IJSA (2025), Vol. 9, No. 2, 274-288
DOI: https://doi.org/10.29244/ijsa.v9i2p274-288

Indonesian Journal of Statistics and Its Applications
Available at https://journal-stats.ipb.ac.id/index.php/ijsa
e-ISSN: 2599-0802

Optimization of Fuzzy C-Means Clustering with Particle Swarm Optimization on Socioeconomic Indicators of ASEAN Countries

Cindy Indriyani[1], Siti Arbaynah[1], Ananda Putra Wijaya[1], Lusi Oktaviani[1], Fadhilah Yumna[1], Norashida Othman[2], and Sachnaz Desta Oktarina[1‡], and Rahma Anisa[1]

[1]Study Program on Statistics and Data Science, IPB University, Indonesia
[2]Department of Economics and Financial Studies, Universiti Teknologi MARA, Malaysia
[‡]corresponding author: sachnazdes@apps.ipb.ac.id

[*]Abstract

Grouping data based on similarity in characteristics is commonly applied in various exploratory analyses. The Fuzzy C-Means algorithm offers flexibility through the degree of membership of data points in each cluster, but it is vulnerable to poor cluster center initialization, which increases the risk of getting trapped in local optima. To enhance the performance of Fuzzy C-Means, this study integrates the Particle Swarm Optimization method for determining cluster centers. The evaluation is conducted by comparing Fuzzy C-Means and Fuzzy C-Means-Particle Swarm Optimization across several cluster counts using three internal validation metrics, namely the silhouette coefficient, partition coefficient, and Xie-Beni Index. The results show that Fuzzy C-Means-Particle Swarm Optimization consistently yields higher silhouette coefficient and partition coefficient values, along with lower Xie-Beni Index values, compared to standard Fuzzy C-Means. This indicates that the integration of Particle Swarm Optimization can improve clustering quality in terms of cluster compactness and separation. This hybrid approach demonstrates significant potential in complex data clustering scenarios.

Keywords: clustering, cluster center optimization, fuzzy c-means, internal validation, particle swarm optimization.

[*] Received: Nov 2025; Reviewed: Nov 2025; Published: Dec 2025

1. Introduction

The ASEAN (Association of Southeast Asian Nations) region is one of the fastest-growing economic blocs in the world. According to World Economics (2023), over the past decade—from 2012 to 2022—ASEAN countries have played a significant role in the global economy, contributing approximately 7% of the world's total gross domestic product and accounting for 9% of global economic growth. These figures not only reflect the region's growing economic capabilities but also indicate a significant shift in the global economic center from developed nations in the northern hemisphere to Asia, particularly Southeast Asia.

The establishment of the ASEAN Economic Community (AEC), also known as the ASEAN Economic Community (AEC), represents ASEAN's integration into the global economy through participation in global supply chains as an effort to reduce development gaps (Ishikawa, 2021). However, behind the narrative of economic integration lies a structural challenge that cannot be overlooked—the high level of social and economic disparity among member states. To address this, one possible approach is to group ASEAN countries based on social and economic indicators. This cluster analysis can serve as a strategic tool for designing and directing policies implemented by ASEAN stakeholders, particularly in setting development priorities. By mapping countries based on similar socio-economic conditions, it becomes possible to identify which nations require greater attention, thereby allowing for more targeted and effective interventions.

In the application of cluster analysis, especially on high-dimensional data, fuzzy-based approaches have proven to be more effective in handling soft clustering (Dogan & Avvad, 2025). One commonly used soft clustering method is Fuzzy C-Means (FCM), as it allows each data point to belong to more than one cluster, with membership levels expressed as continuous values between 0 and 1. This degree of membership offers flexibility in analysis by allowing observations to have a probabilistic association with multiple clusters simultaneously (Ishikawa, 2021). Although Fuzzy C-Means (FCM) is widely used in various cluster analyses, the algorithm has a weakness—it is prone to getting trapped in local optima, which can affect the accuracy of selecting initial cluster centers. Therefore, the Particle Swarm Optimization (PSO) algorithm is employed to optimize the FCM algorithm, with the goal of improving the accuracy of initial cluster center selection in FCM, thereby achieving globally optimal clustering results (Yu et al., 2023).

A previous study conducted by Xia et al., (2019) proposed an enhanced PSO algorithm by taking into account both local and global optimization capabilities. The improved PSO algorithm was used to address the weakness of FCM in manually selecting initial cluster centers. The results of the study indicated that the enhanced PSO algorithm not only successfully avoided convergence to local optima but also achieved higher accuracy and better performance in handling noise compared to the traditional FCM algorithm. Furthermore, a study by (Febriyanti et al., 2023) on the clustering of regencies/cities in Kalimantan Island based on community welfare indicators showed that using the FCM algorithm optimized with PSO resulted in a more optimal cluster validity index.

Given the importance of enhancing development in the ASEAN region, this study aims to cluster ASEAN countries based on socio-economic indicators that influence development. The methods used in this study are Fuzzy C-Means (FCM) and Fuzzy C-Means optimized using Particle Swarm Optimization (PSO). Subsequently, the clustering results from both methods are compared to determine the most optimal cluster validity index.

