Literasi Nusantara
vol.5 no. 2, March 2025 – June 2025
https://journal.citradharma.org/index.php/literasinusantara/index

ISSN 2746 8208 (Online)
ISSN 2746 3575 (Print)

Evaluation of Human Development Index Clustering Results Using Fuzzy C-Means and Possibilistic C-Means

Lativa Yulia Taviani[1*], Eva Yulia Puspaningrum[2], Achmad Junaidi[3]

[1,2,3] Universitas Pembangunan Nasional Veteran, Jawa Timur, Surabaya, Indonesia
21081010294@student.upnjatim.ac.id , evapuspaningrum.if@upnjatim.ac.id , achmadjunaidi.if@unjatim.ac.id

DOI : *https://doi.org/10.56480/jln.v5i2.1571*

Received: April 17, 2025
Revised: May 23, 2025
Accepted: June 04, 2025

Abstract

This study aims to evaluate the clustering results of the Human Development Index (HDI) in East Java using two fuzzy-based algorithms: Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM). The dataset includes key indicators Life Expectancy (LE), Mean Years of Schooling (MYS), Expected Years of Schooling (EYS), and Adjusted Per Capita Expenditure (APCE) sourced from the Central Bureau of Statistics. After preprocessing the data and applying Bayesian Optimization to determine optimal parameters, both clustering methods were executed and evaluated using internal metrics: Partition Coefficient (PC), Partition Entropy (PE), and Modified Partition Coefficient (MPC). The results show that both FCM and PCM successfully formed three meaningful clusters representing different levels of human development. However, PCM achieved higher clustering quality, as indicated by superior PC, PE, and MPC values. These findings highlight the effectiveness of PCM in handling complex, overlapping socio-economic data and offer insights for more accurate regional segmentation. The methodology is also applicable to broader socio-economic clustering tasks beyond HDI.

Keywords– *Human Development Index, Fuzzy C-Means, Possibilistic C-Means, Clustering Evaluation, Bayesian Optimization.*

Literasi Nusantara. vol.5 no. 2, March – June 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

1. Introduction

The Human Development Index (HDI) is a comprehensive indicator used to measure the quality of life and welfare of communities in a region, calculated based on three main dimensions: health, education, and standard of living (Simanjuntak et al., 2024). These dimensions are represented by indicators such as Life Expectancy (LE), Expected Years of Schooling (EYS), Mean Years of Schooling (MYS), and Adjusted Per Capita Expenditure (PCE)(Sari et al., 2022). Increasing the HDI is a primary goal of national and regional development. However, analysis of HDI data at the regional level, especially at the district/city level, often faces challenges in defining clear boundaries between groups, as HDI values tend to be close and the data often contain complexities.

One widely adopted approach for handling data with overlapping characteristics is data mining, particularly clustering techniques (Zai, 2022; Hendrastuty, 2024). Fuzzy clustering methods are especially popular because they allow data points to belong to multiple clusters with varying degrees of membership (Budiawan, 2024). Fuzzy C-Means (FCM) is a frequently used algorithm due to its flexibility in handling unclassified data; however, it has limitations in dealing with noise and outliers (Huddin, 2023; Yasir & Firmansyah, 2024). To address these limitations, Possibilistic C-Means (PCM) was developed with a local membership approach that offers more adaptive handling of complex data and noise (Mubarok, 2024).

Previous studies have explored the application of FCM and PCM in clustering social data, including HDI. For instance, Isah et al. (2024) compared FCM and Fuzzy Probabilistic C-Means (FPCM) in clustering provincial HDI data in Indonesia, while Hartanto et al. (2024) analyzed HDI in East Java using FCM and demonstrated satisfactory clustering results. However, these studies primarily focused on general methodological comparisons and did not provide an in-depth evaluation of clustering outcomes or the practical implications of regional mapping based on HDI. Additionally, some studies still employed Grid Search for parameter optimization, which is less efficient in handling complex data spaces (Zhang et al., 2022).

This research addresses these gaps by focusing on the evaluation of HDI clustering results using two widely used methods: Fuzzy C-Means and Possibilistic C-Means. Unlike previous studies that concentrated on parameter optimization or theoretical comparisons, this study emphasizes the in-depth evaluation of clustering quality using metrics such as Partition Coefficient (PC), Partition Entropy (PE), and Modified Partition Coefficient (MPC) (Faturahman & Hidayati, 2025). Thus, this research not only supports previous findings but also provides a new contribution to understanding the practical and applicative performance of clustering methods, particularly for regional grouping based on HDI.

The objective of this study is to evaluate the clustering outcomes of HDI data using Fuzzy C-Means and Possibilistic C-Means, focusing on the analysis of cluster quality. The results are expected to provide clearer insights into the effectiveness of these methods in supporting data-driven regional development mapping, serving as a reference for policymakers and future development analysis model designs.

2. Method

This study is designed to evaluate the clustering outcomes of Human Development Index (HDI) data using the Fuzzy C-Means and Possibilistic C-Means algorithms. The research process involves several interrelated stages, starting from data collection, data preprocessing, parameter initialization, implementation of clustering algorithms, and finally, evaluation of clustering results. The flow of the research stages is illustrated in Figure 1.

Literasi Nusantara. vol.5 no. 2, March – June 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

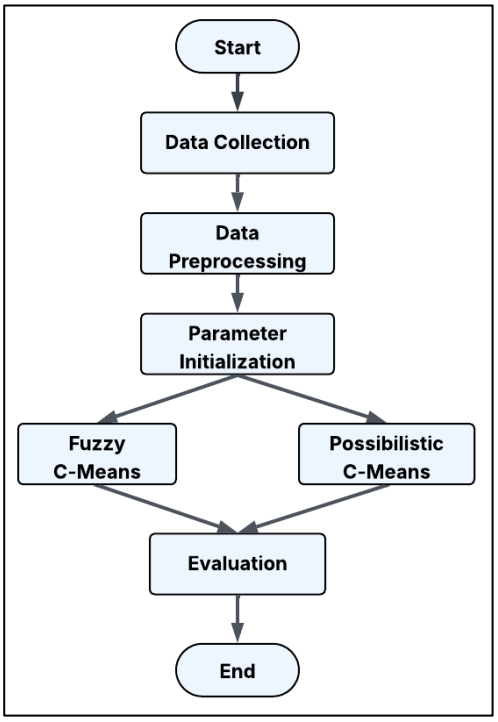

Figure 1. Research Process Flowchart

The following describes each stage of the research process in detail:

a. Data Collection

The data used in this study were obtained from the official Human Development Index (HDI) records published by the Central Bureau of Statistics (BPS) of East Java Province. The dataset includes key variables representing the core dimensions of HDI, namely Life Expectancy (LE), Expected Years of Schooling (EYS), Mean Years of Schooling (MYS), and Adjusted Per Capita Expenditure (APCE) (Badan Pusat Statistik, 2024). The data were collected at the district/city level to capture socio-economic variations across regions.

b. Data Preprocessing

Data preprocessing is a crucial stage in data analysis aimed at transforming raw data into a cleaner, more structured form ready for subsequent analysis (Putra et al., 2023). This stage includes:

1) Data Cleaning : Handling missing values by applying median imputation, where missing entries are replaced with the median value of the respective variable (Putra et al., 2023).

2) Outlier Detection and Handling: Using the boxplot method to detect outliers, where points falling outside the whiskers are considered as potential outliers (Sudipa et al., 2023; Thakkar et al., 2021). Identified outliers were adjusted to minimize their impact on clustering results.
3) Data Normalization: Applying Z-Score normalization to standardize variable scales, ensuring a mean of 0 and a standard deviation of 1, thus allowing balanced analysis across variables with different scales in the clustering process (Setiawan et al., 2023).

c. Parameter Inizialitation

Parameter initialization is an essential stage before the implementation of clustering algorithms. In this phase, the main parameters required for the Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM) algorithms are determined, including the number of clusters (c), the fuzzifier value (m), the weighting power (η) specific to PCM, the maximum number of iterations (MaxIter), and the minimum error threshold (ξ).

The number of clusters (c) used in this study ranges from 3 to 9, while the fuzzifier value (m) is set within the range of 1.5 to 3.0. For the PCM algorithm, the weighting power (η) is set to 2. The maximum number of iterations (MaxIter) is set to 1000, and the minimum error threshold (ξ) is 0.001. These parameters were selected based on prior literature that recommends this range of parameters for fuzzy clustering analysis (Setiawan et al., 2023). Additionally, the parameter ranges are adapted from Table 3.2 in this study, which shows the optimal parameters identified through preliminary exploration.

Although this study employs a manual exploration approach for parameter selection, previous research suggests that optimal parameter selection can be conducted using Bayesian Optimization (Zhang et al., 2022). Bayesian Optimization is a probabilistic optimization method that utilizes a surrogate model, such as Gaussian Process Regression (GPR), to approximate the objective function and select parameters that yield optimal results (Zhang et al., 2022). However, in this study, parameters were determined through a

Literasi Nusantara. vol.5 no. 2, March – June 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

simple exploratory approach based on literature and initial analysis, without employing Bayesian-based optimization.

d. Clustering Algorithm Implementation

Clustering is a data mining technique that groups a set of objects into clusters based on the similarity of their characteristics, and is classified as unsupervised learning (Akbar, 2023; Rahayu et al., 2024). The goal is to group objects that are highly similar within the same cluster while ensuring dissimilar objects are grouped into different clusters. Clustering is commonly applied in fields such as customer segmentation, regional development analysis, and socio-economic studies (Akbar, 2023).

This study focuses on partitioning methods, which divide data into clusters with the aim of maximizing intra-cluster similarity and minimizing inter-cluster similarity. Unlike traditional methods such as K-Means that assign each data point to only one cluster, Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM) allow each data point to have degrees of membership in multiple clusters based on its proximity. These fuzzy-based clustering algorithms are chosen for their ability to handle overlapping data and uncertainty. The clustering algorithms used in this study are:

1) Fuzzy C-Means (FCM)

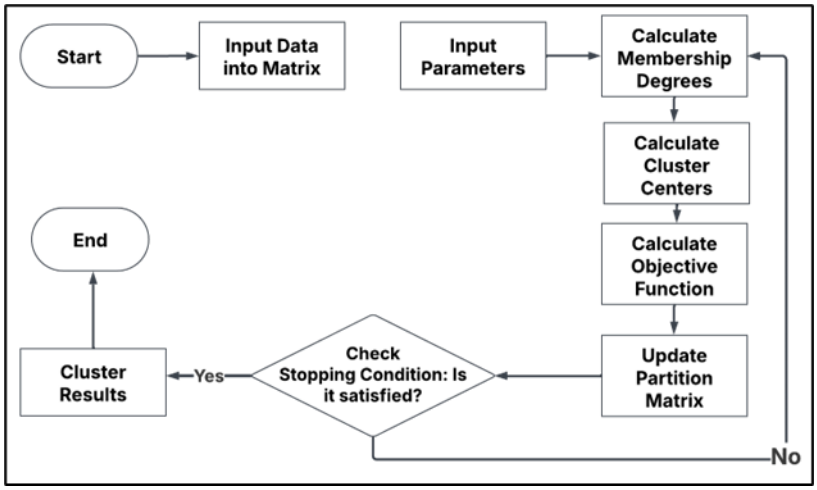

Figure 2. Flowchart of the Fuzzy C-Means (FCM) Algorithm

The Fuzzy C-Means (FCM) algorithm is a fuzzy clustering technique introduced by Dunn in 1973 and later developed by Bezdek in 1981 (Hidayat et al., 2017). Unlike hard clustering methods, FCM allows each data point to belong to multiple clusters by assigning membership degrees ranging from 0 to 1 (Kadja et al., 2023).

The workflow of the FCM algorithm used in this study is illustrated in Figure 2. The algorithm starts by inputting data into a matrix and initializing parameters, including the number of clusters and fuzzifier value. It then iteratively calculates membership degrees, updates cluster centers, and computes the objective function. The partition matrix is updated based on the calculated membership degrees, and the algorithm checks the stopping condition. Once the stopping condition is satisfied, final clusters are determined based on the highest membership values (Lestari et al., 2023).

2) Possibilistic C-Means (PCM)

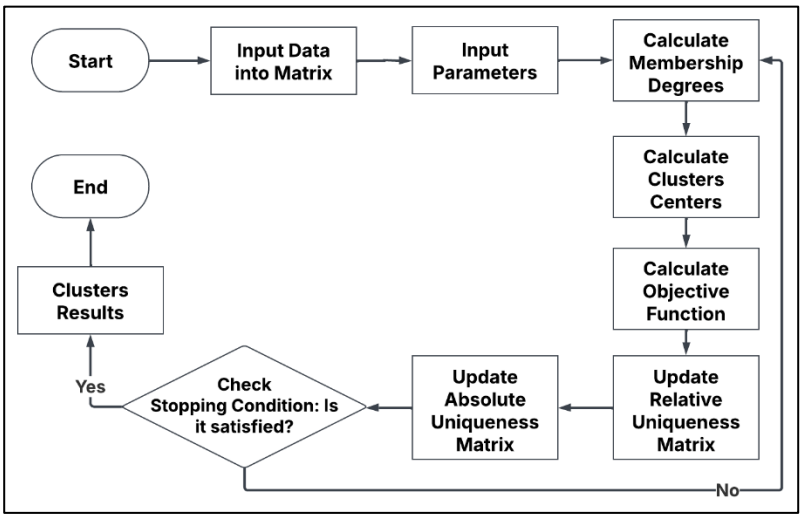

Figure 3. Flowchart of the Possibilistic C-Means (PCM) Algorithm

The Possibilistic C-Means (PCM) algorithm is an extension of Fuzzy C-Means (FCM) that addresses FCM's limitations in handling noisy or outlier data (Lestari et al., 2023). PCM assigns membership degrees based on the probability of each data point belonging to a

Literasi Nusantara. vol.5 no. 2, March – June 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

cluster, independently of other clusters, resulting in more stable clustering outcomes.

The workflow of the PCM algorithm implemented in this study is illustrated in Figure 3. The algorithm begins with inputting data into a matrix and initializing parameters, including the number of clusters, fuzzifier, and weighting power. It then calculates membership degrees, cluster centers, and the objective function. Uniqueness matrices—both relative and absolute—are updated iteratively, and the algorithm checks the stopping condition. Once the stopping condition is satisfied, final clusters are determined based on the highest membership degrees (Lestari et al., 2023).

e. Evaluation

To assess the quality of clustering results, this study employs three internal evaluation metrics: Partition Entropy (PE), Partition Coefficient (PC), and Modified Partition Coefficient (MPC) (Lestari et al., 2023). Partition Entropy (PE) measures the uncertainty or fuzziness of the clustering result. A lower PE value indicates clearer and more distinct cluster boundaries. The metric is defined in Equation (1):

$$PE = -\frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n}\mu_{ik}\ln(\mu_{ik}) \quad (1)$$

This function computes the negative average of the product of membership values and their logarithms. Lower values suggest reduced ambiguity and better clustering quality.

Partition Coefficient (PC) evaluates the sharpness of memberships in the fuzzy partition. Higher PC values, close to 1, indicate that each data point strongly belongs to a single cluster. PC is defined as shown in Equation (2):

$$PC = \frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n}\mu_{ik}^{2} \quad (2)$$

This equation calculates the average squared membership values across all clusters. Higher PC indicates a more confident clustering structure.

Modified Partition Coefficient (MPC) is an adjusted version of PC that addresses its scaling limitations. MPC normalizes the PC value so that it ranges between 0 and 1. It is defined in Equation (3):

$$MPC = 1 - \frac{c-1}{c}(1 - PC) \quad (3)$$

A value of MPC close to 1 indicates high clustering accuracy and low ambiguity. Conversely, values near 0 suggest that the clustering is weak or indistinct.

3. Result and Discussion

After completing the clustering process using Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM) algorithms, the results were analyzed and evaluated using three internal validation metrics: Partition Entropy (PE), Partition Coefficient (PC), and Modified Partition Coefficient (MPC). The following section presents the findings of the clustering process and discusses the comparative performance of FCM and PCM based on the evaluation metrics.

a. Clustering Results

1) Data Preparation and Normalization

Before the clustering process, the dataset was examined for completeness and quality. A review of all variables confirmed that no missing values were present, and therefore, no imputation was required at this stage.

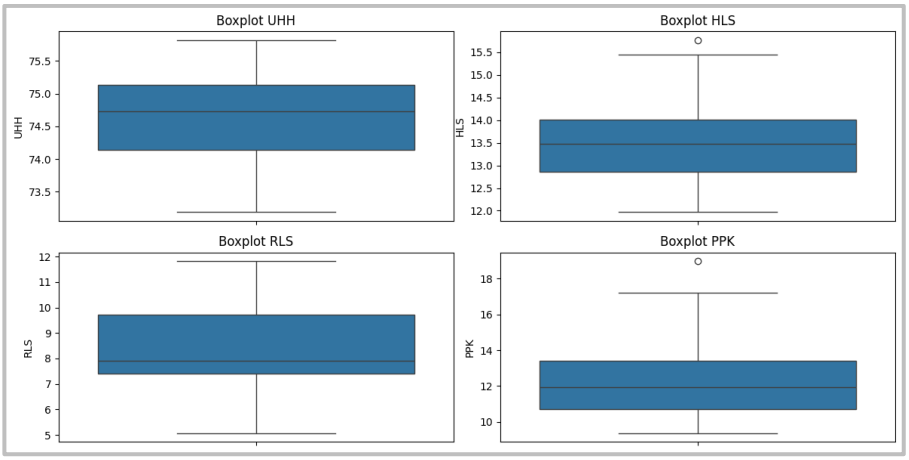

Figure 4. Boxplot Visualization of HDI Variables for Outlier Detection

Literasi Nusantara. vol.5 no. 2, March – June 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

To detect potential outliers, boxplot visualization was used. As shown in Figure 4, extreme values were observed in the variables Expected Years of Schooling (EYS) and Adjusted Per Capita Expenditure (APCE). Further investigation confirmed that these outliers came from Surabaya City and Malang City, respectively. To minimize distortion, the affected values were replaced using median imputation, as shown in Table 1.

Table 1. Boxplot Visualization of HDI Variables for Outlier Detection

Variable	District	Original Value	Treatment
EYS	Kota Surabaya	18.977	13.48
APCE	Kota Malang	15.77	11.952

After treating the outliers, Z-score normalization was applied to all four variables Life Expectancy (LE), Mean Years of Schooling (MYS), Expected Years of Schooling (EYS), and APCE to ensure all features were on the same scale. The transformed dataset had a mean of approximately 0 and a standard deviation of 1 across all variables. A sample of the normalized data is provided in Table 2.

Table 2. Sample of Normalized Data (Z-Score)

District/ City	UHH	HLS	RLS	PPK
Pacitan	-0.04	-0.97	-0.3	-1.23
Ponorogo	0.64	0.39	-0.36	-0.73
Trenggalek	0.77	-1.05	-0.29	-0.83
Tulungagung	0.53	-0.15	0.17	-0.27
Blitar	0.71	-1.01	-0.33	-0.31
Kediri	0.34	0.2	-0.08	-0.08
Malang	0.74	0.03	-0.38	-0.67
Lumajang	-0.28	-1.62	-0.75	-1.21
Jember	-0.82	0.04	-1.12	-0.93
Banyuwangi	-0.94	-0.42	-0.37	0.36

2) Parameter Optimization Using Bayesian Optimization

Bayesian Optimization was used to determine the optimal values for the number of clusters (c) and fuzzifier (m), targeting the minimization of the clustering objective function. The optimal parameters for both FCM and

PCM were identified as c = 3 and m =1.5, and these values were used for further analysis.

3) Clustering Execution and Summary

After parameter optimization, the clustering algorithms were executed using the normalized HDI dataset. Both Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM) were initialized with the optimized parameter c = 3, m = 1.5 identified through Bayesian Optimization. The convergence of both algorithms was assessed based on the behavior of the objective function across iterations. FCM reached convergence after 18 iterations, while PCM required 20 iterations. The final values of the objective functions are presented in Table 3.

Table 3. Final Objective Function Value and Iteration Count

Metode	Iteration Count	Final Objective Value
Fuzzy C-Means	18	43.0366
Possibilictic C-Means	20	75.7025

The results show that both methods successfully minimized their respective objective functions, indicating stable convergence. Although FCM achieved a slightly lower final objective value and faster convergence, these values do not directly reflect clustering quality and must be interpreted alongside evaluation metrics, which are discussed in the next section.

4) Cluster Centers and Interpretation

Upon convergence, the cluster centers (centroids) for each algorithm were calculated to represent the average normalized values of the HDI indicators Life Expectancy (LE), Mean Years of Schooling (MYS), Expected Years of Schooling (EYS), and Adjusted Per Capita Expenditure (APCE) within each cluster.

The centroids produced by Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM) are presented in Table 4 and Table 5, respectively. While the values are not identical due to differences in the membership and typicality computation mechanisms of each algorithm, they show highly similar

Literasi Nusantara. vol.5 no. 2, March – June 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

patterns across clusters. This indicates that both methods produced consistent clustering structures.

Tabel 4. Cluster Centroids from FCM

Klaster	UHH	HLS	RLS	PPK
1	0.2047	-0.3258	-0.1415	-0.1953
2	0.9648	1.2367	1.3767	1.2404
3	-1.3773	-0.5772	-1.1831	-0.9151

Tabel 5. Cluster Centroids from PCM

Klaster	UHH	HLS	RLS	PPK
1	0.2125	-0.3167	-0.1415	-0.1953
2	0.963	1.2457	1.3689	1.2326
3	-1.343	-0.5628	-1.1528	-0.8887

Based on the magnitude and direction of the centroid values, the three clusters can be interpreted in terms of human development levels. Cluster 1 represents regions with medium development, characterized by moderate life expectancy and slightly below average values in education and income indicators. Cluster 2 reflects very high development, with strong performance across all four indicators—life expectancy, years of schooling, and adjusted per capita expenditure—typically representing urban areas with advanced socio-economic conditions. In contrast, Cluster 3 indicates low development, where education and income variables are significantly below average, suggesting underdeveloped regions in need of targeted policy interventions. This interpretation offers a qualitative insight into the structure of the clustering results and reinforces the credibility of the grouping before moving on to the quantitative evaluation of clustering performance in the next section.

This interpretation offers a qualitative insight into the structure of the clustering results and reinforces the credibility of the grouping before moving on to the quantitative evaluation of clustering performance in the next section. The district/city groupings based on these cluster definitions are summarized in Table 6.

Table 6. District/City Grouping Based on Clustering Results

Cluster	Development Level	Districts/Cities
1	Medium Development	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Banyuwangi, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Kota Probolinggo
2	Very High Development	Sidoarjo, Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu
3	Low Development	Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Bangkalan, Sampang, Pamekasan, Sumenep

b. Evaluation Metrics Results

To assess the internal quality of the clustering results, three fuzzy clustering evaluation metrics were employed: Partition Coefficient (PC), Partition Entropy (PE), and Modified Partition Coefficient (MPC). These indices provide insight into the clarity, certainty, and overall structure of the fuzzy partitioning generated by Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM). The results of the evaluation are summarized in Table 6.

Table 6. Fuzzy Clustering Evaluation Metrics for FCM and PCM

Metode	PC	PE	,MPC
Fuzzy C-Means	0.8550	0.2693	0.7824
Possibilictic C-Means	0.8891	0.2058	0.8336

The Partition Coefficient (PC) measures the sharpness of cluster membership assignments. A value closer to 1 indicates that data points strongly belong to a single cluster. PCM achieved a PC of 0.8891, which is higher than FCM's 0.8550, suggesting that PCM produced more definitive cluster memberships.

The Partition Entropy (PE) quantifies the fuzziness or uncertainty of the cluster memberships. Lower values indicate clearer separation between clusters. PCM again outperformed FCM with a lower PE value (0.2058 vs. 0.2693), implying a better-defined clustering structure.

Literasi Nusantara. vol.5 no. 2, March – June 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

The Modified Partition Coefficient (MPC) adjusts the PC to account for the number of clusters, providing a normalized value between 0 and 1. PCM achieved a higher MPC (0.8336) compared to FCM (0.7824), reinforcing its superior performance in terms of clustering validity.

To enhance the interpretability of the results, Figures 4 and 5 present a bar chart comparison of each metric across both algorithms.

