Cluster Analysis of Inclusive Economic Development Using K-Means Algorithm

Riska Yanu Farifah1, Dita Pramesti2

1,2Bachelor of Information System, Telkom University, Bandung, Indonesia

ABSTRACT

This study aims to cluster 38 Districts/Cities in East Java based on the 10 forming indicators of inclusive economic development and to determine the inclusive economic growth of Districts/Cities above or below the total average. 10 indicators used in this study are GRDP per capita, GRDP by business field, Labor force participation rate, Unemployment rate, Gini ratio, Expenditure per capita, the number of poverty, Life expectancy, expectation years of schooling, and mean years of schooling. There are 3 scenarios in this study, namely 2 clusters, 3 clusters, and 4 clusters. The method of clustering in this study is using the K-means algorithm. This study uses the silhouette coefficient to evaluate the best cluster of 3 scenarios. The best k-means algorithm in this study is using 2 clusters with a silhouette coefficient of 0.87. There are 29 Districts/Cities included in cluster 1 with inclusive economic development below the total average and 9 Districts/Cities included in cluster 2 with inclusive economic development above the total average. The members of cluster 1 are mostly district areas and located in coastal or border areas and the members of cluster 2 are mostly urban or industrial areas.

Corresponding Author:
Riska Yanu Farifah
Bachelor of Information System, Telkom University, Bandung, Indonesia
Email: riskayanu@gmail.com

This is an open access article under the CC BY-SA license.
has been used in information technology and information systems, for example, it has been applied to cloud computing (Sharma and Bala, 2020) and cellular network site management (Gbadoubissa et al., 2020). In health, a k-means algorithm has been used to group mutations of the coronavirus (Hozumi et al., 2021). In ecology, a k-means algorithm has been used to partition the open unemployment rate in the South Sulawesi Province (Akramunnisa and Fajriani, 2020). In the economy, a k-means algorithm has been used to partition the open unemployment rate in the South Sulawesi Province (Akramunnisa and Fajriani, 2020). In ecology, a k-means algorithm has been used to determine the spatial pattern and the relationship between controlling factors and toxic elements in the topsoil (Xu et al., 2021). In the economy, a k-means algorithm has been used to partition the open unemployment rate in the South Sulawesi Province (Akramunnisa and Fajriani, 2020).

This study uses the k-means algorithm to cluster inclusive economic development data in East Java. In contrast to the study that has been done by (Akramunnisa and Fajriani, 2020), this study uses 3 scenarios to get the best cluster and has an additional stage that is used to evaluate the cluster results. The method to evaluate the best cluster in this stage is using the silhouette coefficient (Naghizadeh and Metaxas, 2020).

Inclusive economic development aims to create equitable access and opportunities for all levels of society, improve welfare, and reduce the gap between regions. Inclusive economic development is divided into 3 Pillars, namely Pillar 1 concerning economic growth and development, Pillar 2 relates to income distribution and poverty reduction, and Pillar 3 is an opportunity and expanding access. These 3 pillars are used to measure and monitor the level of inclusiveness of development in Indonesia, both on a national and regional scale (Agency, 2021). The higher the percentage of achievement of these 3 pillars, the more inclusive the development will be, thus the more prosperous the population (Statistics-East Java, 2020), (Hapsari, 2019), (Setianingtias et al., 2019). The inclusive economic development index is one of the benchmarks for the success of a region in the welfare of its population. It is known that in 2019, East Java was generally above the national inclusive economic development index. Pillar 1 index of inclusive economic development is 5.72 and national is 5.48, Pillar 2 has an index of 6.56 and national of 6.57, Pillar 3 has index of 7.28 and national is 6.09 (Agency, 2021). However, not all Districts/Cities in East Java have a level of development above the national inclusive development.

In this study, the k-means algorithm is used to cluster inclusive economic development data in East Java. In contrast to the study that has been done by (Akramunnisa and Fajriani, 2020), this study uses 3 scenarios to get the best cluster and has an additional stage that is used to evaluate the cluster results. The method to evaluate the best cluster in this stage is using the silhouette coefficient (Naghizadeh and Metaxas, 2020).

B. LITERATURE REVIEW

K-means algorithm is one of the most frequently used partition-based algorithms for clustering (Jain, 2010). K-Means algorithm studies each object in the data and forms partitions called clusters, representation the members in each cluster having similar characteristics. If the data used is continuous, each cluster is represented by a centroid which is the mean of the cluster members. In categorical data, each cluster is represented by a medoid, which is the object that occurs most frequently. K-Means uses squared Euclidean distances as a measure of similarity for cluster membership (Patel and Kushwaha, 2020). The formula of squared Euclidean distances is:

\[ d(x_i, x_k) = \sum_{j=1}^{p} (x_{ij} - x_{kj})^2 \] (1)

where \(x_{ij}\) is the \(i\)th object in the \(j\)th variable, \(i = 1, 2, \cdots, n, j = 1, 2, \cdots, p\) (the data dimension is \(n \times p\)), and \(x_{kj}\) is the \(k\)th centroid in the \(k\)th cluster. The formula for \(c_{kj}\) is (Kakushadze and Yu, 2017):

\[ c_{kj} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij} \] (2)

The goal of the k-means algorithm is to minimize the sum of the square errors (SSE) of all clusters formed (Jain, 2010), (Gbadoubissa et al., 2020). SSE can be written as follows:

\[ SSE = \sum_{i=1}^{n} \sum_{k=1}^{r} w_{i,k} \| x_i - c_k \|^2 \] (3)

where \(w_{i,k}\) has 2 values, 1 if \(x_i\) is in the \(k\)th centroid and 0 if \(x_i\) is not in the \(k\)th centroid (Patel and Kushwaha, 2020). The k-means algorithm has the following stages (Gbadoubissa et al., 2020): Input : \(X_{n \times p}\) and \(r, r \geq 2\) Output : \(r\) clusters \((c_1, c_2, \cdots, c_r)\) with members of each cluster

Vol. 5, No. 2, April 2022, Hal. 171–178
DOI: https://doi.org/10.30812/varian.v5i2.1894
1. Select \( r \) cluster centroids randomly \((c_1, c_2, \ldots, c_r)\)

2. Repeat

3. For each \( x_i \in X_{n \times p} \) do

4. For all centroids \( c_k, 1 \leq c_k \leq r \) do

5. If \( \| x - c_k \| < \| x - c_l \|, k \neq l, 1 \leq k, l \leq k \) then

6. Assign \( x \) in \( c_k \)

7. End if

8. End for all

9. End each

10. Calculate the latest centroids for each cluster

11. Until the objective function (SSE) is minimized

Based on the pseudocode above, the iteration of cluster formation will stop when the objective function (SSE) is minimum so that the cluster formation process is followed by drinking SSE (Jain, 2010). Evaluation of cluster formation from the k-means algorithm can use a silhouette coefficient. The silhouette coefficient is used in the optimization process of cluster formation to get the best number and members of the cluster (Naghizadeh and Metaxas, 2020). The algorithm of the silhouette coefficient has the following stages:

1. Calculating the mean distance between objects in the same cluster

\[
a_k = \frac{1}{n_j - 1} \sum_{i=1 \atop i \neq j}^{n_j} (x_i - x_{ij})^2
\]  

(4)

2. Calculating the mean distance between objects in different clusters and find the minimum

\[
b_k = \min \frac{1}{n_j} \sum_{i=1 \atop x_i \in c_k}^{n_j} (x_i - x_{ij})^2
\]  

(5)

3. Calculating the silhouette coefficient

\[
s = \begin{cases} 
\frac{b_k - a_k}{\max(a_k, b_k)} & c_k > 1; \\
0 & c_k = 1 
\end{cases}
\]  

(6)

4. Determining the mean of the silhouette coefficients obtained.

C. RESEARCH METHOD

This study will group Districts/Cities in East Java based on the inclusive economic development index. This study uses 2 sub-pillars for each pillar with details, 4 indicators on pillar 1, 3 indicators on pillar 2, and 3 indicators on pillar 3. The data used is secondary data obtained from BPS East Java. The details of the data used are as follows:

| Pilar                          | Sub Pilar           | Variable | Indicator                          |
|-------------------------------|---------------------|----------|------------------------------------|
| Economic Growth and Development | Economic Growth     | \( X_1 \) | GRDP per capita                     |
|                               |                     | \( X_2 \) | GRDP by business field             |
|                               | Employment Opportunity | \( X_3 \) | Labor force participation rate     |
|                               |                     | \( X_4 \) | Unemployment rate                  |
| Income Equity and Poverty Reduction | Inequality     | \( X_5 \) | Gini ratio                         |
|                               |                     | \( X_6 \) | Expenditure per capita              |
|                               | Poverty             | \( X_7 \) | The number of poverty               |
| Expansion of Access and Opportunities | Health         | \( X_8 \) | Life expectancy                     |
|                               |                     | \( X_9 \) | Mean years of schooling             |
|                               | Education           | \( X_{10} \) | Expectation years of schooling     |

This study uses the k-means algorithm to form clusters of Districts/Cities based on the indicators in Table 1. The stages in this analysis are as follows:
1. Specifies the number of clusters. This study uses 3 scenarios, namely using 2, 3, and 4 clusters
2. Choosing an initial centroid randomly
3. Calculating Euclidean distance.
4. Defining new group members.
5. Calculating new centroid
6. Repeating process c to e until there is no change in the members of each cluster and the minimum SSE is obtained
7. Generating cluster members in each scenario
8. Evaluating the cluster formed based on the silhouette coefficient
9. The best scenario is chosen based on point h.

D. RESULTS AND DISCUSSION

1. The Results K-means Algorithm

This study uses 3 scenarios to get the optimal and best cluster based on the objective function criteria and the silhouette coefficient. The first scenario uses 2 clusters, the second uses 3 clusters, and the third uses 4 clusters. Before analyzing, the 10 indicators of inclusive economic development are standardized first because those indicators have different units. The first stage is to select the centroids randomly. A selection of the centroids can be seen in Table 2 below.

| Scenario | District/City | Variable | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 |
|----------|---------------|----------|-----|----|----|----|----|----|----|----|----|-----|
| 1 (2 clusters) | Pacitan | 26999.6 | 14968.4 | 79.55 | 0.95 | 71.52 | 12.61 | 7.19 | 75.86 | 0.35 | 8527 |
| | Surabaya | 186738.9 | 538845.5 | 68.61 | 5.87 | 73.98 | 14.78 | 10.46 | 130.55 | 0.36 | 17157 |
| 2 (3 clusters) | Pacitan | 26999.6 | 14968.4 | 79.55 | 0.95 | 71.52 | 12.61 | 7.19 | 75.86 | 0.35 | 8527 |
| | Surabaya | 186738.9 | 538845.5 | 68.61 | 5.87 | 73.98 | 14.78 | 10.46 | 130.55 | 0.36 | 17157 |
| | Bangkalan | 24360.9 | 23846.7 | 63.11 | 5.84 | 69.94 | 11.58 | 5.33 | 186.11 | 0.29 | 8393 |
| 3 (4 clusters) | Surabaya | 186738.9 | 538845.5 | 68.61 | 5.87 | 73.98 | 14.78 | 10.46 | 130.55 | 0.36 | 17157 |
| | Sampang | 19726.4 | 19105.4 | 66.42 | 2.81 | 67.79 | 11.76 | 4.36 | 202.21 | 0.26 | 8569 |
| | Kediri City | 447215.8 | 127716.8 | 64.6 | 4.22 | 73.8 | 14.96 | 9.91 | 20.54 | 0.33 | 11976 |
| | Batu city | 76004.9 | 15640.9 | 71.01 | 2.48 | 72.37 | 14.04 | 8.77 | 7.89 | 0.33 | 12466 |

Based on Table 2, it is known that Surabaya is selected in all scenarios and Pacitan is selected in scenario 1 and scenario 2. The next stage is creating clusters and determining the members of each cluster based on Euclidean distance. The results of the k-means algorithm of 10 indicators of inclusive economic development are:

| Table 3. Membership of Each Cluster |
|-------------------------------------|
| Scenario | Cluster | Amount | Maximum Iteration |
|----------|---------|--------|-------------------|
| 1 (2 clusters) | Cluster 1 | 29 | 6 |
| | Cluster 2 | 9 | |
| 2 (3 clusters) | Cluster 1 | 16 | |
| | Cluster 2 | 5 | 4 |
| | Cluster 3 | 17 | |
| 3 (4 clusters) | Cluster 1 | 1 | |
| | Cluster 2 | 15 | 5 |
| | Cluster 3 | 1 | |
| | Cluster 4 | 21 | |

The maximum iteration is based on the value of the objective function in the iteration. If SSE is used as an objective function and produces the smallest value, the cluster formation process is stopped (Gbadoubissa et al., 2020). Based on Table 3, scenario 2 reaches the optimum in the fourth iteration, and this is the least number of iterations of all the scenarios. Meanwhile, the first scenario has the most iterations, namely optimum in the sixth iteration. Based on the results of the iteration in Table 3, scenario 2 is the best.

The results of k-means clustering of inclusive economic development for all of the scenarios can be seen in the Figure 1.
Cluster centroids are shown in Table 4. GRDP per capita of cluster 1 and cluster 3 have the same criteria (both values are negative). The details of the final result for the scenario 2 and scenario 3. In scenario 2 and scenario 3, the last centroids have the same criteria. For example in scenario 2, example, GRDP per capita in cluster 1 has a negative value, but in cluster 2 has a positive value. This result is not the same in scenario 1 don’t have the same criteria. In other words, each indicator in cluster 1 and cluster 2 have a different value. For members of a cluster are above the total average. From Figure 1, it is known that the last centroids between cluster 1 and cluster 2 are both above the total average. To determine the type or name of each group in each scenario, it can calculate the average of the resulting centroids in each indicator. Determination of the best scenario of k-means clustering in the grouping of Districts/Cities in East Java Province based on 10 indicators of inclusive economic development, the silhouette value that is in one range, below the total average or above the total average. For example, the GRDP per capita indicator in cluster 1 and cluster 2 are both above the total average. From Table 4 and Figure 1, in the first scenario, the centroid values of the 10 indicators creating clusters in scenario 1 with 1 (2 clusters) Cluster 1 $X_1 \downarrow, X_2 \downarrow, X_3 \uparrow, X_4 \downarrow, X_5 \downarrow, X_6 \downarrow, X_7 \downarrow, X_8 \uparrow, X_9 \downarrow, X_{10} \downarrow$ Cluster 2 $X_1 \uparrow, X_2 \uparrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \downarrow, X_9 \uparrow, X_{10} \uparrow$

2 (3 clusters) Cluster 1 $X_1 \downarrow, X_2 \downarrow, X_3 \uparrow, X_4 \downarrow, X_5 \uparrow, X_6 \downarrow, X_7 \uparrow, X_8 \downarrow, X_9 \uparrow, X_{10} \downarrow$ Cluster 2 $X_1 \uparrow, X_2 \uparrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \downarrow, X_9 \uparrow, X_{10} \uparrow$ Cluster 3 $X_1 \uparrow, X_2 \downarrow, X_3 \downarrow, X_4 \uparrow, X_5 \downarrow, X_6 \downarrow, X_7 \downarrow, X_8 \downarrow, X_9 \downarrow, X_{10} \downarrow$

3 (4 clusters) Cluster 1 $X_1 \uparrow, X_2 \uparrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \uparrow, X_9 \uparrow, X_{10} \uparrow$ Cluster 2 $X_1 \downarrow, X_2 \downarrow, X_3 \uparrow, X_4 \downarrow, X_5 \downarrow, X_6 \downarrow, X_7 \downarrow, X_8 \downarrow, X_9 \downarrow, X_{10} \downarrow$ Cluster 3 $X_1 \uparrow, X_2 \uparrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \downarrow, X_9, X_{10} \uparrow$ Cluster 4 $X_1 \downarrow, X_2 \downarrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \downarrow, X_9 \uparrow, X_{10} \uparrow$

From Table 4 and Figure 1, in the first scenario, the centroid values of the 10 indicators creating clusters in scenario 1 with scenario 2 are in different ranges. In the second scenario, there are several components whose centroid values are in the same range or intersect. For example, the centroid values for the Life Expectancy indicator in cluster 1 and cluster 2 are both negative. However, the centroid values are not the same. Likewise, with scenario 3, there are several indicators in each cluster that have a centroid value that is in one range, below the total average or above the total average. For example, the GRDP per capita indicator in cluster 1 and cluster 2 are both above the total average. To determine the type or name of each group in each scenario, it can calculate the average of the resulting centroids in each indicator. Determination of the best scenario of k-means clustering in the grouping of Districts/Cities in East Java Province based on 10 indicators of inclusive economic development, the silhouette

Based on Figure 1, information about the last centroid of the 10 indicators of inclusive economic development can be obtained. If a centroid is negative, then the cluster members are below the total average and if a centroid is positive, then the members of a cluster are above the total average. From Figure 1, it is known that the last centroids between cluster 1 and cluster 2 of scenario 1 don’t have the same criteria. In other words, each indicator in cluster 1 and cluster 2 have a different value. For example, GRDP per capita in cluster 1 has a negative value, but in cluster 2 has a positive value. This result is not the same in scenario 2 and scenario 3. In scenario 2 and scenario 3, the last centroids have the same criteria. For example in scenario 2, GRDP per capita of cluster 1 and cluster 3 have the same criteria (both values are negative). The details of the final result for the cluster centroids are shown in Table 4.

Table 4. Results of the Cluster Centroids

| Scenario | Cluster | The Value of Indicators |
|----------|---------|-------------------------|
| 1 (2 clusters) | Cluster 1 | $X_1 \downarrow, X_2 \downarrow, X_3 \uparrow, X_4 \downarrow, X_5 \downarrow, X_6 \downarrow, X_7 \downarrow, X_8 \uparrow, X_9 \downarrow, X_{10} \downarrow$ |
| 1 (2 clusters) | Cluster 2 | $X_1 \uparrow, X_2 \uparrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \downarrow, X_9 \uparrow, X_{10} \uparrow$ |
| 2 (3 clusters) | Cluster 1 | $X_1 \downarrow, X_2 \downarrow, X_3 \uparrow, X_4 \downarrow, X_5 \uparrow, X_6 \downarrow, X_7 \uparrow, X_8 \downarrow, X_9 \uparrow, X_{10} \downarrow$ |
| 2 (3 clusters) | Cluster 2 | $X_1 \uparrow, X_2 \uparrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \downarrow, X_9 \uparrow, X_{10} \uparrow$ |
| 2 (3 clusters) | Cluster 3 | $X_1 \uparrow, X_2 \downarrow, X_3 \downarrow, X_4 \uparrow, X_5 \downarrow, X_6 \downarrow, X_7 \downarrow, X_8 \downarrow, X_9 \downarrow, X_{10} \downarrow$ |
| 3 (4 clusters) | Cluster 1 | $X_1 \uparrow, X_2 \uparrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \uparrow, X_9 \uparrow, X_{10} \uparrow$ |
| 3 (4 clusters) | Cluster 2 | $X_1 \downarrow, X_2 \downarrow, X_3 \uparrow, X_4 \downarrow, X_5 \downarrow, X_6 \downarrow, X_7 \downarrow, X_8 \downarrow, X_9 \downarrow, X_{10} \downarrow$ |
| 3 (4 clusters) | Cluster 3 | $X_1 \uparrow, X_2 \uparrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \downarrow, X_9, X_{10} \uparrow$ |
| 3 (4 clusters) | Cluster 4 | $X_1 \downarrow, X_2 \downarrow, X_3 \downarrow, X_4 \uparrow, X_5 \uparrow, X_6 \uparrow, X_7 \uparrow, X_8 \downarrow, X_9 \uparrow, X_{10} \uparrow$ |

From Table 4 and Figure 1, in the first scenario, the centroid values of the 10 indicators creating clusters in scenario 1 with scenario 2 are in different ranges. In the second scenario, there are several components whose centroid values are in the same range or intersect. For example, the centroid values for the Life Expectancy indicator in cluster 1 and cluster 2 are both negative. However, the centroid values are not the same. Likewise, with scenario 3, there are several indicators in each cluster that have a centroid value that is in one range, below the total average or above the total average. For example, the GRDP per capita indicator in cluster 1 and cluster 2 are both above the total average. To determine the type or name of each group in each scenario, it can calculate the average of the resulting centroids in each indicator. Determination of the best scenario of k-means clustering in the grouping of Districts/Cities in East Java Province based on 10 indicators of inclusive economic development, the silhouette
The silhouette coefficient can be used.

2. Cluster evaluation with silhouette coefficient

The silhouette coefficient is one of the measurement criteria to determine the best number of clusters used in clustering (Corporal-Lodango et al., 2014), (Naghizadeh and Metaxas, 2020). The results of the k-means clustering evaluation using the silhouette coefficient in the 3 scenarios used are:

Table 5. Silhouette Coefficients of Each Scenario

| Scenarios | Silhouette Coefficient |
|-----------|------------------------|
| 1 (2 clusters) | 0.87 |
| 2 (3 clusters) | 0.72 |
| 3 (4 clusters) | 0.79 |

The silhouette coefficient has a value from -1 to 1. There are 3 levels in the silhouette coefficient, namely \( s < 0.2 \) is a poor category, \( 0.2 \leq s \leq 0.5 \) is a fair category, and \( s > 0.5 \) is a good category (Naghizadeh and Metaxas, 2020), (Mooi et al., 2011). Table 5 shows that the silhouette coefficients generated from the three scenarios used in the k-means clustering analysis are more than 0.5, meaning that all scenarios produce a good number of clusters. Of the three scenarios, scenario 2 is the best k-means clustering analysis on the grouping of Districts/Cities in East Java based on an inclusive economic development index. The silhouette coefficient is 0.87.

Based on the results listed in Table 5, the k-means clustering analysis on the grouping of the inclusive economic development index in East Java is using 2 clusters. The results are as follows:

Table 6. Membership of 2 Clusters in the First Scenario

| Cluster | Amount | Districts/Cities |
|---------|--------|------------------|
| 1       | 29     | Pacitan, Ponorogo, Trenggalek, Tulunganggung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Bangkalan, Sampang, Pamekasan, Sumenep, Probolinggo City, Batu City |
| 2       | 9      | Sidoarjo, Gresik, Kediri City, Blitar City, Malang City, Pasuruan City, Mojokerto City, Madiun City, Surabaya City |

Table 6 shows that 76% of Districts/Cities in East Java are in cluster 1. Based on Figure 1 and Table 4, cluster 1 is a group with most of the indicators that have a negative value centroid. Indicators in cluster 1 are GRDP per capita, GRDP by business field, unemployment rate, Gini ratio, expenditure per capita, the number of poverty, Expectation years of schooling, and Mean years of schooling. The low unemployment rate and the number of poverty indicate that unemployment and poverty in this cluster are better than in cluster 2. However, 6 of the other indicators that are lower than the total average suggest that inclusive economic growth is lower than cluster 2. Most of the Districts/Cities in cluster 1 are coastal and border areas in East Java, where most of the population work as farmers, fishermen, or laborers (Agency, 2021), (Statistics-East Java, 2020). Determination of the cluster’s name or type can see in the last centroids that result (Corporal-Lodango et al., 2014), (Clayman et al., 2020). This study shows that the overall centroids in cluster 1 are lower than in cluster 2. Cluster 1, the average of all the indicators is lower than the total average of inclusive economic development indexes in East Java. Districts/Cities in cluster 1 are Pacitan, Ponorogo, Trenggalek, Tulunganggung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Bangkalan, Sampang, Pamekasan, Sumenep, Probolinggo City, Batu City.

Based on the description in the paragraph above, cluster 1 in scenario 1 is a group where the average inclusive economic development index of the 10 categories used is below the total average. Then it can be said that cluster 2 is above the total average. Cluster 2 consists of big cities or industrial areas, namely Sidoarjo, Gresik, Kediri City, Blitar City, Malang City, Pasuruan City, Mojokerto City, Madiun City, Surabaya City.

E. CONCLUSION AND SUGGESTION

This study represents a situation of clustering Districts/Cities in East Java Province based on the economic inclusive development data. 76% of Districts/Cities are in cluster 1 and 24% are in cluster 2. Cluster 1 is a group of Districts/Cities with an inclusive economic development index below the total average of inclusive economic development index in Jawa Timur and cluster 2 is a
group of Districts/Cities with an inclusive economic development index above the total average. In other words, cluster 2 shows a higher inclusive economic growth than cluster 1.

