turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Tb Ai Munandar
Assignment title:	Quick Submit
Submission title:	Plagiarism Check TEM V12N2
File name:	TEMJournalMay2023_680_690.pdf
File size:	505.63K
Page count:	11
Word count:	6,508
Character count:	34,986
Submission date:	20-Mar-2024 10:36AM (UTC-0400)
Submission ID:	2325791602

TEM Journal. Volume 12, Issue 2, pages 680-690, ISSN 2217-8309, DOI: 10.18421/TEM122-11, May 2023.

Extracting Patterns of Proximity in Regional Development Inequality Using Hierarchical Agglomerative Clustering

Tb Ai Munandar [1], Dwipa Handayani [1]

[1]Faculty of Computer Science, Informatics Dept. Universitas Bhayangkara Jakarta Raya, Indonesia

Abstract **– Equitable regional development is still difficult to implement. Many interests in determining regional development priorities make it difficult to realize the main goals of national development. Political interests and the underutilization of regional statistical data are often other obstacles in carrying out regional development. This study aims to classify regional development data to be used to analyse existing development inequalities. Gross regional domestic income (GRDP) data is the main data in this study. The researcher used the hierarchical agglomerative clustering (HAC) technique. The results of the cluster can be used as a consideration for local governments to determine future regional development priorities. In addition, visualized cluster results in the form of a dendrogram can show the proximity of development inequality between regions.**

Keywords **– Regional development, development inequality, cluster technique, agglomerative clustering.**

1. Introduction

One of the phenomena arising when carrying out development in various parts of the world is the inequality of development between regions.

DOI: 10.18421/TEM122-11
https://doi.org/10.18421/TEM122-11

Corresponding author: Tb Ai Munandar,
Faculty of Computer Science, Informatics Dept. Universitas Bhayangkara Jakarta Raya, Indonesia
Email: tb.aimunandar@dsn.ubharajaya.ac.id

Received: 06 February 2023.
Revised: 24 April 2023.
Accepted: 10 May 2023.
Published: 29 May 2023.

Inequality here means that there have been differences in the level of achievement of development, welfare, and economic growth in each region [1]. Various strategies have been used to minimize development inequality [2]. Development inequality is influenced by several factors: statistical data on development achievements have not been used optimally [3], the government fails to prioritize areas that will become the main activity in the development process [4] and there are political reasons like the intention to fulfil the interests of certain people or groups and decentralized fiscal policy [5],[6],[7].

There are many methods to identify regional development disparities, especially in the regional economy. One of them is the Klassen typology. The Klassen method classifies regions based on economic growth indicators [8], [9]. The results of the Klassen grouping divide the region into four quadrants. Quadrant I is a group of developed and rapidly growing regions. Quadrant II is a group of developed but stressed regions. Quadrant III is a group of potential areas, meaning that they still have the possibility of developing rapidly. Meanwhile, quadrant IV is a group of relatively underdeveloped areas. To classify regions into these quadrants, one of the indicators used is the value of the rate of development growth. In the Klassen method, the development growth rate is calculated for the analysed area and the reference area.

However, the calculation of the growth rate is considered unfair. The growth rate of the analyzed area is calculated based on its own achievement data. It also applies to the growth rate of the reference area. No proportions of any kind are applied here. The calculation of the growth rate, both the analyzed area and the reference area, seems to be two unrelated things. In fact, data of the reference area are obtained from the total development data of a number of existing analyzed areas. The calculation of the growth rate allows a region to have a slower growth rate than other regions even though the region has a higher GRDP value achievement than other regions. Table 1 illustrates the unfair use of the growth rate.

Plagiarism Check TEM V12N2

by Tb Ai Munandar

Submission date: 20-Mar-2024 10:36AM (UTC-0400)
Submission ID: 2325791602
File name: TEMJournalMay2023_680_690.pdf (505.63K)
Word count: 6508
Character count: 34986

TEM Journal. Volume 12, Issue 2, pages 680-690, ISSN 2217-8309, DOI: 10.18421/TEM122-11, May 2023.

