BAREKENG: Journal of Mathematics and Its Applications
September 2024 Volume 18 Issue 3 Page 1919–1930
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss3pp1919-1930

IMPLEMENTATION OF FUZZY C-MEANS AND FUZZY POSSIBILISTIC C-MEANS ALGORITHMS ON POVERTY DATA IN INDONESIA

Dian Kurniasari [1*], Virda Kurniawati[2], Aang Nuryaman[3], Mustofa Usman[4], Rizki Khoirun Nisa[5]

[1,2,3,4,5]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung*
Jl. Prof. Dr. Soemantri Brodjonegoro No. 1, Bandar Lampung, 35145, Indonesia

Corresponding author's e-mail: ***dian.kurniasari@fmipa.unila.ac.id***

Article History:

Received: 28th February 2024
Revised: 19th April 2024
Accepted: 13th July 2024
Published: 1st September 2024

Keywords:

Cluster Analysis;
Fuzzy C-Means;
Fuzzy Possibilistic C-Means;
Partition Entropy.

ABSTRACT

Cluster analysis involves the methodical categorization of data based on the degree of similarity within each group to group data with similar characteristics. This study focuses on classifying poverty data across Indonesian provinces. The methodologies employed include the Fuzzy C-Means (FCM) and Fuzzy Probabilistic C-Means (FPCM) algorithms. The FCM algorithm is a clustering approach where membership values determine the presence of each data point in a cluster. On the other hand, the FPCM algorithm builds upon FCM and Possibilistic C (PCM) algorithms by incorporating probabilistic considerations. This research compares the FCM and FPCM algorithms using local poverty data from Indonesia, specifically examining the Partition Entropy (PE) index value. It aims to identify the optimal number of clusters for provincial-level poverty data in Indonesia. The findings indicate that the FPCM algorithm outperforms the FCM algorithm in categorizing poverty in Indonesia, as evidenced by the PE validity index. Furthermore, the study identifies that the ideal number of clusters for the data is 2.

How to cite this article:

D. Kurniasari, V. Kurniawati, A. Nuryaman, M. Usman and R. K. Nisa., " IMPLEMENTATION OF FUZZY C-MEANS AND FUZZY POSSIBILISTIC C-MEANS ALGORITHMS ON POVERTY DATA IN INDONESIA," *BAREKENG: J. Math. & App.*, vol. 18, iss. 3, pp. 1919-1930, September, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Poverty is a significant global issue affecting economic, cultural, geographical, and social aspects [1] [2]. The 2022 Multidimensional Poverty Index (MPI) report highlighted that over 1.2 billion people in 111 countries live in poverty. A recent reevaluation by Indonesia's Central Statistics Agency (BPS) in March 2022 found that 26.36 million people in Indonesia were living in poverty. Although there have been claims of a steady decline in poverty since 2011, World Bank data suggests otherwise. According to the World Bank, the number of impoverished people in Indonesia rose from 54 million to 67 million over the same period. This discrepancy arises from different poverty threshold standards used by the World Bank and Indonesia. Consequently, the World Population Review (WPR) ranks Indonesia as the 73rd poorest country in the world.

Poverty represents a significant barrier to regional or national development as it often results in diminished human capital through limited access to education and healthcare [3]. That underscores the critical need for effective government policies aimed at poverty alleviation to foster progress within a region or country. However, poverty rates can vary significantly across regions due to various influencing factors. Therefore, the government must gain a comprehensive understanding of poverty across different provinces in Indonesia by categorizing them based on poverty indicators [4].

Cluster analysis, a statistical technique, involves grouping objects with similar characteristics [5]. The primary objective is to divide a dataset into clusters such that elements within each cluster are more alike to each other compared to elements in different clusters, using predefined criteria [6]. This method offers several advantages, including efficiently organizing extensive data sets with multiple variables and its applicability to data measured on interval and ratio scales [7].

Cluster analysis encompasses a variety of algorithms designed for different objectives. One well-known algorithm is K-means, valued for its straightforwardness and efficiency in clustering extensive datasets. However, it can introduce bias when determining the placement of data points relative to their cluster centres. Consequently, several new algorithms have been proposed to address this issue, such as Fuzzy C-Means (FCM) and Fuzzy Probabilistic C-Means (FPCM). FCM, widely applied in pattern recognition and image processing, has demonstrated superior performance compared to K-means. Nevertheless, this algorithm is notably sensitive to noise [8].

Alternatively, the FPCM algorithm extends both the FCM and PCM algorithms, integrating fuzzy and possibilistic principles to overcome their limitations by utilizing two distinct types of memberships: fuzzy degree of membership and possibilistic absolute specificity values. This approach offers the advantage of effectively handling data that may include errors or outliers [9].

Numerous studies have explored the application of FCM and FPCM algorithms, such as the investigation by Rajkumar et al. [10]. They analyzed and validated the partitioned cluster FCM using a subset of CiteScore data. The study employed fuzzy clustering and FCM algorithms on a portion of the CiteScore dataset to investigate data points with similar distances. Their findings suggest that FCM is superior in effectiveness and efficiency compared to the fuzzy clustering method. Wiharto and Suryani [11] examined the efficacy of K-means and FCM algorithms in segmenting retinal blood vessels. Their performance differs significantly, particularly in establishing thresholds based on mean or median values. The research indicates that the FCM algorithm outperforms the K-means algorithm.

Apsari et al. [12] evaluated and categorized students according to their performance using the FCM and FPCM algorithms. These methods effectively identify high-achieving students by minimizing the objective function. Rahakbaw et al. [13] utilized the FCM algorithm to determine scholarship recipients, considering multiple criteria such as GPA, semester, number of dependents, parental income, and transportation methods. Their findings suggest that this algorithm represents a viable option for evaluating scholarship eligibility. Jayasree and Selvakumari [14] introduced an enhanced FPCM clustering algorithm for predicting students' academic outcomes based on their health status. Their research demonstrates that the proposed FPCM algorithm outperforms other clustering techniques like K-Means, K-Medoids, and FCM, achieving an accuracy rate of up to 93%.

Based on the above description, government strategies and policies to alleviate poverty in Indonesia can be effectively implemented through equitable development. Moreover, the government needs to classify Indonesian provinces based on their specific poverty characteristics to understand poverty dynamics in each region comprehensively. We propose employing clustering analysis using the FCM and FPCM algorithms on various datasets of Indonesian poverty. The poverty variables considered include the poverty line, percentage of the population living in poverty, poverty depth index, poverty severity index, and total number of individuals below the poverty line. This study will compare the two algorithms based on their accuracy and validity. It will utilize the Partition Entropy Index (PEI) to identify the most suitable algorithm for clustering regencies and cities across Indonesia.

2. RESEARCH METHODS

The study utilizes poverty data from Indonesia in 2021, consisting of 514 data points. The variables encompass the poverty line, percentage of the population classified as poor, poverty depth index, poverty severity index, and total number of people living in poverty.

a. The process begins with sourcing data from the Indonesian Central Bureau of Statistics (BPS), encompassing poverty data for each district/city in Indonesia. The variables utilized in this study are as follows: x_1 representing the poverty line, x_2 representing the percentage of the poor population, x_3 representing the poverty depth index, x_4 representing the poverty severity index and x_5 representing the number of poor population.

b. Standardizing poverty statistics involves utilizing z-scores. This process is necessary due to variations in the unit sizes of the variables within the dataset.

c. The clustering process using the FCM and FPCM algorithms involves setting a range for the number of clusters. For instance, the number of clusters considered ranges from two to a maximum relevant to Indonesia's poverty context. Specifically, the number of clusters examined is between 2 and 5, with a weighting exponent of 2. This range of 2 to 5 clusters is chosen because it is considered sufficient to represent the variation in urban poverty data without adding unnecessary complexity to the analysis.

d. Evaluate the clusters formed using PEI.

e. Compare the results of both algorithms to determine which algorithm is most suitable for clustering.

2.1 Data Standardization

Data standardization, called data scaling or normalization, represents an essential preprocessing step in data analysis and machine learning. It involves converting data into a uniform format or scale to facilitate straightforward analysis, comparison, and modeling. The main objective of data standardization is to achieve uniform scaling across all variables or features within a dataset, typically aiming for a mean of 0 and a standard deviation of 1. This approach enables equitable comparison among diverse variables. It prevents any one feature from unduly influencing the analysis based solely on its scale [15].

In common standardization, each variable is typically transformed into standard values known as z-scores [16]. Equation (1) specifies the procedure for standardizing the data.

$$z = \frac{x_i - \mu}{\sigma} \tag{1}$$

2.2 Fuzzy C-Means (FCM)

The FCM clustering algorithm holds substantial prominence and is widely utilized across various domains, including pattern recognition, machine learning, and data mining [17]. Initially introduced by Dunn [18], the algorithm was subsequently refined by Bezdek et al. [19].

Shan et al. [20] categorized the FCM algorithm as a form of soft clustering addressed through iterative updates of membership matrices and cluster centroids. In mathematical terms, the centroid of a cluster in the FCM method can be computed using Equation (2):

$$v_{kj} = \frac{\sum_{k=1}^{n}((\mu_{ik})^w x_{ij})}{\sum_{k=1}^{n}(\mu_{ik})^w} \tag{2}$$

In this context, v_{kj} represents the j-th component of the centroid for cluster k. The term n denotes the number of data points. The value μ_{ik} is the membership degree of data point i in cluster k, ranging from 0 to 1. The parameter w is the fuzziness factor, usually greater than 1. Lastly, x_{ij} refers to the j-th component of the i-th data point.

The value μ_{ik} indicates the strength of membership of the i-th data point in cluster k. It ranges from 0 to 1, with values closer to 1 signifying a solid membership and values closer to 0 indicating a weak membership. **Equation (3)** defines how to calculate this membership degree.

$$\mu_{ik} = \left[\frac{\left[\sum_{j=1}^{m}(x_{ij}-v_{kj})^2\right]^{\frac{1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}(x_{ij}-v_{kj})^2\right]^{\frac{1}{w-1}}} \right]^{-1} \quad (3)$$

The summation symbol, $\sum_{j=1}^{m}(x_{ij} - v_{kj})$ used to compute the overall difference between each data component and the centroid cluster component across all dimensions. This method is implemented on every individual data point inside the collection. The data components x_{ij} represent the values assigned to each dimension, while the centroid cluster components v_{kj} the values that indicate the centre of each cluster across all dimensions. The parameter w is a critical determinant in the FCM algorithm, controlling the level of fuzziness in clustering. Usually, the value is more than 1, and as the value increases, the boundaries between the resulting clusters get increasingly muddled. The FCM algorithm uses this Equation to iteratively compute the cluster centers and the degree of membership of each data point to each cluster. This information can generate more adaptable data segmentation than classical clustering approaches.

An iteration process to minimize the objective function is required to obtain accurate cluster centers and membership degrees. The formula for the objective function in the FCM technique can be derived using **Equation (4)**.

$$J_w(U,V) = \sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik})^w d_{ik}^2 \quad (4)$$

Equation (4) computes the objective function in the FCM algorithm by summing the weighted squared distances between data points and cluster centroids, considering membership degrees. Here, n represents the total data points, c is the total clusters, μ_{ik} indicates data point i membership degree to cluster k, d_{ik} signifies the Euclidean distance between data point i and cluster centroid k, and w is the fuzziness parameter controlling clustering's fuzziness degree. The aim is to minimize $J_w(U,V)$ to attain precise cluster centres and membership degrees through iterative processes. Lesser $J_w(U,V)$ values imply better cluster quality.

The Euclidean distance equation utilized in this study is defined by **Equation (5)**:

$$d_{ik} = \sqrt{\sum_{j=1}^{m}(x_{ij} - v_{kj})^2} \quad (5)$$

where m denotes the dimensions, x_{ij} represents data point i's component j, and v_{kj} indicates cluster k's centroid component j. It measures dissimilarity between data points and cluster centroids.

2.3 Fuzzy Possibilistic C-Means (FPCM)

The FPCM represents a progression from the algorithms utilized in FCM and PCM. It incorporates cluster centres, specificity, and objective function values **[21]**. The cluster's centre in this method can be found using **Equation (6)**.

$$v_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik}^{w} + t_{ik}{}^{\eta})x_{ij})}{\sum_{i=1}^{n}\mu_{ik}^{w} + t_{ik}{}^{\eta}} \quad (6)$$

The matrix T, referred to as the absolute distinctiveness matrix, is derived from the ultimate combination of the membership degree (μ_{ik}) and cluster centre (v_{kj}) values in the FCM algorithm. **Equation (7)** presents the formal mathematical formulation employed to calculate this absolute distinctiveness matrix within the FPCM methodology.

$$\boldsymbol{T} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1k} \\ t_{21} & t_{22} & \dots & t_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ t_{i1} & t_{i2} & \dots & t_{ik} \end{bmatrix} \tag{7}$$

In the Equation provided above, there exist matrix elements denoted as $t_{11}, t_{12}, \dots t_{ik}$. Each element in the absolute distinctiveness matrix can be computed using **Equation (8)**:

$$t_{ik} = \left[\frac{\left[\sum_{j=1}^{m} (x_{ij} - v_{kj})^2\right]^{\frac{1}{\eta-1}}}{\sum_{i=1}^{n} \left[\sum_{j=1}^{m} (x_{ij} - v_{kj})^2\right]^{\frac{1}{\eta-1}}} \right]^{-1} \tag{8}$$

The objective function of the FPCM algorithm is defined by **Equation (9)**:

$$J_{w,\eta}(U,T,V) = \sum_{k=1}^{c} \sum_{i=1}^{n} (\mu_{ik}{}^{w} + t_{ik}{}^{\eta}) \sqrt{\sum_{j=1}^{m} (x_{ij} - v_{kj})}^2 \tag{9}$$

3. RESULTS AND DISCUSSION

3.1 Data Input

This study aims to cluster each province in Indonesia based on 2021 poverty data at the district level in 2021, using a variety of relevant variables. **Table 1** provides a detailed summary of the poverty statistics at the district level in Indonesia for 2021.

Table 1. Poverty data at the district level in Indonesia

District Name	X_1	X_2	X_3	X_4	X_5
Simeulue	458896	18.98	2.37	0.50	18.25
Aceh Singkil	487249	20.36	3.67	0.91	25.48
Aceh Selatan	418689	13.18	1.69	0.40	32.25
Aceh Tenggara	404725	13.41	2.51	0.71	29.31
……	……	……	……	……	……
Puncak	721541	36.26	9.92	3.49	40.78
Dogiyai	557009	28.81	3.68	0.78	28.38
Intan Jaya	725106	41.66	5.92	1.05	21.31
Deiyai	627742	40.59	3.79	0.67	30.83
Kota Jayapura	1051297	11.39	2.53	0.79	34.79

3.2 Data Standardization

This data standardization phase aims to homogenize the units of measurement across different data sets. As described in Section. 2, the research data includes variables with varying units of measurement. Therefore, it is essential to use the z-score method for data normalization. This normalization aims to achieve more accurate and reliable clustering results. **Table 2** presents the standardized poverty data at the district level in Indonesia.

Table 2. Standardization of poverty data at the district level in Indonesia

X_1	X_2	X_3	X_4	X_5
-0,01528	0,89919	0,15204	-0,08261	-0,58672
0,24073	1,08421	0,82583	0,42850	-0,46667

X_1	X_2	X_3	X_4	X_5
-0,37834	0,12158	-0,20040	-0,20727	-0,35426
……	……	……	……	……
0,87064	2,21712	0,83102	0,26644	-0,41852
2,38850	3,93994	1,99202	0,60302	-0,53591
1,50934	3,79648	0,88803	0,12931	-0,37784
5,33391	-0,11841	0,23497	0,27890	-0,31209

3.3 Clustering Process

The next step involves performing clustering using the FCM and FPCM methods. In the FCM method, a weighting exponent (w) of 2 is applied, and the number of clusters varies from 2 to 5. The process includes 1000 iterations using the Euclidean distance metric. **Table 3** presents the list of Indonesian provinces assigned to each cluster.

Table 3. List of provinces in Indonesia

No	Provinces	No	Provinces
1	Aceh	18	Nusa Tenggara Barat
2	Sumatera Utara	19	Nusa Tenggara Timur
3	Sumatera Barat	20	Kalimantan Barat
4	Riau	21	Kalimantan Tengah
5	Jambi	22	Kalimantan Selatan
6	Sumatera Selatan	23	Kalimantan Timur
7	Bengkulu	24	Kalimantan Utara
7	Bengkulu	24	Kalimantan Utara
8	Lampung	25	Sulawesi Utara
9	Kep. Bangka Belitung	26	Sulawesi Tengah
10	Kep. Riau	27	Sulawesi Selatan
11	DKI Jakarta	28	Sulawesi Tenggara
12	Jawa Barat	29	Gorontalo
13	Jawa Tengah	30	Sulawesi Barat
14	DI Yogyakarta	31	Maluku
15	Jawa Timur	32	Maluku Utara
16	Banten	33	Papua Barat
17	Bali	34	Papua

Figure 1 to **Figure 4** offers a detailed depiction of poverty clustering in Indonesia using the FCM method, presented through graphical plots. These visuals utilize principal component analysis to streamline dimensions or variables into a more succinct format.

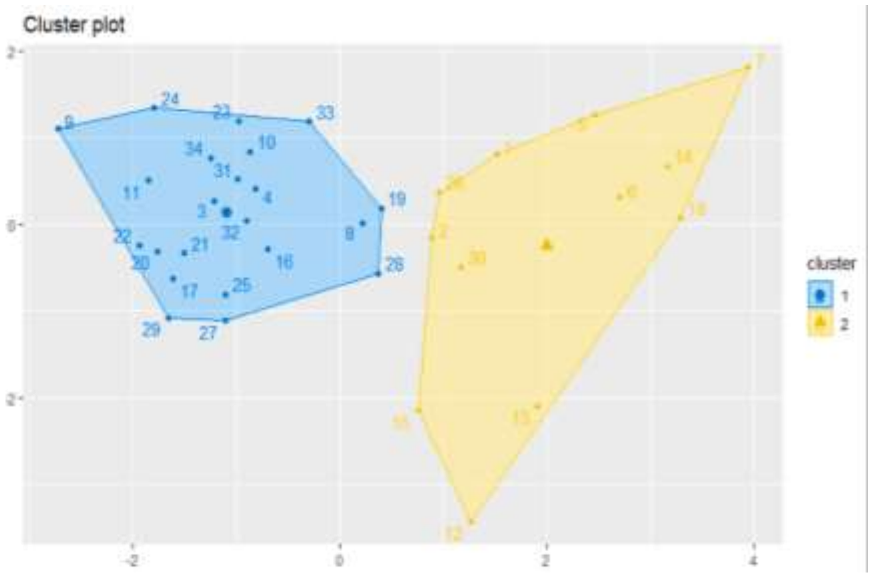

Figure 1. FCM method plot result with 2 cluster

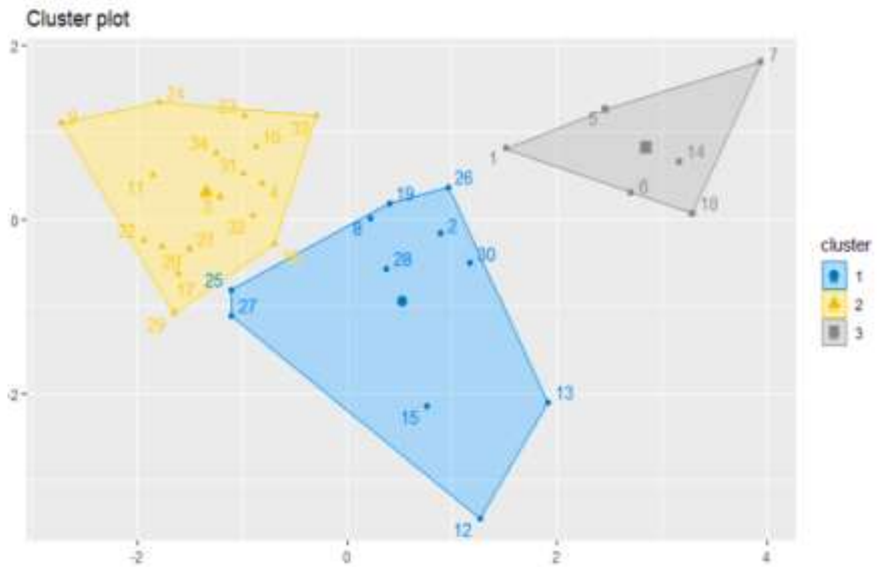

Figure 2. FCM method plot result with 3 cluster

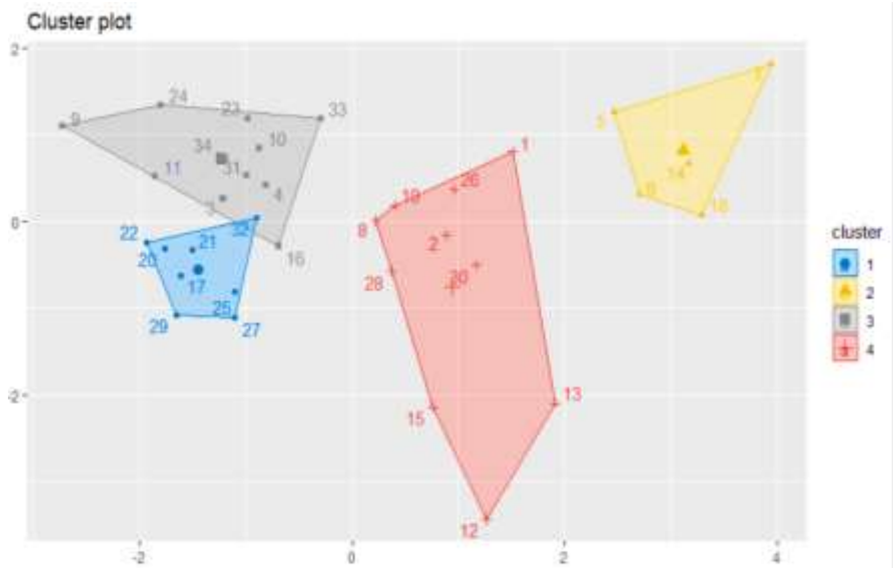

Figure 3. FCM method plot result with 4 cluster

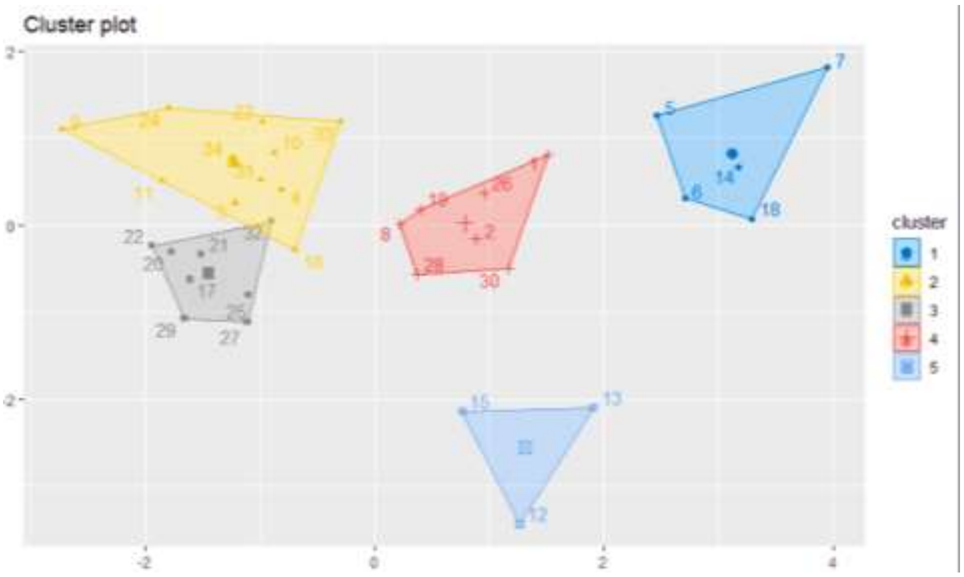

Figure 4. FCM method plot result with 5 cluster

Above of Figure 1 to Figure 4 show that the results of clustering poverty data at the provincial level in Indonesia using the FCM method. Each figure displays points representing provinces grouped within each cluster, with their respective order detailed in Table 1. In Figure 1, Cluster 1 exhibits points closer together than Cluster 2, suggesting more significant similarity in characteristics within Cluster 1 than Cluster 2. Similarly, in Figure 2, Cluster 2 shows closer points than Clusters 1 and 3, indicating more similarity within Cluster 2. Figure 3 demonstrates that Clusters 1 and 3 points are closer together than Clusters 2 and 4, highlighting similarities between Clusters 1 and 3. In Figure 4, Clusters 2, 3, and 4 points are closer together than in Clusters 1 and 5, emphasizing similarities among Clusters 2, 3, and 4. Subsequently, further elaboration on poverty clustering in Indonesia using the FPCM method will be presented through visualization in Figures 5-8.

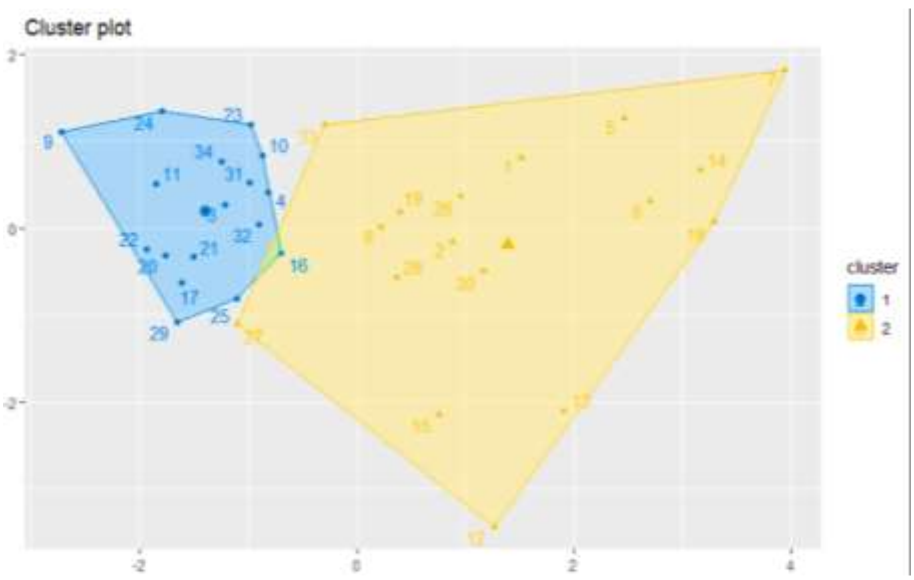

Figure 5. FPCM method plot results for 2 cluster

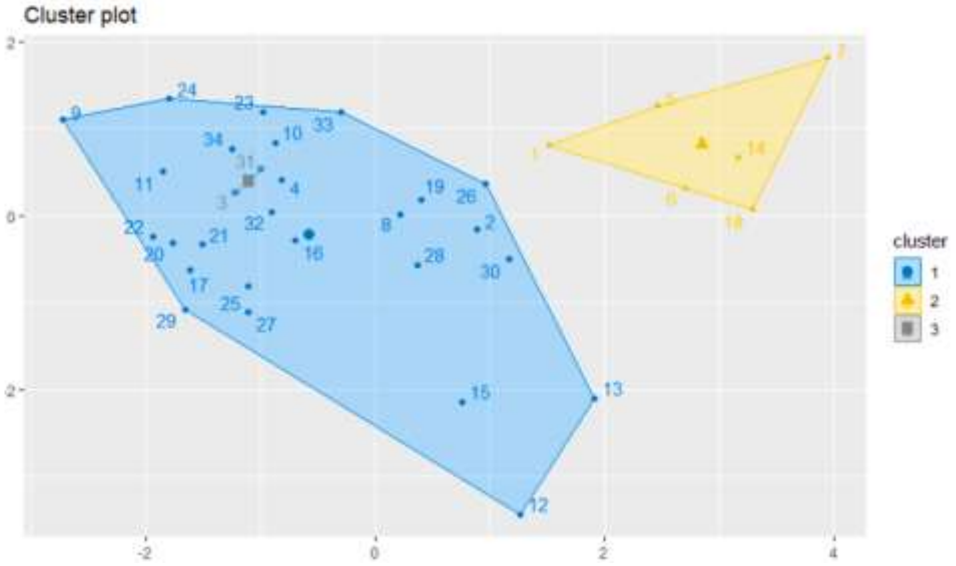

Figure 6. FPCM method plot results for 3 cluster

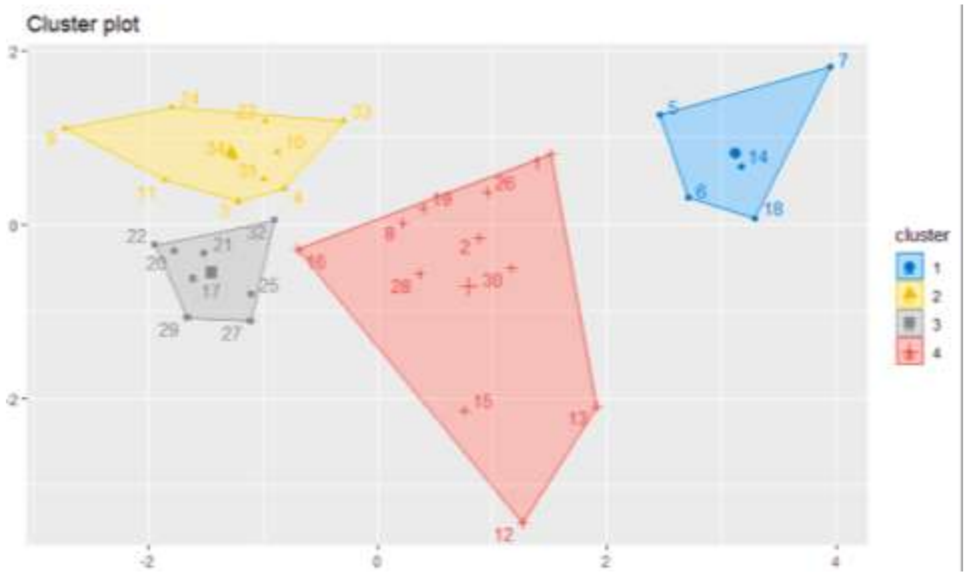

Figure 7. FPCM method plot results for 4 cluster

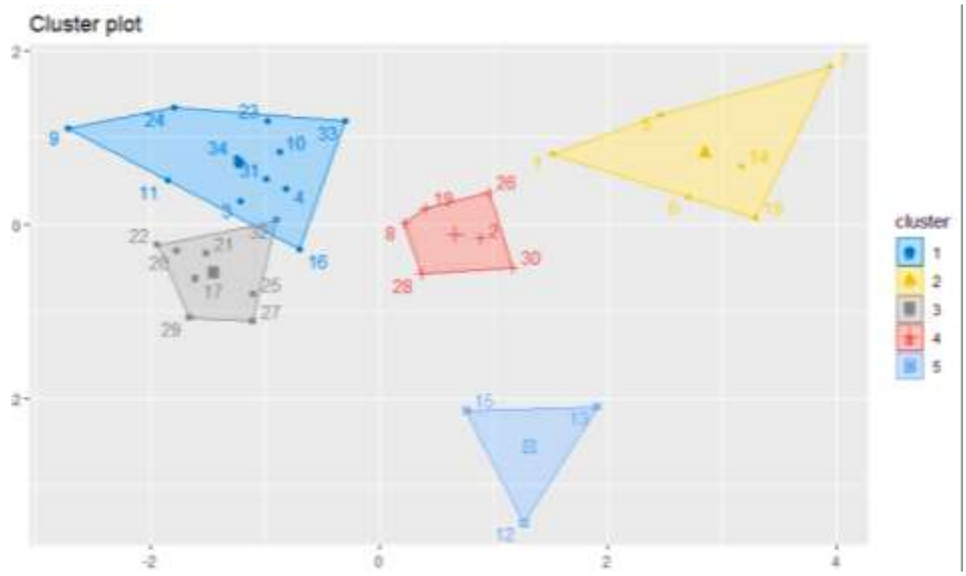

Figure 8. FPCM method plot results for 5 cluster

Cluster 1 in **Figure 5** loser dots than Cluster 2, implying that Cluster 1 shares more similar characteristics among its members than Cluster 2. **Figure 6** illustrates that the points within Cluster 1 are closely grouped in contrast to Clusters 2 and 3, suggesting greater homogeneity in the characteristics of Cluster 1 compared to Clusters 2 and 3. **Figure 7** demonstrates that the points in Clusters 2 and 3 are closer together than in Clusters 1 and 4, indicating higher similarity in the characteristics of Clusters 2 and 3 relative to Clusters 1 and 4. **Figure 8** reveals that Clusters 1, 3, and 4 points are closer together than in Clusters 2 and 5, implying that the properties of Clusters 1, 3, and 4 are more comparable than those of Clusters 2 and 5.

After generating clustering results using FCM and FPCM methods, cluster evaluation was performed to determine the optimal algorithm for poverty clustering based on the PE value. **Table 4** presents the results of the poverty data cluster evaluation using PE.

Table 4. Evaluation results of cluster methods FCM and FPCM

Number of Clusters	*Partition Entropy* (PE)	
	FCM	FPCM
Cluster 2	0,4000031	0,1560257
Cluster 3	0,6827615	0,1615916
Cluster 4	0,7853068	0,2813303
Cluster 5	0,8064033	0,3128189

Table 4 shows that the PE validity index value in the FCM method consistently exceeds that in the FPCM method across clusters 2 to 5. The FPCM approach outperformed the FCM method in classifying poverty in Indonesia.

After selecting the best method, the next step is to determine the optimal number of clusters using the FPCM method, which is the best method identified. This determination is based on the PE validity index value, where the lowest PE value indicates the most optimal cluster. The criterion for the PE validity index in determining the optimal number of clusters is that a smaller PE value, closer to zero, indicates better results.

