | **0.66134** |
| Tanah Datar | 0.18066 | **0.81934** |
| Padang Pariaman | 0.28913 | **0.71087** |
| Agam | 0.39162 | **0.60838** |
| Lima Puluh Kota | 0.09944 | **0.90056** |
| Pasaman | 0.17813 | **0.82187** |
| South Solok | 0.19682 | **0.80318** |
| Dharmasraya | **0.69703** | 0.30297 |
| West Pasaman | 0.14166 | **0.85834** |
| Padang | **0.85355** | 0.14645 |
| Solok | **0.88525** | 0.11475 |
| Sawahlunto | **0.91211** | 0.08789 |
| Padang Panjang | **0.73278** | 0.26722 |
| Bukittinggi | **0.89084** | 0.10916 |
| Payakumbuh | **0.73890** | 0.26110 |
| Pariaman | **0.96282** | 0.03718 |

TABLE 4.
CLUSTERING MEMBERSHIP DEGREE VALUE USING SFCM METHOD

| Districs / City | Cluster 1 | Cluster 2 |
|---|---|---|
| Kepulauan Mentawai | 0.41082 | **0.58918** |
| South Pesisir | 0.48914 | **0.51086** |
| Solok | 0.02222 | **0.97778** |
| Sijunjung | 0.29573 | **0.70427** |
| Tanah Datar | 0.17666 | **0.82334** |
| Padang Pariaman | 0.25978 | **0.74022** |
| Agam | 0.38323 | **0.61677** |
| Lima Puluh Kota | 0.03358 | **0.96642** |
| Pasaman | 0.11275 | **0.88725** |
| South Solok | 0.15499 | **0.84501** |
| Dharmasraya | **0.67977** | 0.32023 |
| West Pasaman | 0.16616 | **0.83384** |
| Padang | **0.92177** | 0.07823 |
| Solok | **0.94140** | 0.05860 |
| Sawahlunto | **0.92041** | 0.07959 |
| Padang Panjang | **0.70155** | 0.29845 |
| Bukittinggi | **0.96640** | 0.03360 |
| Payakumbuh | **0.75006** | 0.24994 |
| Pariaman | **0.96396** | 0.03604 |

In Table 1 can be seen that the four principal components obtained an eigenvalue of more than 1 with a cumulative percentage of diversity obtained of 85.4%. So, clustering was carried out as many as two to four clusters in this study.

*B. Optimum Number of Clusters*

To determine the optimal number of clusters, it is crucial to take into account the MPC validity index. This index offers valuable insights into the efficacy of various clustering techniques and aids in identifying the ideal cluster count. To gain a deeper understanding of this notion, we can scrutinize Table 2's presented outcomes. The table outlines both FCM and sFCM methods' MPC validity index values for all conceivable cluster configurations.

Upon meticulous examination, it becomes evident that two clusters consistently yielded the highest MPC index values. Specifically, two-cluster setups using either FCM or sFCM techniques produced polished MPC index values of 0.380095412 and 0.454651895 respectively, indicating their effectiveness. As such, based on these findings, we can conclude that utilizing two clusters would likely result in superior outcomes when employing clustering methodologies with the MPC validity index as a performance evaluation benchmark.

*C. Results of Fuzzy C-Means (FCM) Clustering Method*

To ascertain the optimal cluster affiliation, one must assess the level of cluster membership attained in the concluding iteration. This pertinent data is readily available in Table 3, which showcases the degree value of clustering membership ascertained via employment of FCM methodology. Scrutinizing this information with precision is crucial to ensure sound and precise deductions are drawn from the clustering process. By conscientiously considering this pivotal data, researchers can make judicious determinations regarding cluster membership and further comprehend how various variables influence overall clustering outcomes.

*D. Results of Spatial Fuzzy C-Means (sFCM) Clustering Method.*

The culmination of this particular methodology results in a definitive value for the degree of membership. This noteworthy piece of data can be located within the designated Table 4, which exhibits the result of this procedure.

Upon scrutinizing Table 4, it was discovered that the clustering outcomes derived from the FCM technique exhibited a minor fluctuation. Notably, there was one specific region which pertained to a distinct cluster. In contrast to the FCM approach, Pesisir Selatan was allocated to Cluster 2 instead of Cluster 1 by means of the sFCM methodology. This deviation can be attributed to the fact that during analysis, sFCM computation took into account both spatial and non-spatial parameters. The incorporation of these supplementary aspects led to a more nuanced and refined clustering outcome, underscoring the significance of considering multiple variables when executing data analysis processes.

*E. Characteristics of Clustering Results*

After the cluster was formed, the average of each indicator of people's welfare will be taken, the average value of the nineteen districts/cities per indicator was denoted by $\overline{X}_j$. Then for each cluster an average for all districts/cities per indicator of people's welfare will also be taken, the average value per indicator for each cluster was denoted by $\overline{X}_{jc}$ where $j$ = 1,2,…,3 and $c$ = 1,2. The characteristics of the clustering results will be determined by marking each indicator in the cluster, if the value was $\overline{X}_{jc} > \overline{X}_j$ then it was given a positive sign (+) and if the value was $\overline{X}_{jc} < \overline{X}_j$ then it was given a negative sign (-).

In this study, the HDI indicator ($X_1$), women use birth control ($X_4$), birth assisted by health personnel ($X_5$), literacy rate ($X_6$), the average length of schooling ($X_7$), APM ($X_8$), expenditure per capita per month ($X_{11}$), proper drinking water ($X_{12}$), proper sanitation ($X_{13}$), and lighting sourced from PLN ($X_{14}$). If it was positive, it can be said that the level of welfare was relatively better in that cluster. Meanwhile, for the indicators of health complaints ($X_2$), morbidity ($X_3$), TPAK ($X_9$), TPT ($X_{10}$), and poverty ($X_{15}$), if they are opposing, it can be said that the level of welfare was relatively better in that cluster.

In order for two different methods to produce the same clustering characteristics, it is important to analyze the characteristics of each cluster on people's welfare indicators. This analysis can be seen in Table 5, which outlines the clustering results. Specifically, in the first cluster, success was achieved for people's welfare in indicators $X_1, X_4, X_5, X_6, X_7, X_8, X_{11}, X_{12}, X_{13}$ and $X_{14}$. However, there were still several indicators that did not achieve success within this cluster: namely indicators $X_2, X_3, X_9, X_{10}$ dan $X_{15}$.

On the other hand, within the second cluster there were different indicators that could be said to be good. Specifically, indicators $X_2, X_4, X_{10}, X_{11}$ dan $X_{12}$ all showed positive results for people's welfare. However, there were still number of indicators that did not reach welfare figures within this cluster including: indicators $X_1, X_3, X_5, X_6, X_7, X_8, X_9, X_{13}, X_{14}$ and final indicator $X_{15}$. Overall it is clear that careful analysis of these clusters is necessary to ensure that both methods are producing similar clustering characteristics with respect to people's welfare indicators.

## IV. Conclusions

The primary aim of this investigation was to apply fuzzy clustering techniques for the purpose of grouping districts and cities in West Sumatra Province based on their residents' welfare indicators. Two distinct types of fuzzy clustering methods, namely Fuzzy C-Means (FCM) and Spatial Fuzzy C-Means (sFCM), were employed in order to achieve the desired outcomes. The superior cluster was determined by utilizing the Modified Partition Coefficient (MPC) validity index calculation with a total number of clusters set at 2. This comprehensive analysis allowed for a more nuanced understanding of the economic and social landscape of West Sumatra Province, which can inform future policy decisions aimed at improving the overall welfare and well-being of its citizens.

In the FCM method, the Cluster 1 consisted of nine regions including South Pesisir, Dharmasraya, Padang, Solok, Sawahlunto, Padang Panjang, Bukittinggi, Payakumbuh, and Pariaman. The Cluster 2 consisted of ten regions including Mentawai Islands, Solok, Sijunjung, Tanah Datar, Padang Pariaman, Agam, Lima Puluh Kota, Pasaman, South Solok and West Pasaman.