Extracting Patterns of Proximity in Regional Development Inequality Using Hierarchical Agglomerative Clustering

Tb Ai Munandar [1], Dwipa Handayani [1]

[1]*Faculty of Computer Science, Informatics Dept. Universitas Bhayangkara Jakarta Raya, Indonesia*

Abstract **– Equitable regional development is still difficult to implement. Many interests in determining regional development priorities make it difficult to realize the main goals of national development. Political interests and the underutilization of regional statistical data are often other obstacles in carrying out regional development. This study aims to classify regional development data to be used to analyse existing development inequality. Gross regional domestic income (GRDP) data is the main data in this study. The researcher used the hierarchical agglomerative clustering (HAC) technique. The results of the cluster can be used as a consideration for local governments to determine future regional development priorities. In addition, visualized cluster results in the form of a dendrogram can show the proximity of development inequality between regions.**

Keywords **– Regional development, development inequality, cluster technique, agglomerative clustering.**

1. Introduction

One of the phenomena arising when carrying out development in various parts of the world is the inequality of development between regions. Inequality here means that there have been differences in the level of achievement of development, welfare, and economic growth in each region [1]. Various strategies have been used to minimize development inequality [2]. Development inequality is influenced by several factors: statistical data on development achievements have not been used optimally [3], the government fails to prioritize areas that will become the main activity in the development process [4] and there are political reasons like the intention to fulfil the interests of certain people or groups and decentralized fiscal policy [5],[6],[7].

There are many methods to identify regional development disparities, especially in the regional economy. One of them is the Klassen typology. The Klassen method classifies regions based on economic growth indicators [8], [9]. The results of the Klassen grouping divide the region into four quadrants. Quadrant I is a group of developed and rapidly growing regions. Quadrant II is a group of developed but stressed regions. Quadrant III is a group of potential areas, meaning that they still have the possibility of developing rapidly. Meanwhile, quadrant IV is a group of relatively underdeveloped areas. To classify regions into these quadrants, one of the indicators used is the value of the rate of development growth. In the Klassen method, the development growth rate is calculated for the analysed area and the reference area.

However, the calculation of the growth rate is considered unfair. The growth rate of the analyzed area is calculated based on its own achievement data. It also applies to the growth rate of the reference area. No proportions of any kind are applied here. The calculation of the growth rate, both the analyzed area and the reference area, seems to be two unrelated things. In fact, data of the reference area are obtained from the total development data of a number of existing analyzed areas. The calculation of the growth rate allows a region to have a slower growth rate than other regions even though the region has a higher GRDP value achievement than other regions. Table 1 illustrates the unfair use of the growth rate.

DOI: 10.18421/TEM122-11
https://doi.org/10.18421/TEM122-11

Corresponding author: Tb Ai Munandar,
Faculty of Computer Science, Informatics Dept. Universitas Bhayangkara Jakarta Raya, Indonesia
Email: tb.aimunandar@dsn.ubharajaya.ac.id

Received: 06 February 2023.
Revised: 24 April 2023.
Accepted: 10 May 2023.
Published: 29 May 2023.

Table 1. The illustration of the unfair use of the growth rate

Region	Previous year's GDP	Current Year's GRDP	Development growth rate
Region A	247,686,648	265,892,081	0.0735
Region B	15,038,460	16,176,981	0.0757
Region C	10,119,547	11,217,718	0.1085

Table 1 shows that Region A has a higher GRDP than Regions B and C but has a lower development growth rate. Based on this reason, this research was conducted by providing alternative modelling of regional development inequality grouping using a data mining approach. Data mining provides a grouping solution based on the similarity between one data and another. This similarity value is obtained by calculating the distance between two objects. This way, inequality in the growth rate between regions with differences in the value of the GRDP indicator can be minimized. Moreover, the classification of development inequalities between regions can be processed fairly. Data mining is a technique commonly used to find new patterns and trends in data and the correlations in it [10]. In a more specific context, data mining aims to understand large amounts of unsupervised data [11]. In various cases, data mining has been widely implemented in various domains, such as education [12], [13], health [14], [15], prediction of landslide movement [16], extraction of weblog data patterns to create adaptive websites [17], and grouping of drug supplies at community health centers [18].

For the regional economy, especially the grouping of development achievements or regional development imbalances, data mining has been widely used for the past few years. Some of its uses are to analyze the development inequality with fuzzy clustering in Ukraine [19], to explore the regional inequality in the Republic of Croatia and Portugal using socio-economic indicators [20], [21], and to analyze the inequality in European Union states based on regional competitiveness [22], [23]. In addition, data mining is also used to analyze regional inequality in the Czech Republic [24] and Pakistan [25] and to provide development policy recommendations based on the results of the grouping of development inequality [26].