In Indonesia's poverty context, optimizing clusters is crucial for identifying groups of people with similar poverty characteristics. Based on **Table 4**, it can be concluded that the optimal number of clusters is 2. That is because, among all the clusters tested (ranging from 2 to 5), the two-cluster solution using the FPCM method has the lowest PE value, 0.1560257.

Therefore, using this method, two clusters can provide the most precise and most informative segmentation of the poor population in Indonesia. The first cluster might include individuals or families experiencing the most severe poverty. In contrast, the second cluster could consist of those on the brink of poverty who still require assistance and intervention. By identifying and understanding the characteristic differences between these two clusters, more effective and targeted programs and policies can be designed to reduce poverty in Indonesia.

4. CONCLUSIONS

After closely analyzing and comparing the results of the FCM and FPCM methods discussed earlier, it becomes evident that the FPCM approach proves significantly more effective for clustering provincial-level poverty data in Indonesia. Evaluating this method's effectiveness entails examining the PE validity index. Specifically, in the FPCM method, the PE validity index is lower than that of the FCM method, indicating that cluster 2 exhibits the lowest validity index values among various clusters, specifically 0.46827 and 0.02074, respectively. This finding highlights that the FPCM method has achieved more accurate and reliable clustering results.

REFERENCES

[1] R. Probosiwi, “Pengangguran dan Pengaruhnya Terhadap Tingkat kemiskinan,” *J. Penelit. Kesejaht. Sos.*, vol. 15, no. 02, 2016.

[2] B. Gweshengwe and N. H. Hassan, "Defining the characteristics of poverty and their implications for poverty analysis," *Cogent Social Sciences*, vol. 6, no. 1. 2020. doi: 10.1080/23311886.2020.1768669.

[3] T. Agus Triono and R. C. Sangaji, “Faktor Mempengaruhi Tingkat Kemiskinan di Indonesia: Studi Literatur Laporan Data

Kemiskinan BPS Tahun 2022," *J. Soc. Bridg.*, vol. 1, no. 1, pp. 59–67, 2023, doi: 10.59012/jsb.v1i1.5.

[4] N. Afira and A. W. Wijayanto, "Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019," *Komputika J. Sist. Komput.*, vol. 10, no. 2, pp. 101–109, 2021, doi: 10.34010/komputika.v10i2.4317.

[5] D. Setiawan and A. Zahra, "Pengelompokan Kemiskinan di Indonesia Menggunakan Time Series Based Clustering," *Inferensi*, vol. 6, no. 1, p. 83, 2023, doi: 10.12962/j27213862.v6i1.14969.

[6] P. Dubey and R. K. Chakrawarti, "Analysis of Clustering-Algorithms for Efficient Data Mining," *Int. Conf. Electr. Electron. Comput. Sci. Math. Phys. Educ. Manag.*, no. April, 2023, [Online]. Available: https://www.researchgate.net/publication/369819077

[7] M. S. Yana, L. Setiawan, E. M. Ulfa, and A. Rusyana, "Penerapan Metode K-Means dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018," *J. Data Anal.*, vol. 1, no. 2, pp. 93–102, 2018, doi: 10.24815/jda.v1i2.12584.

[8] X. Geng, Y. Mu, S. Mao, J. Ye, and L. Zhu, "An Improved K-Means Algorithm Based on Fuzzy Metrics," *IEEE Access*, vol. 8, pp. 217416–217424, 2020, doi: 10.1109/ACCESS.2020.3040745.

[9] N. Grover, "A study of various Fuzzy Clustering Algorithms," *Int. J. Eng. Res.*, vol. 3, no. 3, pp. 177–181, 2014, doi: 10.17950/ijer/v3s3/310.

[10] K. V. Rajkumar, A. Yesubabu, and K. Subrahmanyam, "Fuzzy clustering and Fuzzy C-Means partition cluster analysis and validation studies on a subset of CiteScore dataset," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 4, pp. 2760–2770, 2019, doi: 10.11591/ijece.v9i4.pp2760-2770.

[11] W. Wiharto and E. Suryani, "The comparison of clustering algorithms K-means and fuzzy C-means for segmentation retinal blood vessels," *Acta Inform. Medica*, vol. 28, no. 1, pp. 42–47, 2020, doi: 10.5455/AIM.2020.28.42-47.

[12] G. R. Apsari, M. S. Pradana, and N. E. Chandra, "Implementasi Fuzzy C-Means dan Possibilistik C-Means Pada Data Performance Mahasiswa," *Unisda J. Math. Comput. Sci.*, vol. 6, no. 2, pp. 39–48, 2020, doi: 10.52166/ujmc.v6i2.2392.

[13] D. L. Rahakbauw, V. Y. I. Ilwaru, and M. H. Hahury, "Implementasi Fuzzy C-Means Clustering Dalam Implementation Of Fuzzy C-Means Clustering In," *J. Ilmu Mat. dan Terap.*, vol. 11, pp. 1–12, 2019, [Online]. Available: https://media.neliti.com/media/publications/277582-implementasi-fuzzy-c-means-clustering-da-3afa5ba1.pdf

[14] R. Jayasree and N. A. S. Selvakumari, "Analyzing Student Performance using Fuzzy Possibilistic C-Means Clustering Algorithm," *Indian J. Sci. Technol.*, vol. 16, no. 38, pp. 3230–3235, Oct. 2023, doi: 10.17485/ijst/v16i38.226.

[15] T. A. S. Srinivas, Y. Sravanthi, Y. V. Kumar, and V. D. Shirith, "Data Standardization Key to Effective Data Integration," *Adv. Innov. Comput. Program. Lang.*, vol. 6, no. 1, 2023.

[16] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability & Statistics for Engineers Scientists Probability & Statistics for Engineers & Scientists*, vol. 6. Pearson Prentice Hall, 2007. [Online]. Available: http://www.amazon.com/Probability-Statistics-Engineers-Scientists-8th/dp/0131877119/ref=sr_1_3?ie=UTF8&s=books&qid=1257621824&sr=1-3

[17] B. A. Pimentel, R. De Amorim Silva, and J. C. S. Costa, "Fuzzy C-Means Clustering Algorithms with Weighted Membership and Distance," *Int. J. Uncertainty, Fuzziness Knowldege-Based Syst.*, vol. 30, no. 4, pp. 567–594, 2022, doi: 10.1142/S0218488522500143.

[18] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, 1973, doi: 10.1080/01969727308546046.

[19] 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: 10.1016/0098-3004(84)90020-7.

[20] Y. Shan, S. Li, F. Li, Y. Cui, and M. Chen, "Dual-level clustering ensemble algorithm with three consensus strategies," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-49947-9.

[21] M. H. F. Zarandi, S. Sotudian, and O. Castillo, "A new validity index for fuzzy-possibilistic c-means clustering," *Sci. Iran.*, vol. 28, no. 4, pp. 2277–2293, 2021, doi: 10.24200/SCI.2021.50287.1614.