2. Methodology

2.1 Data

The data used in this study are secondary data obtained from The World Bank at https://data.worldbank.org/, Human Development Reports at https://hdr.undp.org/, and statista.com at https://www.statista.com/ in 2022. The data includes socio-economic information from ten countries in the Southeast Asia region with a total of 10 observations and 10 numeric variables. Details of the variables used in the analysis are presented in Table 1. The entire data analysis process was carried out using RStudio software.

Table 1: Details of the variables used in the study.

Variables	Description	Unit	Reference
X_1	Gross Domestic Product (GDP) per capita	Dollar	(Hartini, 2017)
X_2	Unemployment rate	Percent	(Azzahra et al., 2024)
X_3	Life expectancy	Years	(Simanjuntak et al., 2024)
X_4	Infant mortality rate	Percent	(Simanjuntak et al., 2024)
X_5	Internet users	Percent	(Rizqi & Sidiq, 2023)
X_6	Urbanization rate	Percent	(Shinta, 2024)
X_7	Government health expenditure	Percent	(Nasution et al., 2020)
X_8	Expected years of schooling	Percent	(Hepi & Zakiah, 2018)
X_9	Mean years of schooling	Years	(Hepi & Zakiah, 2018)
X_{10}	Inflation rate	Percent	(Salim & Fadilla, 2021)

2.2 Research Methods

The stages of analysis in research are as follows:

1. Conducting data exploration using a correlation matrix to detect possible multicollinearity between variables, boxplots to identify outliers, and descriptive statistics to obtain an overview of the characteristics of the data.
2. Standardizing data using the z-score method so that all variables have an average of 0 and a standard deviation of 1. The z-score standardization can be obtained from equation 1 with x being the data value, μ being the average, and σ being the standard deviation.

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

3. Conducting principal component analysis (PCA) if possible multicollinearity is found. Principal Component Analysis (PCA) or PCA is carried out to reduce the dimensions of the data while maintaining most of the variations in the original data (Eliyanto & Suparman, 2019). PCA forms principal components as linear combinations of uncorrelated original variables (Mayapada et al., 2019) and is obtained through the decomposition of the covariance matrix into values and eigenvectors (Nasution, 2019). The resulting principal components can be used as new predictor variables with the selection of the number of components based on the proportion of explained variance, usually 70–85% (Jolliffe, 2002). Main components (KU) can be obtained from equation 2 with $j = 1, 2, 3, \ldots, p$ and $e_{1j},\ e_{2j}, \ldots, e_{pj}$ are elements of the eigenvector corresponding to KU (Y_j) (Rosyada & Utari, 2024).
$$Y_j = e_{1j}X_1 + e_{2j}X_2 + \cdots + e_{pj}X_p \tag{2}$$
4. Determine the initial parameters before optimization such as the number of clusters (c), inertia weight (z), fuzzy weight (m), maximum speed (v_{max}), and the maximum number of iterations.
5. Using the FCM-PSO algorithm to find the optimal combination of parameters (c and m). FCM is a partition-based clustering algorithm that allows each data to have a degree of membership to more than one cluster (Eliyanto & Suparman, 2019). FCM operates by minimizing an objective function that depends on the distance between the data point and the cluster center and the fuzzy degree of membership of each data (Bezdek et al., 1981; Sanusi et al., 2020). This algorithm groups data based on the level of membership of each data in a particular group with a value between 0 and 1 (Sa'diyah et al., 2020). The higher the membership value, the greater the similarity of the data to the related group. The objective function used in FCM to determine the best clustering can be written as follows (Sa'diyah et al., 2020):
$$J(U,V) = \sum_{j=1}^{m} \sum_{i=1}^{n} (u_{ij})^m {d_{ij}}^2 \tag{3}$$
where u_{ij} indicates the degree of membership of the i-th data in the j-th cluster, m is the fuzzifier value that regulates the level of cluster fuzziness, and d_{ij} is the distance between the data and the cluster center. The iteration process in FCM aims to minimize the value of the objective function so that the clustering results are optimal. The FCM algorithm can be explained as follows (Sa'diyah et al., 2020):
   a. Determine the initial parameters, namely the number of cluster c, the fuzzifier value m, the smallest error threshold, and the maximum iteration limit.
   b. Initialize the membership matrix $U = \{u_{ij}\}$ of size $c \times n$ randomly.
   c. Calculate the cluster center $V = v_i$ using the following equation 4:
$$v_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j}{\sum_{j=1}^{n} (u_{ij})^m} \tag{4}$$
   d. Evaluate the objective function $J(U,V)$ using the formula in equation 3.
   e. Update the membership matrix U based on the following equation 5:

$$u_{ij} = \frac{[\sum_{i=1}^{n}(d_{ij})]^{\frac{-1}{m-1}}}{\sum_{j=1}^{c}[\sum_{i=1}^{n}(d_{ij})]^{\frac{-1}{m-1}}} \quad (5)$$

PSO is a population-based optimization algorithm first introduced by Kennedy and Eberhart in 1995 (Kennedy & Eberhart, 1995). In the PSO algorithm, each potential solution is represented as a particle that has two main attributes, namely position (X) and speed or velocity (V) (Sa'diyah et al., 2020). At each iteration, the position and velocity of the particle will be updated to approach the optimal solution. Each particle has a memory of the best position it has ever reached (P_b) and accesses the best position of all particles in the population (g_b). The process of updating the speed and position of each particle follows the following formula:

$$p_b = \arg\min f(X_i(t)) \quad (6)$$

$$g_b = \arg\min f(p_b(t)) \quad (7)$$

$$V_i(t+1) = w \times V_i(t) + c_1 \times r_1 \times (p_b - X_i(t)) + c_2 \times r_2 \times (g_b - X_i(t)) \quad (8)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (9)$$

where $\arg\min f(X_i(t))$ is the position of particle X_i at $f(X_i(t))$ which is minimum until iteration t, w is the inertia weight that controls the influence of previous speed, c_1 and c_2 are learning factors that indicate how much influence individual and social experience has, r_1 and r_2 are random numbers between 0 and 1, and P_b and g_b are the personal and global best positions.

Inertia weights play an important role in balancing exploration (global search) and exploitation (local search). This weight can be linearly decreased as iterations increase to encourage convergence. The adjustment of the inertia weight is done with the following equation:

$$w(i) = w_{max} - \frac{w_{max} - w_{min}}{iterasi_{max}} \times iterasi \quad (10)$$

where $w(i)$ is the inertia weight at iteration i and $iter_{max}$ is the maximum number of iterations. The FCM-PSO algorithm can be explained by the following steps:

a. Initialize the population membership matrix U.
b. Calculate the cluster center of each particle.
c. Calculate the objective function of each particle.
d. Calculate the fitness value $f_s = \frac{1}{J(U,V)}$ based on research conducted by Febriyanti et al., (2023).
e. Determine the best individual ($pbest$) and global ($gbest$) values.
f. Update the particle speed and position.
g. Stop iteration if the value $|gbest - gbest_{t-1}| < \varepsilon$. If not, repeat the process from the second stage.
h. Take the optimal parameters from the $gbest$ value.

6. Perform the clustering process with the FCM algorithm using the optimal parameters from PSO, then compare it with the FCM results using random parameters.

7. Evaluate the validity of the cluster formed with the silhouette coefficient (SC), partition coefficient (PC), and Xie-Beni Index (XBI) values.
   a. SC is used to measure how well data fits its own cluster compared to the closest other cluster. The SC value is in the range of −1 to 1 with values close to 1 indicating that the object is very suitable for its cluster, while negative values indicate the possibility of incorrect clustering. SC is calculated based on the following equation 11 (Rahmawati et al., 2024):
   $$S_i = \frac{b_i - a_i}{MAX(a_i, b_i)} \tag{11}$$
   where a_i is the average distance between the i-th and other data in the same cluster, b_i is the smallest average distance between the i-th data and data in other different clusters, and $MAX(a_i, b_i)$ indicates the maximum value between the average distance within the cluster itself and the average distance with the nearest different cluster.
   b. PC is a measure used to assess the level of clarity of partitions in the fuzzy clustering process (Kalla et al., 2022). This index represents the extent to which an object's membership is in a particular cluster. The higher the PC value (approaching 1), the more deterministic or clear the partition formed. The PC value is in the range of $\frac{1}{c}$ to c with c stating the number of clusters used, N stating the number of research objects, and μ_{ik} stating the degree of membership of the k-th object to the i-th cluster center. The PC calculation is carried out based on equation 12 as stated by (Bezdek et al., 1981).
   $$PC(c) = \frac{1}{N}\sum_{i=1}^{C}\sum_{k=1}^{N} {\mu_{ik}}^2 \tag{12}$$
   a. XBI aims to calculate the ratio between the total variation within a group (intra-cluster compactness) and the level of separation between groups (inter-cluster separation) in fuzzy clustering (Widiyanto, 2019). This index evaluates the quality of the partition by considering the balance between the proximity of data in one cluster and the separation between clusters. The XBI value does not have a standard range. However, the smaller the value of this index, the better the quality of the cluster formed, namely more compact internally and more separated externally. The XBI formula was formulated by (Xie & Beni, 1991) and is shown in equation 13.
   $$XB(c) = \sum_{i=1}^{c} \frac{\sum_{i=1}^{C}\sum_{k=1}^{N} {\mu_{ik}}^m \, ||x_k - v_i||^2}{N min_{ik}||v_i - v_k||^2} \tag{13}$$
   where N represents the total number of objects in the study and c is the number of clusters used. The notation μ_{ik} represents the degree of membership of the k-th object to the i-th cluster center. The parameter m is a fuzzifier, which is a constant that regulates the level of fuzziness in the clustering process. The symbol $||x_k - v_i||$ shows the euclidean distance between data point x_k and cluster center v_i, while $||v_i - v_k||$ indicates the euclidean distance between the k-th and i-th cluster center.
8. Interpret and visualize clustering results.

3. Results and Discussions

3.1 Data Exploration

Data exploration was conducted to detect multicollinearity, identify outliers, and obtain a general overview of the characteristics of the data. Boxplot was used to observe the distribution pattern of a variable. Boxplots can also indicate the presence of outliers in the data. Figure 1 presents the boxplots of variables X_1 to X_5.

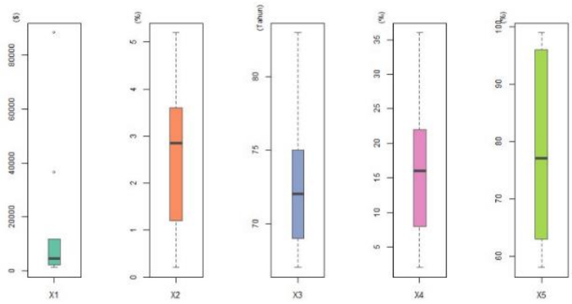


Figure 1: Boxplot of variables $\boldsymbol{X_1}$ to $\boldsymbol{X_5}$.

Based on Figure 1, it can be concluded that only X_1 contains outliers and its distributions is skewed to the right. In contrast, variables X_2 to X_5 do not have any outliers. The distributions of X_2 to X_5 are also relatively symmetrical. Next, the distributions of variables X_2 to X_5 will be presented in Figure 2.

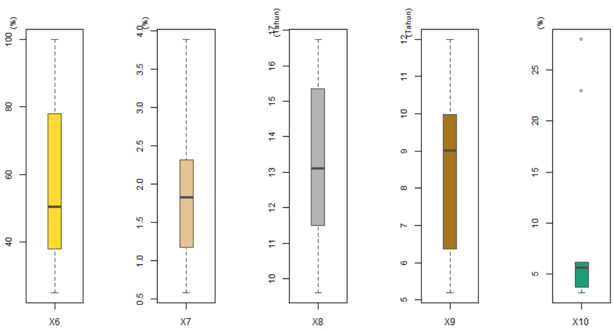


Figure 2: Boxplot of variables $\boldsymbol{X_6}$ to $\boldsymbol{X_{10}}$.

In Figure 2, it can be seen that X_6, X_8, and X_9 have relatively symmetrical distributions and do not contain outliers. Meanwhile, X_7 has a less symmetrical distribution, tending to be skewed to the right, but it does not have any outliers. X_{10} has a highly right-skewed distribution and also contains upper outliers. In addition to

exploration using boxplots, further analysis was carried out using a correlation matrix to examine the relationships among the independent variables. The correlation matrix is used to detect multicollinearity among the independent variables. The stronger the correlation values, the higher the tendency for multicollinearity to occur. Figure 3 below presents the correlation matrix among the independent variables.

Figure 3: Correlation matrix between independent variables.

The correlation matrix in Figure 3 shows that the strongest relationship among the independent variables is between X_4 and X_5, with a correlation of –0,88. The next strongest correlations are between X_5 and X_6 at 0,86, and between X_6 and X_9 at 0,86. The high correlations among these four variables may indicate the presence of multicollinearity. One way to address multicollinearity is through Principal Component Analysis (PCA). Table 2 presents the results of the PCA.

Table 2: PCA results.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**
Proportion of variance	0,6707	0,1497	0,07708	0,04251	0,0283
Cumulative proportion	0,6707	0,8204	0,89749	0,9400	0,9683

Based on Table 2, the first principal component (PC) is able to explain approximately 67% of the total data variance. Subsequently, the second principal component increases the proportion of explained variance to 82%. The principal components used are up to the third PC, as the total explained variance reaches 89% with the third component. Meanwhile, the fourth principal component shows a decreased gain in explained variance, and thus is not included.

3.2 Clustering Process

The initial process in applying the FCM method combined with PSO begins with determining the basic parameters that play an important role in the clustering process. These parameters include the number of clusters (c), the degree of fuzziness (m), the inertia weight (w), the maximum particle movement speed (v_{max}), and the maximum number of iterations. This study uses parameters ranging from 2 to 5 clusters (c = 2, 3, 4, 5) fuzzy weights from 1 to 3, inertia weight (w = 5) , maximum velocity (v_{max}= 5) convergence threshold 0,5, and a maximum of 1.000 iterations. The selection of the cluster number range (c) between 2 and 5 is based on an analysis of ten Southeast Asian countries that are the objects of this study, with the aim of obtaining optimal clustering results. Meanwhile, the variation in the value of m is used to explore the most suitable fuzzy parameter, considering that the value of m affects the degree of membership in fuzzy clustering.

The optimization process using the FCM-PSO algorithm is carried out for a maximum of 1.000 iterations or until convergence is achieved. During the iterations, each particle evaluates the fitness value by applying FCM to the data using the selected parameters. These particles then update their positions based on their individual best information (p_{best}), and the best information from the entire population (g_{best}), with the aim of finding the parameter combination that produces the most optimal clustering. In this study, convergence was achieved at iteration 122, where the fitness value showed stability at the minimum point, indicating that the algorithm successfully found an effective solution. The process of finding the optimal solution by PSO is shown in Figure 4.