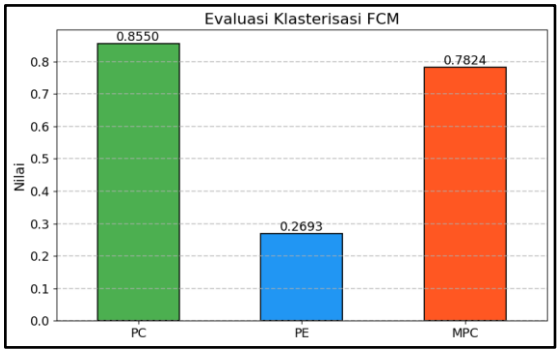

Figure 5. Evaluation Metrics – FCM

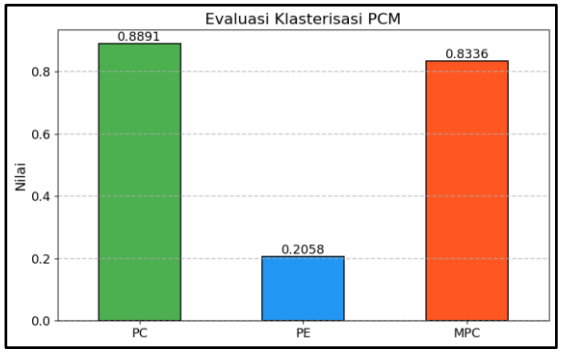

Figure 6. Evaluation Metrics – PCM

These results collectively demonstrate that PCM outperformed FCM across all three evaluation metrics, indicating that the clusters formed by PCM were more distinct and exhibited stronger membership clarity. While the computational time and convergence behavior of FCM were slightly better, the overall clustering quality based on fuzzy validity indices favors PCM.

c. Discussion of Practical Implications

The results of this study carry several practical implications, particularly for regional development planning in East Java. By applying fuzzy clustering methods to HDI data, policymakers can identify districts or cities with similar

development profiles—not only in absolute levels but also in nuanced socio-economic characteristics.

The findings suggest that regions classified in Cluster 3 (Low Development) exhibit consistent underperformance in education and income, requiring targeted investment in infrastructure, education, and economic programs. Meanwhile, Cluster 2 (Very High Development) comprises urban areas with strong performance across all indicators, which can serve as benchmarks or models for best practices.

Furthermore, the use of PCM over FCM offers a more robust understanding of development disparities, as it minimizes the ambiguity in classification. This is especially beneficial when designing interventions for transitional or borderline regions (e.g., cities that fall between clusters under different algorithms). The stronger cluster definition produced by PCM can improve precision in resource allocation and monitoring, supporting evidence-based governance.

Lastly, the clustering approach used in this study may be extended to other socio-economic indicators or regions. It enables scalable, data-driven segmentation without relying solely on fixed administrative boundaries or arbitrary thresholds.

4. Conclusion

This study evaluated the clustering results of Human Development Index (HDI) data in East Java using Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM) algorithms. Both methods successfully identified three clusters representing low, medium, and very high development levels. Although the clustering structures were largely similar, the internal evaluation metrics—Partition Coefficient (PC), Partition Entropy (PE), and Modified Partition Coefficient (MPC)—indicated that PCM produced superior clustering quality with clearer membership boundaries and lower uncertainty.

The findings suggest that PCM offers a more robust approach for analyzing socio-economic disparities, particularly when data involve ambiguity or

Literasi Nusantara. vol.5 no. 2, March – June 2025
ISSN 2746 8208 (Online) and ISSN 2746 3575 (Print)

overlapping characteristics. In contrast, FCM demonstrated faster convergence but slightly lower clustering precision. This insight provides guidance for future applications of fuzzy clustering, especially in policy-oriented domains where interpretability and accuracy are essential.

The broader implication of this research lies in its ability to support data-driven development planning. By clustering regions based on HDI indicators, policymakers can design more targeted and equitable interventions. Furthermore, the integration of Bayesian Optimization in parameter selection enhances methodological rigor and ensures optimal clustering performance, making this framework adaptable for similar analyses in other regional or national contexts.

References

Budiawan, J. C. (2024). Fuzzy Clustering untuk Text Dimensionality Reduction.