The sFCM approach demonstrated a disparity in the positioning of clustering outcomes as compared to the FCM method. Notably, South Pesisir was incorporated in Cluster 2 instead of Cluster 1. Consequently, sFCM categorized eight regions under Cluster 1 and eleven regions under Cluster 2. These results provide valuable perceptions into district/city clustering techniques that policymakers can utilize when devising regional development strategies. It is crucial for decision-makers to take such research into account while making informed judgments concerning regional development policies that have extensive implications for both societies and economies.

## REFERENCES


[1] A. Aitken, "Measuring Welfare Beyond GDP," *Natl. Inst. econ. rev.*, vol. 249, pp. R3–R16, Aug. 2019, doi: 10.1177/002795011924900110.

[2] R. Sabri, A. Y. S. Hamid, J. Sahar, and Besral, "The effect of culture-based interventions on satisfaction and quality of life of elderly at social welfare institution in West Sumatera," *Enfermería Clínica*, vol. 29, pp. 619–624, Sep. 2019, doi: 10.1016/j.enfcli.2019.06.010.

[3] F. Yanuar, M. Q. Shobri, R. F. Mabrur, I. H. Putri, and A. Zetra, "Classification of Death Risk for COVID-19 Patients Using Bayesian Logistic Regression and Naive Bayes Classifier," *IAENG International Journal of Computer Science*, vol. 50, no. 3, pp. 1–6, 2023.

[4] V. Torra, Y. Endo, and S. Miyamoto, "Computationally intensive parameter selection for clustering algorithms: The case of fuzzy c-means with tolerance," *Int. J. Intell. Syst.*, vol. 26, no. 4, pp. 313–322, Apr. 2011, doi: 10.1002/int.20467.

[5] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, vol. 30, no. 1, pp. 9–15, Jan. 2006, doi: 10.1016/j.compmedimag.2005.10.001.

[6] F. Di Martino and S. Sessa, "The extended fuzzy C-means algorithm for hotspots in spatio-temporal GIS," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11829–11836, Sep. 2011, doi: 10.1016/j.eswa.2011.03.071.

[7] A. A. Hassan, W. Md Shah, M. F. Iskandar Othman, and H. A. Hussien Hassan, "Evaluate the performance of K-Means and the fuzzy C-Means algorithms to formation balanced clusters in wireless sensor networks," *IJECE*, vol. 10, no. 2, pp. 1515–1523, Apr. 2020, doi: 10.11591/ijece.v10i2.pp1515-1523.

[8] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, "Clustering models for hospitals in Jakarta using fuzzy c-means and k-means," *Procedia Computer Science*, vol. 216, pp. 356–363, 2023, doi: 10.1016/j.procs.2022.12.146.

[9] F. Yanuar, T. Abrari, I. Rahmi Hg, and A. Zetra, "Spatial Autoregressive Quantile Regression with Application on Open Unemployment Data," *Sci. Technol. Indones*, vol. 8, no. 2, pp. 321–329, Apr. 2023, doi: 10.26554/sti.2023.8.2.321-329.

[10] F. Yanuar, T. Abrari, A. Zetra, I. Rahmi Hg, D. Devianto, and S. Ahda, "Examining regional factors on malnutrition rate in Indonesia using spatial autoregressive approach," *Commun. Math. Biol. Neurosci.*, vol. 60, pp. 1–21, 2023, doi: 10.28919/cmbn/7955.

[11] J. Kang, L. Min, Q. Luan, X. Li, and J. Liu, "Novel modified fuzzy c-means algorithm with applications," *Digital Signal Processing*, vol. 19, no. 2, pp. 309–319, Mar. 2009, doi: 10.1016/j.dsp.2007.11.005.

[12] F. Zhao, L. Jiao, H. Liu, and X. Gao, "A novel fuzzy clustering algorithm with non local adaptive spatial constraint for image segmentation," *Signal Processing*, vol. 91, no. 4, pp. 988–999, Apr. 2011, doi: 10.1016/j.sigpro.2010.10.001.