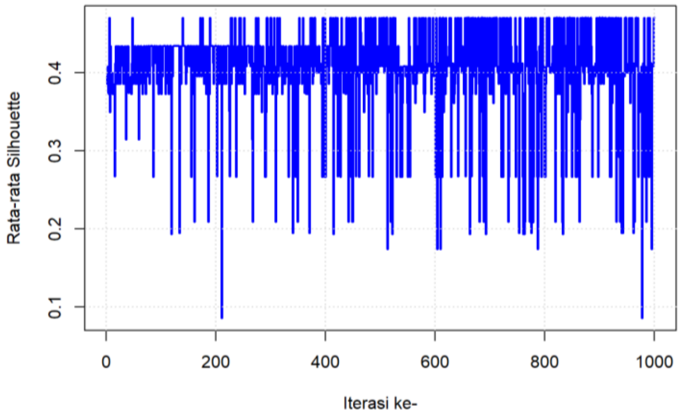
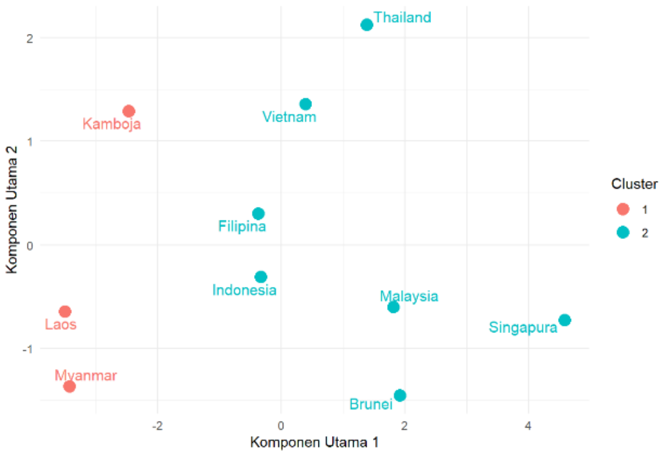


Figure 4: Graph of convergence values for PSO optimal solution search.

The optimal parameters obtained are the optimal number of clusters ($c = 5$) and fuzzy weight (m = 2,18). These parameters are used as the primary measures in the clustering process. Additionally, during the clustering process, analyses were conducted with 2, 3, and 4 clusters. The purpose of this approach is to compare the clustering results obtained using the PSO optimisation parameters with the results obtained when only using the FCM method without optimisation. The addition of 2, 3, and 4 clusters was added as random parameters. The fuzzy weight value of m = 2,18 was applied throughout the clustering process, as this value was proven to yield the best performance among the various parameter values explored previously and aligns with the optimization results obtained through the PSO algorithm.

3.2.1. Clustering with c = 2

Clustering with the FCM algorithm was carried out using parameters c = 2 and m = 2,18 to group ASEAN countries based on socio-economic indicators. The results of this clustering are visualized by the biplot in Figure 5 using three main components of PCA. Based on the biplot, cluster 1 shows that the countries in it have similar socio-economic conditions and tend to be at a lower level



compared to other countries in the ASEAN region. Meanwhile, countries in cluster 2 are in a wider distribution of positions in the two-dimensional space. This cluster tends to have a better level of socio-economic development.

Figure 5: Two-dimensional plot of PCA results with c = 2.

3.2.2. Clustering with c = 3

Clustering with the FCM algorithm was also carried out using parameters c = 2 and m = 2,18 to group ASEAN countries based on socio-economic indicators. The results of this clustering are visualized by the biplot in Figure 6 using three main components of PCA. Based on the biplot, it can be seen that cluster 1 has low socio-economic conditions. Cluster 2 shows high and good socio-economic conditions. Meanwhile, cluster 3 consists of countries that are at the middle level and show quite good development but are not yet on par with cluster 2.

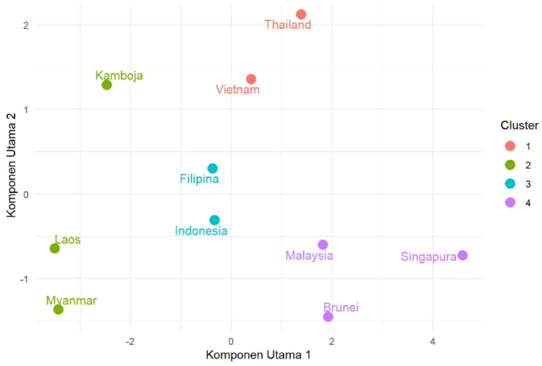


Figure 6: Two-dimensional plot of PCA results with c = 3.

3.2.3. Clustering with c = 4

Clustering with the FCM algorithm was also carried out using parameters c = 4 and m = 2,18 to group ASEAN countries based on socio-economic indicators. The results of this clustering are visualized by the biplot in Figure 7 using three main components of PCA. Based on the biplot, cluster 1 shows relatively good socio-economic characteristics and tends to be superior compared to other countries. Cluster 2 describes countries with a relatively low level of socio-economic development. Cluster 3 shows countries in the middle position which reflects socio-economic conditions that are developing and have the potential to move up to a higher class. On the other hand, cluster 4 shows high values in the main components so that the socio-economic conditions in its countries tend to be good and stable.

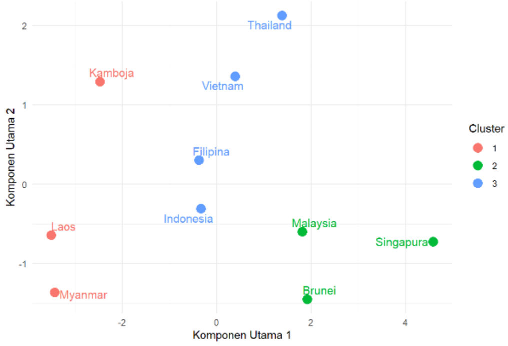


Figure 7: Two-dimensional plot of PCA results with c = 4.

3.2.4. Clustering with c = 5

Figure 8 shows the visualization results of the clustering of ASEAN countries using FCM that has been optimized with the best parameters of PSO. This clustering process produces five main clusters based on the similarity of characteristics of socio-economic conditions in ASEAN. Cluster 1 represents countries with medium socio-economic conditions. Cluster 2 reflects a group of countries with a relatively high socio-economic level. Cluster 3 consists of countries with socio-economic conditions that tend to be left behind. Meanwhile, cluster 4 includes countries with developing socio-economic conditions and cluster 5 includes countries with a very high level of socio-economic development.

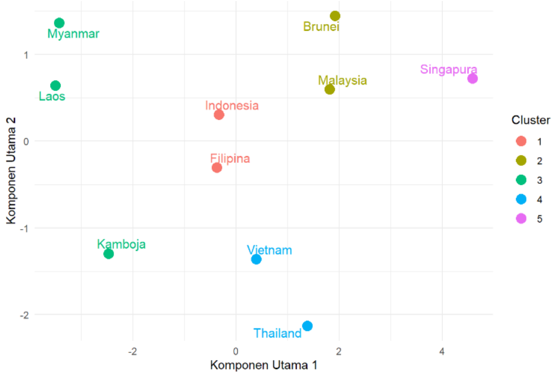


Figure 8: Two-dimensional plot of PCA results with c = 5.

3.3 Cluster Evaluation

The determination of the optimal cluster was conducted by comparing the values of the Silhouette Index (SI), Partition Coefficient (PC), and Xie-Beni Index (XBI). Based on the clustering evaluation results, the grouping that used the optimal parameters obtained from PSO, namely (c = 5, m = 2,18), demonstrated the best performance among the tested scenarios. These results confirm that the PSO algorithm can be effectively utilized as an optimization mechanism to enhance the performance of the FCM algorithm in the clustering process (Yu et al., 2023).

Table 3: Cluster evaluation on the FCM-PSO algorithm and the FCM algorithm.

Validity Metrics	**FCM-PSO (c = 5)**	**FCM (c = 4)**	**FCM (c = 3)**	**FCM (c = 2)**
Silhouette Coefficient	**0,4702**	0,3724	0,3857	0,4343
Partition Coefficient	**0,7610**	0,6956	0,6645	0,7070
Xie-Beni Index	**0,0579**	0,1346	0,1866	0,1536

Based on the table, the highest Silhouette Coefficient value of 0,4702 obtained by the FCM-PSO method (c = 5) indicates that the observational data is well-separated between clusters and exhibits high similarity within each cluster. Additionally, the

Partition Coefficient also reaches its maximum value of 0,7610 in the FCM-PSO method, indicating that the formed clusters do not overlap and have clear membership levels. Meanwhile, the lowest Xie-Beni Index value of 0,0579 achieved by FCM-PSO suggests that the clustering has more optimal separation, as the within-cluster variation is smaller and the between-cluster differences are greater. This clustering evaluation demonstrates that the PSO algorithm effectively optimizes the parameters, thereby significantly enhancing the clustering performance of the FCM method.

3.4 Interpretation of Best Clustering Result

The best clustering result was obtained when c = 5 based on the PSO outcome. The five resulting clusters each contain between 1 to 3 members with relatively similar characteristics. Cluster 1 consists of Indonesia and the Philippines. Cluster 2 consists of Brunei Darussalam and Malaysia. Cluster 3 includes Myanmar, Laos, and Cambodia. Cluster 4 is composed of Vietnam and Thailand. Lastly, cluster 5 consists solely of Singapore. Table 4 presents the characteristics of each formed cluster.

Table 4: Cluster characteristics based on social and economic indicators.

Characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
GDP per capita($)	Low	High	Very low	Moderate	Very high
Unemployment rate(%)	Moderate	Very High	Low	Very low	High
Life expectancy (years)	Low	Moderate	Very low	High	Very high
Infant mortality rate(%)	High	Low	Very high	Moderate	Very low
Internet users(%)	Low	Very high	Very low	Moderate	High
Urbanization rate(%)	Moderate	High	Very low	Low	Very high
Government health expenditure(%)	Low	Low	Very low	Very high	High
Expected years of schooling(years)	Low	Low	Very low	High	Very high
Mean years of schooling(years)	Moderate	High	Very low	Low	Very high
Inflation rate(%)	Moderate	Very low	Very high	Low	High

Based on the characteristics of each cluster presented in Table 4, cluster 1, which consists of Indonesia and the Philippines, has economic and social indicators that are generally in the low to moderate range. The high infant mortality rate and low life expectancy indicate challenges in the health sector. Cluster 2, which consists of Malaysia and Brunei, has high levels of internet access, urbanization, and GDP, reflecting strong infrastructure. However, the countries in cluster 2 face high unemployment rates, indicating labor market challenges.