Most research applies the exclusive clustering technique, where areas are grouped with a firm degree of membership. The results of the grouping are not able to find patterns of proximity to inequality and development achievements between regions. Therefore, this study presents alternative modelling of regional clustering using an overlapping clustering technique. This research begins with a question about how to group development inequalities between regions and at the same time knowing how close the inequality is to the overall data analyzed (in this case, regional development data). So that policymakers (such as the government) know that an area can not only be grouped into a cluster, but can also find out the closeness of development inequality between regions, both inside and outside the cluster that is formed. One alternative grouping technique that can be used is hierarchical agglomerative clustering (HAC). This technique is not only able to classify inequalities between regions but also visualizes the pattern of the proximity of inequalities. The grouping of data is based on the similarity value between the data used. Thus, grouping development inequalities using the HAC technique is believed to give more representative results.

There are four HAC techniques used, namely single, complete, average, and ward linkage methods. This study uses GRDP data from one of the provinces in Indonesia, namely Banten Province. There are eight regions (consisting of regencies and municipalities) which are the main focus of this research. The study results show that the eight regencies/ municipalities in Banten Province are generally grouped into two large cluster groups. The first cluster group (C1) consists of Tangerang City and Tangerang Regency. Meanwhile, the second cluster group (C2) consists of Lebak Regency, Serang City, Pandeglang Regency, Cilegon City and South Tangerang City. The results of the cluster show that one region with another region has a close inequality of development. For example, the Tangerang City and Tangerang Regency, have close inequality of development based on the clusters formed. Furthermore, if we look at the GRDP data for the two regions, they both have adjacent GRDP data. This also applies to other cluster results. The historical aspect and geographical location of the regions do not significantly influence the results of the cluster.

2. Research Method and Data

This section is comprised of four subsections. The first subsection describes the phases of the research; the second, the problem-solving techniques utilized in the research; the third, the data requirements; and the fourth, the analytical tool employed.

2.1. Research Method

This research begins with several preliminary activities, such as literature study related to the concept of development inequality, the use of analytical methods from the regional economic perspective, and the use of computational techniques for data grouping. The next stage is data collection.

After data collection, the next step is to apply general data mining, starting from data cleaning, data pre-processing, and transformation to data analysis activities, namely grouping GRDP data using hierarchical agglomerative clustering (HAC). There are two outcomes resulting from the grouping activities in this study: first, information on regional inequality clusters; second, the pattern of regional inequality. The grouping results are then interpreted to gain new knowledge by comparing the original data and the actual conditions of the research area. In general, the model for clustering patterns of proximity to regional development inequality based on data mining in this study is shown in Figure 1.

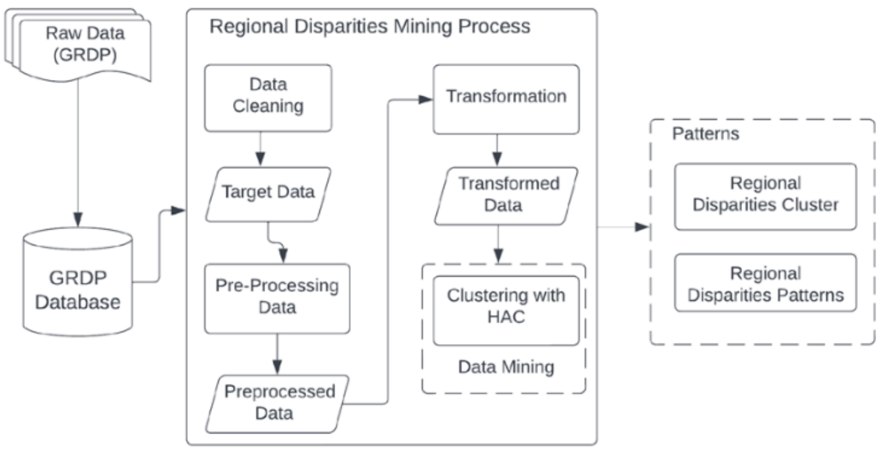

Figure 1. Model of Grouping Patterns of Proximity of Regional Development Inequality

2.2. