VARIANSI: Journal of Statistics and Its Application on Teaching and Research **ISSN 2684-7590 (Online)**
Vol. 8 No. 1 (2026), 8-20

DOI: 10.35580/variansiunm451

Clustering of Regencies/Cities in Jambi Province Based on The Human Development Index Indicators Using The K-Medoids Clustering

Dinda Aulia Rifani[1]* & Cut Multahadah[2]

Department of Mathematics, Jambi University, Indonesia

Keywords: Clustering, Human Development Index Indicator, K-Medoids, Silhouette Coefficient

Abstract:

The Human Development Index (HDI) is an index that used to measure the quality of human resources and societal welfare. The HDI value in Jambi Province keeps increasing. Nevertheless, there was a decline in the growth of HLS and RLS in 2024. In addition, the RLS of regencies/cities in Jambi Province just approaching 9 years. Therefore, further analysis of HDI indicators in each regency/city is needed to support the improvement of human development in regencies/cities in Jambi Province. This study aims to group the regencies/cities in Jambi Province based on HDI indicators. The grouping uses the k-medoids clustering algorithm along with silhouette coefficient method to determine the optimal number of clusters. The clustering results formed 2 clusters. Cluster 1 consists of 9 regencies/cities with HDI indicators lower than cluster 2, which consists of Jambi City and Sungai Penuh City. Therefore, the results indicate that cluster 2 has better human development quality compared to cluster 1.

1. Introduction

The Human Development Index (HDI) is an approach that can assess the success of a region's development. The measurement of HDI is determined through four main indicators, namely Life Expectancy of Birth (LEB), Expected Years of Schooling (EYS), Mean Years of Schooling (MYS), Adjusted Expenditure Per Capita (EPC) (BPS, 2025).

These indicators have a significant influence in improving the HDI value. High value of each HDI indicator indicates that the population can obtain better access in terms of income, health, and education. This supports sustainable economic growth (Fahrurrozi et al., 2023). Quoting from (BPS, 2024b), the HDI value in Jambi Province has continued to increase every year. It can be seen that in 2020, it was initially 72,29, and then reached 74,36 in 2025. This improvement is certainly supported by the increase in the HDI indicators. Although Jambi Province's HDI value continues to show improvement every year, there has been a decline in the growth of two HDI indicators. This decline in growth occurred in the MYS and EYS indicators. EYS growth decreased from 0,61 percent in 2023 to 0,08 percent in 2024. Similarly, the MYS indicator experienced a decline from 1,50 percent in 2023 to 1,02 percent in 2024. In addition, EYS growth in 10 out of 11 regencies/cities in Jambi is still below 1 percent. The MYS value of the regencies is also just approaching 9 years.

The HDI of Jambi Province is still below the national HDI value. The national HDI stands at 75,02, while the HDI of Jambi Province is only 74,36. This value is also far below the value of DKI Jakarta Province, which has the highest HDI of 84,15. This is also reflected in the HDI components of Jambi Province, which are still below national indicators value. Data from BPS shows that the Life Expectancy at Birth (LEB) in Jambi Province is still below the national LEB, at 72,04 years, compared to the national LEB of 74,15 years. Furthermore, the Expected Years of Schooling (EYS) in Jambi Province is 13,14 years, which is also lower than national EYS of 13,21 years. Similarly, the adjusted Expenditure Per Capita (EPC) in Jambi Province is only 11.621 thousand rupiahs, which is still lower than the national EPC of 12.341 thousand rupiahs (BPS, 2024a, 2024c). Improving the HDI value is very important for all regions in Indonesia, including Jambi Province. It is shown by Aprilianti & Harkeni (2021), who showed that the increasing HDI value of Jambi Province has positive impact on the welfare of the population in Jambi Province, whereby regional disparities can decrease along with the improvement of HDI.

Based on this, it can be seen that the growth of each HDI indicator is very important for increasing the HDI value in Jambi Province. HDI development is certainly supported by the 11 regencies/cities in Jambi Province. Therefore, further analysis of HDI indicators in each region is required to enhance human development growth in Jambi Province. In efforts to support human development growth in Jambi Province, it is necessary to group the regencies/cities based on the constituent indicators of HDI. This grouping aims to observe the extent of success in human development in Jambi Province. Cluster analysis is a method to group data based on the similarities of the characteristics of each object. This study performs grouping using the k-medoids cluster method. According to Han et al. (2012), the k-medoids method is a clustering method that groups data by addressing issues related to outlier data. The k-medoids method was chosen because it is considered suitable for the HDI indicator in Jambi Province, which contains outliers. Outliers are observations with characteristics that are very different and far from most of the data in a dataset. Outliers can cause problems that interfere with statistical tests (Hair Jr et al., 2019).