Cluster 3, consisting of Myanmar, Laos, and Cambodia, has very low economic and social indicators. These low indicators reflect countries with severely limited social and economic infrastructure. The countries in cluster 3 can be categorized as least developed countries (LDCs). In contrast to cluster 3, cluster 4 consists of stable developing countries. Nations such as Thailand and Vietnam demonstrate a good balance between economic and social aspects. High health expenditure and life expectancy reflect a strong focus on well-being. However, urbanization and education still have room for improvement. Lastly, cluster 5, which includes only one country—Singapore—is classified as a developed country. This cluster displays excellent socio-economic indicators, reflecting a high quality of life, strong infrastructure, and optimal human development.

The integration of multidimensional socioeconomic indicators within an optimized fuzzy clustering framework captures the heterogeneity of development levels among ASEAN countries, ranging from least developed to highly developed economies, which is not always represented by rigid clustering approaches. The resulting clusters reflect varying stages of socioeconomic development with overlapping characteristics across countries. Compared to standard Fuzzy C-Means, the FCM–PSO approach demonstrates improved cluster quality for this study, as indicated by higher Silhouette Index and Partition Coefficient values and a lower Xie–Beni Index. In contrast to previous studies that relied on limited indicators and rigid cluster structures (Afrianita et al., 2024; Arisman, 2018), this study provides a more flexible and nuanced representation of ASEAN socioeconomic patterns through the incorporation of a broader set of indicators and a fuzzy clustering framework.

4. Conclusion and Suggestion

4.1 Conclusion

The integration of the Particle Swarm Optimization (PSO) algorithm into the Fuzzy C-Means (FCM) method provides a significant improvement in the quality of clustering results. This is proven through three main evaluation metrics. First, the silhouette coefficient value increased by 26,25%, indicating that the cluster structure became clearer and more distinct between clusters. Second, the partition coefficient value also showed an increase in the certainty of data membership to clusters. Third, the Xie-Beni index value decreased by 68,96%. Overall, these results reinforce the effectiveness of PSO in optimizing cluster centers in FCM. The best results were obtained with five clusters, each containing 1 to 3 countries with similar socioeconomic characteristics. The clusters reflect the actual conditions of countries, ranging from underdeveloped developing countries to developed countries, with Singapore as the sole member of the most advanced cluster.

4.2 Suggestion

Further research is recommended to test the combination of FCM-PSO on various types of data and different numbers of clusters to determine the consistency of its performance. In addition, exploration of other optimization algorithms such as genetic algorithms or differential evolution can also be carried out to compare their effectiveness with PSO. Improving computational efficiency and sensitivity analysis of PSO parameters can also be the focus of further development.

References

Afrianita, S., Ridwansyah, M., & Umiyati, E. (2024). Kesejahteraan masyarakat ASEAN : dimensi beberapa variabel makroekonomi dan kependudukan. *Jurnal Sumberdaya Dan Lingkungan*, *13*(2), 79–92.

Arisman, A. (2018). Determinant of human development index in ASEAN countries.

Signifikan: Jurnal Ilmu Ekonomi, *7*(1), 113–122. https://doi.org/10.15408/sjie.v7i1.6756

ASEAN States. (2023). https://www.worldeconomics.com/Regions/ASEAN/

Azzahra, S. F., Putri, L. D., Purba, F. Y., Tanjung, D., Rezkitaputri, A., & Zulva, R. Z. D. (2024). Dampak pengangguran terhadap stabilitas sosial dan perekonomian Indonesia. *MENAWAN : Jurnal Riset Dan Publikasi Ilmu Ekonomi*, *2*(4), 220–233. https://doi.org/10.61132/menawan.v2i4.719

Bezdek, J. C., Ehrlich, R., & Full, W. (1981). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2–3), 191–203. https://doi.org/10.1016/0098-3004(84)90020-7

Dogan, O., & Avvad, H. (2025). Fuzzy clustering based on activity sequence and cycle time in process mining. *Axioms*, *14*(5), 1–22. https://doi.org/10.3390/axioms14050351

Eliyanto, J., & Suparman. (2019). Reduksi dimensi untuk meningkatkan performa metode fuzzy klastering pada big data. *Science, Technology, Engineering, Economics, Education, and Mathematics*, *1*(1), 27–36.

Febriyanti, N. A., Goejantoro, R., & Prangga, S. (2023). Optimasi Fuzzy C-Means menggunakan particle swarm optimization Untuk pengelompokan kabupaten / kota di pulau Kalimantan (studi kasus : data indikator kesejahteraan rakyat tahun 2020). *Jurnal EKSPONENSIAL*, *14*(1), 31–40. http://jurnal.fmipa.unmul.ac.id/index.php/exponensial

Hartini, N. T. (2017). Pengaruh PDRB per kapita, investasi dan IPM terhadap ketimpangan pendapatan antar daerah di Provinsi DIY tahun 2011-2015. *Jurnal Pendidikan Dan Ekonomi*, *6*(6), 530–539.