Based on the descriptions above, k-means results can describe the spread of inclusive economic development in East Java. The results of this clustering can be used as a reference for local and provincial governments as a basis for policy/decision making in improving the quality and quantity of indicators for inclusive economic development, particularly in Districts/Cities with indicators of inclusive economic development below the total average.

REFERENCES

Agency, N. D. P. (2021). *Indeks Pembangunan Ekonomi Inklusif*. http://inklusif.bappenas.go.id/indeks.

Akramunnisa, A. and Fajriani, F. (2020). K-Means Clustering Analysis pada Persebaran Tingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan. *Jurnal Varian*, 3(2):103–112.

Clayman, C. L., Srinivasan, S. M., and Sangwan, R. S. (2020). K-means Clustering and Principal Components Analysis of Microarray Data of L1000 Landmark Genes. *Procedia Computer Science*, 168:97–104.

Corporal-Lodangco, I. L., Richman, M. B., Leslie, L. M., and Lamb, P. J. (2014). Cluster analysis of north atlantic tropical cyclones. *Procedia Computer Science*, 36:293–300.

Gbadoubissa, J. E. Z., Ari, A. A. A., and Gueroui, A. M. (2020). Efficient K-Means Based Clustering Scheme for Mobile Networks Cell Sites Management. *Journal of King Saud University-Computer and Information Sciences*, 32(9):1063–1070.

Hapsari, W. R. (2019). Analisis Pertumbuhan Ekonomi Inklusif Kabupaten/Kota di Provinsi Jawa Tengah. *Jurnal Litbang Sukowati: Media Penelitian dan Pengembangan*, 3(1):11–11.

Hozumi, Y., Wang, R., Yin, C., and Wei, G.-W. (2021). UMAP-Assisted K-Means Clustering of Large-Scale SARS-CoV-2 Mutation Datasets. *Computers in biology and medicine*, 131:104264.

Jain, A. K. (2010). Data Clustering: 50 Years Beyond K-Means. *Pattern recognition letters*, 31(8):651–666.

Kakushadze, Z. and Yu, W. (2017). *K-Means and Cluster Models for Cancer Signatures*. *Biomolecular detection and quantification*, 13:7–31.

Mooi, E., and Sarstedt, M. (2011). *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics (1st ed.)*. Springer-Verlag Berlin Heidelberg.

Naghizadeh, A. and Metaxas, D. N. (2020). Condensed Silhouette: An Optimized Filtering Process for Cluster Selection in K-Means. *Procedia Computer Science*, 176:205–214.

Patel, E. and Kushwaha, D. S. (2020). Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. *Procedia Computer Science*, 171:158–167.

Qureshi, M. N. and Ahamad, M. V. (2018). An Improved Method for Image Segmentation using K-Means Clustering with Neutrosophic Logic. *Procedia computer science*, 132:534–540.

Setianingtias, R., Baiquni, M., Kurniawan, A., et al. (2019). Pemodelan Indikator Tujuan Pembangunan Berkelanjutan di Indonesia. *Jurnal Ekonomi Dan Pembangunan*, 27(2):61–74.

Sharma, V. and Bala, M. (2020). An Improved Task Allocation Strategy in Cloud using Modified K-Means Clustering Technique. *Egyptian Informatics Journal*, 21(4):201–208.

Statistica-East Java (2020). *Perkembangan beberapa Indikator Utama Sosial Ekonomi Provinsi Jawa Timur 2020*.

Webster, M. (2021). *Cluster Analysis*. https://www.merriam-webster.com/dictionary/cluster analysis.

Xie, H., Zhang, L., Lim, C. P., Yu, Y., Liu, C., Liu, H., and Walters, J. (2019). Improving K-means Clustering with Enhanced Firefly Algorithms. *Applied Soft Computing*, 84:105763.
Xu, H., Croot, P., and Zhang, C. (2021). Discovering Hidden Spatial Patterns and Their Associations with Controlling Factors for Potentially Toxic Elements in Topsoil using Hot Spot Analysis and K-Means Clustering Analysis. *Environment International, 151*:106456.