The k-medoids cluster has previously been used in grouping regions in NTT based on HDI indicators, resulting in 4 different clusters (Wicaksono & Yolanda, 2021). A comparison of hierarchical clustering, k-means dan k-medoids method conducted by Luthfi & Wijayanto (2021), showed k-medoids as the best clustering method because it produced the smallest Sw/Sb ratio. Research by Mustajab et al. (2021) on the application of k-medoids with the silhouette method grouped regencies/cities in West Java into 3 different clusters according to the level of HDI indicators.

Meanwhile, this study will focus on the grouping of regencies/cities in Jambi Province, with the aim of better understanding the condition of the HDI indicator and knowing the extent of success of human development in each regency/city. The results of this study are expected to be useful in describing the level of success of human development in each region, so that it can be used as evaluation material and reference for local governments in determining appropriate policies, and can also be used as a recommendation for determining the allocation of general allocation funds to meet development needs, in order to achieve more structured and equitable development according to the need of each region.

2. Literature Review

2.1 Human Development Index

Human Development Index (HDI) is an index that used as a development standard that measures the quality of human resources (Yogopriyatno & Aziman, 2024). The existence of the HDI emphasizes that the development of a region is not only evaluated based on economic growth, but also based on the quality of the people in a region, including health, education and standard of living (Visbal-Cadavid et al., 2025). The HDI comprehensively represents human quality of life through three main dimensions namely a long and healthy life, knowledge, and a decent standard of living. These three are represented by four indicators, Life Expectancy of Birth (LEB), Expected Years of Schooling (EYS), Mean Years of Schooling (MYS), Adjusted Expenditure Per Capita (EPC) (BPS, 2025).

The Life Expectancy of Birth (LEB) is a value that describes the average estimated number of years a person can live from birth, encompassing the number of births and deaths each year (Wicaksono & Yolanda, 2021). This indicator assesses the extent of the government's success in improving public welfare in terms of health (Astuti & Wijaya, 2024). Mean Years of Schooling (MYS) is a value that measures how long formal education has been pursued by individuals aged 25 and over. The Expected Years of Schooling (EYS) is a parameter that describes how long the education will be taken by children aged 7 in the future. These two indicators describe the development of education in a region. (BPS, 2025). Then, in the economic dimension, it is represented by Adjusted Expenditure Per Capita (EPC), which is the consumption cost for all household members in a full month divided by number of members in the family (Wicaksono & Yolanda, 2021). The HDI value can be used as a benchmark for the quality of life of the community and the level of success of government policies in efforts to improve human development (Arifin & Fadlan, 2021).

2.2 Data Standardization

Data standardization aims to ensure that the data being analyzed has the same range of values and that indicators with different units can be compared fairly (Ersawahyuni et al., 2025; Yogopriyatno & Aziman, 2024). The Z-Score equation for data standardization is as follows (Hair Jr et al., 2019):

$$Z_i = \frac{x_i - \bar{x}}{S} \quad (1)$$

where

Z_i : The i-th data standardized value

x_i : The i-th data value of variable x

$\bar{x}$: mean value of variable x

S : standard deviation of variable x

2.3 Outlier

Outliers are observations that have values that are very different and far from most of the data on one or a combination of several variables. Outlier detection can be carried out by using a boxplot, which represents the distribution of data for each variable (Hair Jr et al., 2019). The presence of outliers can cause abnormal impacts in statistical analysis (Astivia, 2024).

2.4 Cluster Analysis

Cluster analysis is a method that groups data into k clusters based on the similarity of data characteristics. There are two types of cluster analysis, namely hierarchical and non-hierarchical clusters. The hierarchical method is grouping using a dendrogram (Sharma, 1996). Types of hierarchical methods include *single-linkage method, complete-linkage method, average-linkage method, centroid method* dan *ward's method*. Meanwhile, the non-hierarchical method is a method of grouping data with the number of clusters must be determined before clustering (Hair Jr et al., 2019). Popular non-hierarchical methods are k-means and k-medoids. The measure of data similarity is obtained by involving distance measurement. One of distance measurement method is *euclidean distance*. Equation (2) is used to calculate the *euclidean distance* (Han et al., 2012).

$$d_{(i,k)} = \sqrt{\sum_{j=1}^{m} \left(x_{ij} - o_{kj}\right)^2} \quad (2)$$

where:
$d_{(i,k)}$: The distance between the i-th data point and cluster center k
x_{ij} : the i-th data of the j-th variable
o_{kj} : cluster center or medoid of cluster k in the j-th variable

2.5 Assumptions in Cluster Analysis

There are several assumptions in cluster analysis, that must be qualified before performing clustering. The assumptions are representativeness of sample and non-multicollinearity (Hair Jr et al., 2019).

a. Representativeness of the sample

Representative sample testing is conducted using the *Kaiser Meyer Olkin* (KMO) test. The *Kaiser Meyer Olkin* (KMO) test is used to examine whether the data adequately represents the population and are suitable for further analysis. The hypothesis in this test is as follows:

H_0 : Representative sample

H_1 : Non-representative sample

Critical area: Reject the null hypothesis (H_0) if $KMO \leq 0{,}5$

The KMO statistic is defined as:

$$KMO = \frac{\sum_j^m \sum_l^m {r_{jl}}^2}{\sum_j^m \sum_l^m {r_{jl}}^2 + \sum_j^m \sum_l^m {a_{jl}}^2} \tag{3}$$

where:

j : $1, 2, 3, \dots, m.$

l : $1, 2, 3, \dots, m.$

m : The number of variables

r_{jl} : Correlation coefficient between variables x_j and x_l

a_{jl} : Partial correlation coefficient between variables x_j and x_l

KMO value >0,5 means that the sample used is considered sufficient to represents the population and is suitable for cluster analysis (Amiruddin, 2023).

b. There is no multicollinearity among the variables

Multicollinearity detection aims to determine whether there is a strong linear relationship among independent variables or not (Ersawahyuni et al., 2025). The detection of multicollinearity can be determined through *Variance Inflation Factor* (VIF) (Hair Jr et al., 2019). The calculation of VIF and tolerance values use the following equation (4):

$$VIF_j = \frac{1}{TOL_j} = \frac{1}{1 - R_j^2} \tag{4}$$

where:

j : $1, 2, 3, \dots, m$

R_j^2 : Coefficient of determination of variable j when regressed on the remaining independent variable

TOL_j: Tolerance value of variable j

According to Hair Jr et al. (2019) VIF value >10 or $tolerance < 0{,}10$ indicates the presence of multicollinearity in that variable. One way to address multicollinearity is by reducing the variables that have multicollinearity symptoms (Gujarati & Porter, 2009).

2.6 Silhouette

The average silhouette value is used to assess the quality of clusters in grouping data. A silhouette value approaching 1 means that within the same cluster, objects are closely grouped, and objects are far from different clusters. Whereas a negative silhouette value indicates that object is closer to objects in another cluster than to objects in the same cluster. The quality of clusters in grouping can be observed using the average silhouette value (Han et al., 2012). the optimal number of k clusters is chosen from k clusters with the highest average silhouette (Rahmawati et al., 2024).

The evaluation criteria for assessing the quality of clusters based on the silhouette coefficient score are presented in Table 1 below (Kaufman & Rousseeuw, 1990).

Table 1. *Silhouette Coefficient Criteria*

SC	Interpretation
0,71-1,00	Strong Structure
0,51-0,70	Reasonable Structure
0,26-0,50	Weak Structure
≤0,25	No Substantial Structure

2.7 K-Medoids Cluster

K-medoids is a non-hierarchical clustering method that works by selecting a number of objects as medoids, then allocating other objects to the cluster that is closest to those medoids (Jollyta et al., 2020). This method is more resistant to outlier data. The clustering process begins by randomly selecting k objects as the objects that represent the cluster center points (Han et al., 2012). In the k-medoids method, poor initialization of the initial medoids does not have a significant impact on the quality of the clustering results, but it can make the clustering processing time longer (Schubert & Rousseeuw, 2021). This is because the cluster center is chosen from actual data points within the dataset. The following are the steps in performing k-medoids cluster (Jollyta et al., 2020):

1. Select medoids or cluster centers from the objects in the data, corresponding to the desired number of clusters k.
2. Calculate the distance between each object and medoids using euclidean distance.
3. Distribute objects to clusters based on the nearest distance according to the euclidean distance measurement results.
4. Select object randomly in each cluster to be determined as new medoids
5. Calculate the distance of each object in each cluster to the new medoid.
6. Calculate the total deviation (S) by subtracting the total new distance from the total old distance. If S < 0, replace the medoid with the new object as the new medoid for each cluster.
7. Repeat the process from steps 3 to 5 until there is no change in the medoids..

3. Methodology

The data used in this study are secondary data in the form of HDI indicators in regencies/cities in Jambi Province obtained from the official website of the Central Bureau of Statistics (BPS). The observation objects consist of 11 regencies/cities in Jambi Province in 2024. The research variables are presented in Table 2.

Table 2. Research Variables

No.	Variable	Information	Unit
1.	X_1	Life Expectancy of Birth (LEB)	Year
2.	X_2	Expected Years of Schooling (EYS)	Year
3.	X_3	Mean Years of Schooling (MYS)	Year
4.	X_4	Adjusted Expenditure Per Capita (EPC)	Thousand Rupiahs

The data is processed with the help of RStudio software. The data is analyzed using the k-medoids method with distance measurement using Euclidean distance and using the silhouette method to determine the optimal number of clusters. The following are the steps of the data analysis that are carried out:

1. Conduct descriptive statistical analysis to examine the characteristics of the HDI indicator variables.
2. Standardize data across all variables.
3. Detect outliers using boxplot.
4. Perform cluster assumption tests, including the Kaiser-Meyer-Olkin (KMO) test to assess representativeness of the sample and detect multicollinearity through Variance Inflation Factor (VIF) or tolerance values.
5. Determine the number of clusters (k) using the silhouette coefficient method.
6. Conduct k-medoids cluster analysis as follows:
   a) Select the initial medoids randomly by choosing k objects, which are considered as different clusters.
   b) Calculate the distance of each object to the selected initial medoids using the Euclidean distance measure.
   c) Determine the members of each cluster based on the closest distance to the medoid.
   d) Calculate the total distance of non-medoid objects to the medoids.
   e) Randomly select a non-medoid object from each cluster to become a new medoid, ensuring not to choose objects that have already been a medoid in the previous iteration.
   f) Calculate the distance of each non-medoid object to the new medoid.
   g) Determine the members of each cluster based on the closest distance to the new medoid.
   h) Calculate the total closest distance of non-medoid objects to the new medoid.
   i) Calculate the total deviation (S) obtained from the result of subtracting the total closest distance at the new medoid from the total closest distance at the initial medoid.
   j) If $S < 0$, then select the new medoid and repeat steps e) to i), but, if $S \geq 0$, the process is complete and can be stopped.
7. Evaluate the cluster results
8. Interpret the cluster results.

4. Results and Discussion

4.1. Descriptive Statistical Analysis

In this study, descriptive statistical calculations are carried out on each variable to determine the characteristics of each human development index indicator. The descriptive statistical analysis conducted includes the mean, minimum and maximum values, variance, and standard deviation. The results of the descriptive statistical analysis can be seen in Table 3. Life Expectancy of Birth (LEB), Expected Years of Schooling (EYS), Mean Years of Schooling (MYS), Adjusted Expenditure Per Capita (EPC)

Table 3. Descriptive Statistics

Variable	Minimum	Maximum	*Mean*	Variance	Standard Deviation
Life Expectancy of Birth (LEB) (X_1)	67,26	73,6	70,79454	3,874667273	1,968417
Expected Years of Schooling (EYS) (X_2)	12,21	15,55	13,27272	1,160201818	1,077127
Mean Years of Schooling (MYS) (X_3)	7,700	11,510	8,82272	1,283381818	1,132864
Adjusted Expenditure Per Capita (EPC) (X_4)	10.007	13.388	11.506,36364	1135630,455	1.065,659634

Based on Table 3 above, the characteristics of each variable are obtained. The mean of Life Expectancy of Birth (LEB) (X_1) in the regencies/cities of Jambi Province 70,79454 years, with a minimum value of 67,26 years in East Tanjung Jabung Regency and the maximum value hold by Jambi City at 73,66 years, There is a difference of 6,4 years between Life Expectancy of Birth of East Tanjung Jabung Regency and Jambi City. The variable LEB has variance of 3,874667273 and standard deviation of 1,968417. The mean of Expected Years of Schooling (EYS) (X_2) is 13,27272 with minimum value of 12,21 years in East Tanjung Jabung Regency and maximum value of 15,55 years in Jambi City, with difference of 3,34 years between them. The variable EYS has variance of 1,160201818 and standard deviation of 1,077127. Next, for the variable Mean Years of Schooling (MYS) (X_3), the regencies/cities in Jambi Province have mean value of 8,82272 years with minimum value in East Tanjung Jabung Regency of 7,7 years and the highest MYS hold by Jambi City of 11,51 years. This indicates that there are still regencies/cities in Jambi Province that have not yet reached 9 years of education. MYS has variance of 1,283381818 and standard deviation of 1,132864. Then for the adjusted expenditure per capita (EPC) (X_4), regencies/cities in Jambi Province have mean value of 11.506,36364 thousand rupiahs with minimum value of 10.007 thousand rupiahs in Muaro Jambi Regency and maximum value in Jambi City of 13.388 thousand rupiahs, with difference of 3.381 thousand rupiahs. The EPC variable has variance of 113.5630,455 and standard deviation of 1.065,659634 thousand rupiahs.

4.2. Data Standardization

The four variables have different units which can affect the clustering result, therefore the data must be standardized first. Data standardization is carried out using the z-score formula in equation (1). The results of the data standardization are shown in Table 4.

Table 4. Data Standardization

Regency/City	Z_1 (LEB)	Z_2 (EYS)	Z_3 (MYS)	Z_4 (EPC)
Kerinci	0,002771031	0,6009254	-0,1966054	-0,061336316
Merangin	0,586996698	-0,9402119	-0,3731490	-0,076350491
Sarolangun	-0,444288610	-0,6245573	-0,4172849	1,100385458
Batang Hari	0,383787770	-0,2810508	-0,4172849	-0,595277907
Muaro Jambi	0,658119823	0,0903076	-0,1083336	-1,406981731
East Tanjung Jabung	-1,795627979	-0,9866317	-0,9910517	-1,105759850
West Tanjung Jabung	-0,957391152	-0,5502856	-0,4437665	-0,760433830
Tebo	-0,053111424	-0,3181866	-0,6909276	-0,085734350
Bungo	-1,058995616	-0,5595695	-0,1966054	1,232697873
Jambi City	1,455714864	2,1142107	2,3721044	1,765701078
Sungai Penuh City	1,222024597	1,4550496	1,4629047	-0,006909933

Table 4 shows that all variables have been standardized so that they are within the same value range, and subsequently, the standardized data can be used for further analysis.

4.3. Outlier Detection

Outlier detection is carried out using *boxplot*. Figure 1 below presents boxplots for the four variables.

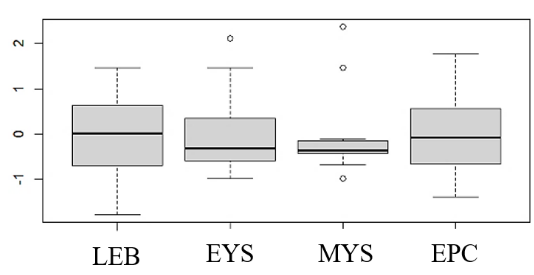

Figure 1. *Boxplot* Data *Outlier*

The boxplot results in Figure 1 show the presence of outlier data. The outliers are represented by points outside the boxplot box. The boxplot results indicate the existence of outliers in the Expected Years of Schooling (EYS) and

Mean Years of Schooling (MYS) variables. However, the presence of outliers is not a significant problem, because in this study, using the k-medoids method, which is more resistant to outliers.

4.4. The Cluster Assumption Tests

In cluster analysis, there are two assumptions, namely non-multicollinearity and representativeness of the sample. The following tests are conducted to examine whether the research data meet these assumptions or not.

a) Representativeness of the sample

The Kaiser Meyer Olkin (KMO) test is used to test the assumption of the representativeness of the sample. Based on the KMO test, a KMO value of 0,66. This value is greater than the minimum threshold of 0.5 so H_0 failed to be rejected. It means that the sample is representative, able to represent the existing population, and is eligible for cluster analysis. Amiruddin (2023) states that KMO value > 0,5 indicates that the sample is sufficient and has represented the population, thus the sample is feasible for cluster analysis

b) There is no multicollinearity

In cluster analysis, the variables should not exhibit multicollinearity. The examination is carried out by calculating the VIF and tolerance values for each variable. The VIF and tolerance values obtained are presented in Table 5.

Table 5. VIF and Tolerance

Variable	R_i^2	VIF	Tolerance
Z_1	0,6001	2,500384	0,3999
Z_2	0,8824	8,502353	0,1176
Z_3	0,9091	10,999894	0,0909
Z_4	0,4172	1,715943	0,5828

Based on Table 5 above, it can be seen that in variables Z_1 (LEB), Z_2 (EYS) dan Z_4 (EPC) there is no multicollinearity, indicated by VIF value < 10. However, in variable Z_3 (MYS) VIF value>10, which means that there is an indication of multicollinearity. If there is still multicollinearity in the variable, it is necessary to handle multicollinearity so that the non-multicollinearity assumption can be met. So that a reduction is carried out on the variable that experience multicollinearity, namely variable Z_3. The following Table 6 presents the VIF values of each variable after the RLS variable is removed.

Table 6. VIF and *Tolerance* Values After Variable Reduction

Variable	R_i^2	VIF	Tolerance
Z_1	0,5591	2,2680	0,4409
Z_2	0,6076	2,5484	0,3924
Z_4	0,1664	1,1996	0,8336

Table 6 shows that after the reduction of the MYS variable, there is no more multicollinearity. In addition, there is a decrease in the VIF value for each variable. After the reduction of MYS variable, the VIF value of LEB variable decreases from 2,5006 to 2,2680. The VIF value of the HLS variable also decreases drastically from 8,5034 to 2,5484. Similarly, the VIF of the EPC variable, which was originally 1,71585 becomes 1,1996. But, after the reduction of variable, there is a change in the KMO value from initially 0,66 to 0,53 and this value still qualify the representativeness of the sample test because still greater than the minimum threshold of 0.5.

The reduction of the MYS variable from the cluster analysis resulted in the resulting clusters not being able to provide a comprehensive picture of the education dimension. Based on the HDI calculation methodology according to BPS (2025), the education dimension in measuring HDI is measured through two education indicators: MYS and EYS. Therefore, the cluster results do not fully represent the HDI construction, but rather represent groupings that represent some of the HDI's constituent indicators, namely LEB, EYS, and EPC.

4.5. Determination of The Optimal Number of Clusters Using Silhouette Coefficient

Before performing k-medoids clustering, the optimal number of clusters is determined using the Silhouette Coefficient method.

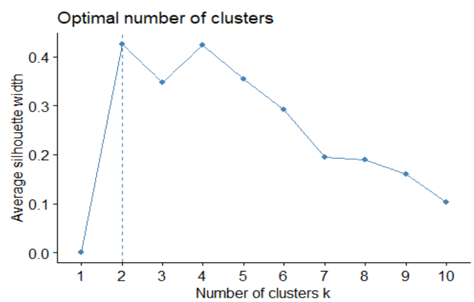

Figure 2. Average Silhouette Graph for Each k

According to Rahmawati et al. (2024) the number of clusters (k) with the highest average silhouette value is the most optimal number of clusters. Based on Figure 2 above, the optimal number of clusters is $k = 2$ which is the peak point in the graph above, this shows the highest average value of the silhouette coefficient among the other clusters.

4.6. K-Medoids Cluster

The clustering process is carried out using the k-medoids method, where the number of clusters is predetermined at two clusters. This process selects cluster centers, or medoids, from the objects in the dataset. The following are the initial medoids selected in the k-medoids clustering process.

Table 7. Initial *Medoid*

K	***Medoid***	Z_1	Z_2	Z_4
1	Tebo	-0,05311142	-0,3181866	-0,085734350
2	Sungai Penuh City	1,22202460	1,4550496	-0,006909933

Table 7 shows the objects selected as initial cluster centers or medoids. Tebo Regency is selected as medoid cluster 1, while Sungai Penuh City is selected as medoid cluster 2. After that, Calculate the distance of each object to the selected initial medoids using the Euclidean distance measure. Objects will be assigned to the cluster with the closest distance to the medoid.

Figure 3. Cluster Results Visualization

Based on Figure 3 above, it can be seen that cluster 1 is represented in red, while cluster 2 is represented in blue. In summary, the results of the k-medoids cluster can be seen in Table 8.

Table 8. The Results of Regency/City Clusters in Jambi Province Based on the Human Development Index (HDI) Indicator

Cluster	*Regencies/Cities*	Total Members
1	Kerinci, Merangin, Sarolangun, Batang Hari, Muaro Jambi, East Tanjung Jabung, West Tanjung Jabung, Tebo, Bungo	9
2	Jambi City, Sungai Penuh City	2

Table 8 shows that 9 regencies/cities are grouped in cluster 1, while the remaining 2 regencies/cities are in cluster 2. The areas in cluster 1 include Kerinci Regency, Merangin Regency, Sarolangun Regency, Batang Hari Regency, Muaro Jambi Regency, East Tanjung Jabung Regency, West Tanjung Jabung Regency, Tebo Regency, and Bungo Regency. Cluster 2 includes Jambi City and Sungai Penuh City.

4.7. The Evaluation of Cluster Results

The cluster results obtained are then evaluated to determine the quality of the resulting clusters. The silhouette coefficient method is used to evaluate the cluster results.

Table 9. ***Silhouette Coefficient Value***

The Number of Clusters k	*Silhouette Coefficient Value*
2	0,4261490
3	0.3471577
4	0.4247572
5	0.3542434
6	0.2562979

Table 9 shows the results of the cluster quality test with clusters $k = 2$ to $k = 6$. The evaluation results show that the cluster carried out with $k = 2$ has the best cluster quality compared to other k clusters, with the highest silhouette coefficient value of 0,4261490 with a weak cluster structure, because according to Kaufman & Rousseeuw (1990) the silhouette coefficient value in the range of $0{,}26 - 0{,}5$ indicates a weak structure.

4.8. The Interpretation of Cluster Results

The grouping of regencies/cities in Jambi Province produces two clusters. The characteristics of each cluster can be seen by calculating the average value of each HDI indicator in each cluster.

Table 10. The Average Value of Each Variable in Each Cluster

Variable	Standardization Data		Original Data	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2
LEB	-0,297526607	1,338869731	70,20888889	73,43
EYS	-0,396584489	1,78463015	12,84555556	15,195
EPC	-0,195421238	0,879395573	11.298,11111	12.443,5

Based on the average values in Table 10, it can be seen that the average value for each variable in cluster 1 is negative, which means that all the average values of the HDI indicators in cluster 1 are below the overall average value, while the variables in cluster 2 have an average value that is positive, which means that all the average values of the HDI indicators in the regions classified in cluster 2 are above the average value of the overall HDI indicators. Table 10 shows that cluster 1 contains regencies/cities in Jambi Province with Life Expectancy of Birth (LEB), Expected Years of Schooling (EYS), and Adjusted Expenditure Per Capita (EPC) indicators that are lower than the HDI indicators in regencies/cities in cluster 2.

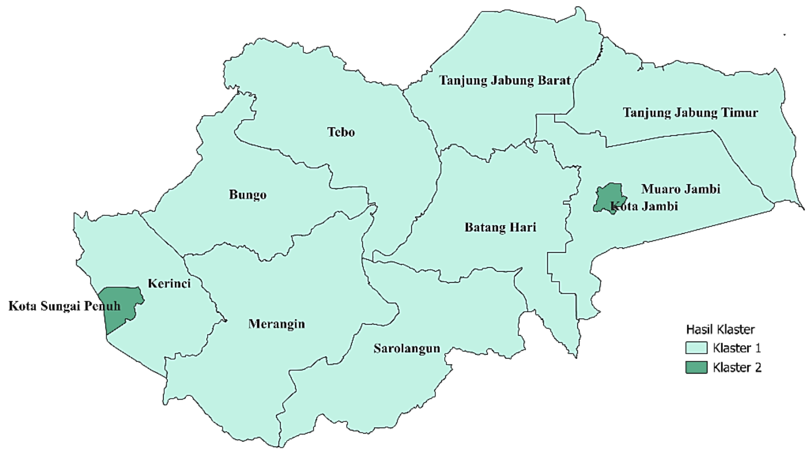

Figure 4. Map of Regency/City Clusters in Jambi Province

Figure 4 shows the cluster mapping in regencies/cities in Jambi Province, with the lighter colored parts representing regencies/cities in Jambi Province that are included in cluster 1, while the darker colored parts represent regencies/cities in Jambi Province that are included in cluster 2. The level of the HDI indicator in cluster 2 is better than cluster 1. This occurs because cluster 2 contains urban areas that usually have health facilities, education and economic centers that tend to be more advanced than regency areas in cluster 1. This result is in line with the concept of cumulative causation proposed by Myrdal (1957) regarding the tendency of more developed regions to develop more rapidly than less developed regions. This statement is supported by the backwash effect, which suggests that professional resources tend to move to more developed regions with more adequate facilities (Priatna et al., 2025). Research by Jagódka (2025) also shows that there is a greater withdrawal of human resources in more developed regions, which can lead to inequality between regions.

The EYS value in cluster 1 tends to be lower due to difficult and limited access to education and fewer school choices at the regency level. Easier access to technology also supports higher EYS achievement in cluster 2. The LEB value in cluster 2 is higher than the LEB in cluster 1 because health facilities in cities are more adequate and easily accessible than health facilities in cluster 1. Facilities in cluster 1 tend to be more limited. The low expenditure per capita in cluster 1 may occur because the income of residents and consumption levels in that area are lower than those of residents in cluster 2. Urban areas such as those in cluster 2 are usually the center of the economy, so that the expenditure in cluster 2 tends to be higher than in cluster 1.

5. Conclusion

Based on the results of the research that has been conducted, the results of the grouping of regencies/cities in Jambi Province based on the HDI indicator using k-medoids with silhouette coefficients produce optimal clusters of 2 clusters. Cluster 1 consists of 9 regencies/cities, while cluster 2 consists of 2 cities, namely Jambi City and Sungai Penuh City. Cluster 2 has average value of the HDI indicator that is higher than the average value of the HDI indicator in cluster 1, so it is concluded that cluster 2 contains areas with better HDI indicator conditions than cluster 1. In other words, the quality of human development in cluster 2 is better than cluster 1. Future research can add other variables related to human welfare and development in addition to the main HDI indicators, such as poverty rate, poverty level, infrastructure availability, and so on. Moreover, it is advisable to compare the k-medoids method with other methods, such as fuzzy c-mean, k-means, and other hierarchical methods to obtain the best results among several methods. Furthermore, distance measurements can be measured using other methods, such as the Manhattan method, to obtain the most optimal distance measurement results, and cluster quality can be measured using cluster validation, such as the Dunn Index and the Davies Bouldin Index.

References

Amiruddin. (2023). *Pertumbuhan Ekonomi, Mobilitas Sosial dan Perdagangan Melalui Transportasi Laut Implikasi Faktor Pengaruh Kinerja ASN Joint Inspection*. Deepublish. https://www.google.co.id/books/edition/Pertumbuhan_Ekonomi_Mobilitas_Sosial_dan/Z3Q-EQAAQBAJ?hl=id&gbpv=1

Aprilianti, V., & Harkeni, A. (2021). Pengaruh Indeks Pembangunan Manusia Terhadap Ketimpangan Wilayah di Provinsi Jambi. *Khazanah Intelektual*, *5*(2), 1142–1160. https://doi.org/https://doi.org/10.37250/newkiki.v4i1.111

Arifin, S. R., & Fadlan. (2021). Pengaruh Indeks Pembangunan Manusia (IPM) dan Tingkat Pengangguran Terhadap Pertumbuhan Ekonomi di Provinsi Jawa Timur Tahun 2016-2018. *IQTISHADIA Jurnal Ekonomi & Perbankan Syariah*, *8*(1), 38–59. https://doi.org/10.1905/iqtishadia.v8i1.4555

Astivia, O. L. O. (2024). A Method to Simulate Multivariate Outliers with Known Mahalanobis Distances for Normal and Non-Normal Data. *Methods in Psychology*, *11*, 1–7. https://doi.org/10.1016/j.metip.2024.100157

Astuti, E. D., & Wijaya, R. S. (2024). Pengaruh Indikator Indeks Pembangunan Manusia dan Jumlah Penduduk Terhadap Pertumbuhan Ekonomi di Kabupaten Sampang. *Jambura Economic Education Journal*, *6*(2), 397–418.

BPS. (2024a). *Indeks Pembangunan Manusia (IPM) Indonesia tahun 2024 mencapai 75,02, meningkat 0,63 poin atau 0,85% dibandingkan tahun sebelumnya yang sebesar 74,39*. Badan Pusat Statistik. https://www.bps.go.id/id/pressrelease/2024/11/15/2296/indeks-pembangunan-manusia--ipm--indonesia-tahun-2024-mencapai-75-02--meningkat-0-63-poin-atau-0-85-persen-dibandingkan-tahun-sebelumnya-yang-sebesar-74-39-.html

BPS. (2024b). *Indeks Pembangunan Manusia Provinsi Jambi 2024*. Badan Pusat Statistik Provinsi Jambi. https://jambi.bps.go.id/id/publication/2024/12/27/dee0caf718f83d7b98a4e993/indeks-pembangunan-manusia-provinsi-jambi-2024.html

BPS. (2024c). *IPM Provinsi Jambi 2024 Mencapai 74,36 Meningkat 0,63 poin (0,85 persen) Dibandingkan Tahun Sebelumnya (73,73)*. Badan Pusat Statistik Provinsi Jambi. https://jambi.bps.go.id/id/pressrelease/2024/12/02/746/ipm-provinsi-jambi-2024-mencapai-74-36-meningkat-0-63-poin%25E2%2580%25930-85-persen%25E2%2580%2593dibandingkan-tahun-sebelumnya%25E2%2580%259373-73-.html

BPS. (2025). *Indeks Pembangunan Manusia 2024*. Badan Pusat Statistik. https://www.bps.go.id/id/publication/2025/05/15/dd9a3ce7dfae1c733e46338f/indeks-pembangunan-manusia-2024.html

Ersawahyuni, A., Martha, S., & Perdana, H. (2025). Pengelompokan Indeks Pembangunan Manusia Di Indonesia Menggunakan Metode K-Medoids Dengan Evaluasi Davies Bouldin Index. *Buletin Ilmiah Math. Stat Dan Terapannya (Bimaster)*, *14*(2), 207–216.

Fahrurrozi, M., Mohzana, M., Haritani, H., Yunitasari, D., & Basri, H. (2023). Peningkatan Indeks Pembangunan Manusia Regional Dan Implikasinya Terhadap Ketahanan Ekonomi Wilayah (Studi Di Kabupaten Lombok Timur, Nusa Tenggara Barat). *Jurnal Ketahanan Nasional*, *29*(1), 70–89. https://doi.org/http://dx.doi.org/10.22146/jkn.83425

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. Douglas Reiner.

Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis*. Annabel Ainscow.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. Morgan Kaufmann.

Jagódka, M. (2025). Typification of Polish Regions Based on Human Capital and Innovativeness: A Cluster Analysis Approach. *Transforming Government: People, Process and Policy*, *19*(3), 614–637. https://doi.org/10.1108/TG-02-2025-0050

Jollyta, D., Ramdan, W., & Zarlis, M. (2020). *Konsep Data Mining dan Penerapan*. Deepublish.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Inc.

Luthfi, E., & Wijayanto, A. W. (2021). Analisis Perbandingan Metode Hirearchical, K-Means, dan K-Medoids Clustering Dalam Pengelompokkan Indeks Pembangunan Manusia Indonesia. *Inovasi*, *17*(4), 761–773. https://journal.feb.unmul.ac.id/index.php/INOVASI/article/view/10106/1513

Mustajab, R., Aristawidya, R., Puspita, L., & Widodo, E. (2021). Aplikasi Metode K-Medoid pada Pengelompokan Kabupaten/Kota di Jawa Barat Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2020. *Jurnal Statistika Dan Aplikasinya*, *5*(2), 221–229.

Priatna, I. A., Widada, R., The, H. Y., Fitriani, Atmaja, U., Amelia, S., Guswandi, Wibowo, T., Prayugo, A., Ahmad, D., Djumaty, B. L., & Dey, N. P. H. (2025). *Pembangunan Daerah*. Widina Media Utama.

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

Schubert, E., & Rousseeuw, P. J. (2021). Fast and Eager K-Medoids Clustering: O(k) Runtime Improvement of The PAM, CLARA, and CLARANS Algorithms. *Information Systems*, *101*, 101804. https://doi.org/10.1016/j.is.2021.101804

Sharma, S. (1996). *Applied Mutivariate Techniques*. John Wiley & Sons, Inc.

Visbal-Cadavid, D., Delahoz-Domínguez, E., & Mendoza-Mendoza, A. (2025). A Multiple Factor Analysis and Hierarchical Clustering of Global Logistics Governance and Development. *Decision Analytics Journal*, *15*(November 2024), 100579. https://doi.org/10.1016/j.dajour.2025.100579

Wicaksono, A. S., & Yolanda, A. M. (2021). Pengelompokkan Kabupaten / Kota di Provinsi Nusa Tenggara Timur Berdasarkan Indikator Indeks Pembangunan Manusia Menggunakan K-Medoids Clustering. *Statistika Terapan*, *1*(1), 79–90.

Yogopriyatno, J., & Aziman, M. F. (2024). *Statistika untuk Administrasi Publik*. PT. Adab Indonesia.