Hepi, & Zakiah, W. (2018). Pengaruh angka harapan hidup dan rata-rata lama sekolah terhadap PDRB perkapita serta pertumbuhan ekonomi di provinsi kalimantan tengah 2011-2015. *Palangka Raya*, *4*(1), 56–68.

Ishikawa, K. (2021). The ASEAN Economic Community and ASEAN economic integration The ASEAN Economic Community and ASEAN economic. *Journal of Contemporary East Asia Studies*, *10*(1), 24–41. https://doi.org/10.1080/24761028.2021.1891702

Jolliffe, I. (2002). *Principal Component Analysis* (2nd ed.). Springer New York. https://doi.org/10.1007/b98835

Kalla, N. I., Annas, S., & Fahmuddin, M. (2022). Metode subtractive fuzzy c-means (SFCM) dalam pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan indikator kemiskinan. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(2), 95–108. https://doi.org/10.35580/variansiunm25

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 1942–1948. https://doi.org/10.1007/978-3-031-17922-8_4

Mayapada, R., Tinungki, G. M., & Sunusi, N. (2019). Penerapan sparse principal component analysis dalam menghasilkan matriks loading yang sparse. *Jurnal Matematika Statistika Dan Komputasi*, *15*(2), 44–54. https://doi.org/10.20956/jmsk.v15i2.5713

Nasution, D. P., Daulay, M. T., & Handani, E. (2020). Pengaruh pengeluaran pemerintah sektor kesehatan dan pendidikan terhadap kemiskinan di provinsi Lampung. *Jurnal Abdi Ilmu*, *10*(1), 33–49. https://doi.org/10.35448/jequ.v10i1.8594

Nasution, M. Z. (2019). Penerapan principal component analysis (PCA) dalam penentuan faktor dominan yang mempengaruhi prestasi belajar siswa (studi kasus : SMK Raksana 2 Medan). *Jurnal Teknologi Informasi*, *3*(1), 41–48. https://www.mdpi.com/2075-1680/14/5/351?

Rahmawati, T., Wilandari, Y., & Kartikasari, P. (2024). Analisis perbandingan silhouette coefficient dan metode elbow pada pengelompokkan provinsi di Indonesia berdasarkan indikator ipm dengan k-medoids. *Jurnal Gaussian*, *13*(1), 13–24. https://doi.org/10.14710/j.gauss.13.1.13-24

Rizqi, M., & Sidiq, S. (2023). Pengaruh penggunaan internet dan variabel makro terhadap PDB lima negara ASEAN. *Jurnal Kebijakan Ekonomi Dan Keuangan*, *2*(2), 115–125. https://doi.org/10.20885/jkek.vol2.iss2.art1

Rosyada, I. A., & Utari, D. T. (2024). Penerapan principal component analysis untuk reduksi variabel pada algoritma K-Means clustering. *Jambura Journal of Probability and Statistics*, *5*(1), 6–13. https://doi.org/10.34312/jjps.v5i1.18733

Sa'diyah, N., Supianto, A. A., & Dewi, C. (2020). Implementasi algoritme fuzzy c-means dengan particle swarm optimization (FCMPSO) untuk pengelompokan proses berpikir siswa dalam proses belajar. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *4*(6), 1625–1632. http://j-ptiik.ub.ac.id

Salim, A., & Fadilla. (2021). Pengaruh inflasi terhadap pertumbuhan ekonomi Indonesia. *Ekonomica Sharia: Jurnal Pemikiran Dan Pengembangan Ekonomi Syariah*, *7*(1), 17–28. www.bps.go.id,

Sanusi, W., Zaky, A., & Afni, B. N. (2020). Analisis fuzzy c-means dan penerapannya dalam pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan faktor-faktor penyebab gizi buruk. *Journal of Mathematics Computations and Statistics*, *2*(1). https://doi.org/10.35580/jmathcos.v2i1.12458

Shinta. (2024). Pengaruh urbanisasi terhadap perubahan kondisi sosial dan ekonomi di Indonesia. *Swarnabhumi*, *9*(1), 47–55.

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*(2), 1062–1069.

Widiyanto, M. T. A. C. (2019). Perrbandingan validitas fuzzy clustering pada fuzzy c-means dan particle swarms optimazation (PSO) pada pengelompokan kelas. *Jurnal Informatika Sunan Kalijaga*, *4*(1), 22–37. https://doi.org/10.14421/jiska.2019.41-03

Xia, P., Lin, Y., & Li-Hua, Z. (2019). An improved PSO-FCM algorithm for image segmentation. *IOP Conference Series Earth and Environmental Science*, *267*(4), 1–7. https://doi.org/10.1088/1755-1315/267/4/042081

Xie, X. L., & Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(8), 841–847. https://doi.org/10.1109/34.85677

Yu, G., Cui, Z., & Yuan, Q. (2023). Image segmentation method based on improved PSO optimized FCM algorithm and its application. *Journal of Computers*, *34*(2), 1–17. https://doi.org/10.53106/199115992023043402001