[13] M. H. Aarabi, A. F. Kazerooni, P. Khateri, M. R. Ay, and H. S. Rad, "A robust MR-based attenuation map generation in short-TE MR images of the head employing hybrid spatial fuzzy C-means clustering and intensity inhomogeneity correction," *EJNMMI Phys*, vol. 1, no. S1, pp. A49, 2197-7364-1-S1-A49, Dec. 2014, doi: 10.1186/2197-7364-1-S1-A49.

[14] P. Roy and J. K. Mandal, "A Novel Selective Scale Space based Fuzzy C-means Model for Spatial Clustering," *Procedia Technology*, vol. 10, pp. 596–603, 2013, doi: 10.1016/j.protcy.2013.12.400.

[15] F. Yanuar and A. Zetra, "Length-of-Stay of Hospitalized COVID-19 Patients Using Bootstrap Quantile Regression," *IAENG International Journal of Applied Mathematics*, vol. 51, no. 3, pp. 799–810, 2021.

[16] F. Yanuar, H. Yozza, and A. Zetra, "Modified Quantile Regression for Modeling the Low Birth Weight," *Front. Appl. Math. Stat.*, vol. 8, p. 890028, Jun. 2022, doi: 10.3389/fams.2022.890028.

[17] M. Liao, Y. Li, F. Kianifard, E. Obi, and S. Arcona, "Cluster analysis and its application to healthcare claims data: a study of end-stage renal disease patients who initiated hemodialysis," *BMC Nephrol*, vol. 17, no. 1, p. 25, Dec. 2016, doi: 10.1186/s12882-016-0238-2.

[18] M. Bennasar-Veny *et al.*, "Cluster Analysis of Health-Related Lifestyles in University Students," *IJERPH*, vol. 17, no. 5, p. 1776, Mar. 2020, doi: 10.3390/ijerph17051776.

[19] A. Wartelle, F. Mourad-Chehade, F. Yalaoui, J. Chrusciel, D. Laplanche, and S. Sanchez, "Clustering of a Health Dataset Using Diagnosis Co-Occurrences," *Applied Sciences*, vol. 11, no. 5, p. 2373, Mar. 2021, doi: 10.3390/app11052373.

[20] A. Ultsch and J. Lötsch, "Euclidean distance-optimized data transformation for cluster analysis in biomedical data (EDOtrans)," *BMC Bioinformatics*, vol. 23, no. 1, p. 233, Dec. 2022, doi: 10.1186/s12859-022-04769-w.

[21] V. Bruni, M. L. Cardinali, and D. Vitulano, "A Short Review on Minimum Description Length: An Application to Dimension Reduction in PCA," *Entropy*, vol. 24, no. 2, p. 269, Feb. 2022, doi: 10.3390/e24020269.

[22] E. Eyal *et al.*, "Principal Component Analysis of Dynamic Contrast Enhanced MRI in Human Prostate Cancer:," *Investigative Radiology*, vol. 45, no. 4, pp. 174–181, Apr. 2010, doi: 10.1097/RLI.0b013e3181d0a02f.

[23] N. Migenda, R. Möller, and W. Schenck, "Adaptive dimensionality reduction for neural network-based online principal component analysis," *PLoS ONE*, vol. 16, no. 3, p. e0248896, Mar. 2021, doi: 10.1371/journal.pone.0248896.

[24] Z. Cebeci and F. Yildiz, "Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures," *JAI*, vol. 6, no. 3, pp. 13–23, Oct. 2015, doi: 10.17700/jai.2015.6.3.196.

[25] N. A. Ali, A. El Abbassi, and O. Bouattane, "Performance evaluation of spatial fuzzy C-means clustering algorithm on GPU for image segmentation," *Multimed Tools Appl*, vol. 82, no. 5, pp. 6787–6805, Feb. 2023, doi: 10.1007/s11042-022-13635-z.