Faturahman, R. D., & Hidayati, N. (2025). Implementasi fuzzy c-means dalam pengelompokan tingkat kemiskinan pada kabupaten/kota di Provinsi Jawa Tengah. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 10(1), 137–149. https://doi.org/10.29100/jipi.v10i1.5747

Hartanto, S., Adzan, M. S., Haq, D. Z., & Novitasari, D. C. R. (2024). Analisis indeks pembangunan manusia di Jawa Timur tahun 2022–2023 berdasarkan indikator menggunakan metode fuzzy c-means. *INTEK: Jurnal Informatika dan Teknologi Informasi,* 7(2), 45–54. https://doi.org/10.37729/intek.v7i2.5358

Hendrastuty, N. (2024). Penerapan data mining menggunakan algoritma k-means clustering dalam evaluasi hasil pembelajaran siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Huddin, S. (2023). Penerapan fuzzy c-means pada klasterisasi penerima bantuan pangan non tunai. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(1), 453–461.

Isah, I., Estri, M. N., & Larasati, N. (2024, October). Perbandingan metode fuzzy c-means dan fuzzy probabilistic c-means dalam pengelompokan provinsi di Indonesia berdasarkan indeks pembangunan manusia tahun 2022. *Prosiding Seminar Nasional Sains Data,* 4(1), 832–841. https://doi.org/10.33005/senada.v4i1.346

Kadja, M. T. G., Rumlaklak, N. D., & Djahi, B. S. (2023). Penerapan metode fuzzy c-means dalam penentuan penerima beasiswa program Indonesia pintar (PIP). *J-ICON: Jurnal Komputer dan Informatika*, 11(1), 37–43. https://doi.org/10.35508/jicon.v11i1.9846

Lestari, M., Kartini, D., Budiman, I., Faisal, M. R., & Muliadi, M. (2023). Comparison of industrial business grouping using fuzzy c-means and fuzzy possibilistic c-means methods. *Telematika,* 16(2), 91–102. https://doi.org/10.35671/telematika.v16i2.2548

Mubarok, M. N. (2024). Implementasi fuzzy possibilistic c-means dengan validitas modified partition coefficient pada clustering prevalensi balita stunting di Indonesia *(Doctoral dissertation*, Universitas Islam Negeri Maulana Malik Ibrahim).

Putra, R. F., Zebua, R. S. Y., Budiman, B., Rahayu, P. W., Bangsa, M. T. A., Zulfadhilah, M., ... & Andiyan, A. (2023). *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia.

Rahayu, P. W., Sudipa, I. G. I., Suryani, S., Surachman, A., Ridwan, A., Darmawiguna, I. G. M., ... & Maysanjaya, I. M. D. (2024*). Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia.

Sari, A. I. C., A'ini, Z. F., & Tukiran, M. (2022). Pengaruh anggaran pendidikan dan kesehatan terhadap indeks pembangunan manusia di Indonesia. *JABE (Journal of Applied Business and Economic),* 9(2), 127–136. https://doi.org/10.30998/jabe.v9i2.15082

Setiawan, Z., Fajar, M., Priyatno, A. M., Putri, A. Y. P., Aryuni, M., Yuliyanti, S., ... & Wijaya, A. (2023). *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia.

Simanjuntak, T. F. B., Zuhriadi, M., Habeahan, J., Lubis, R. J., Hutapea, T. P. U., & Sirait, M. M. (2024). Pengaruh angka harapan hidup dan kemiskinan terhadap indeks pembangunan manusia di Indonesia. *Jurnal Intelek dan Cendikiawan Nusantara*, 1(3), 4012–4019.

Sudipa, I. G. I., Sarasvananda, I. B. G., Prayitno, H., Putra, I. N. T. A., Darmawan, R., & WP, D. A. (2023). *Teknik Visualisasi Data*. PT. Sonpedia Publishing Indonesia.

Thakkar, H., Shah, V., Yagnik, H., & Shah, M. (2021). Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis. *Clinical eHealth*, 4, 12–23. https://doi.org/10.1016/j.ceh.2020.11.001

Yasir, A., & Firmansyah, A. U. (2024). Implementasi metode fuzzy c-means dan metode AHP dalam pemilihan promosi jabatan karyawan berbasis web (studi kasus: PT. Tunas Dwipa Matra Sekampung*). Journal of Science and Social Research,* 7(4), 1616–1619.

Zai, C. (2022). Implementasi data mining sebagai pengolahan data. *Jurnal Portal Data,* 2(3).

Zhang, C., Xue, J., & Gu, X. (2022). An online weighted Bayesian fuzzy clustering method for large medical data sets. *Computational Intelligence and Neuroscience*, 2022, 6168785. https://doi.org/10.1155/2022/6168785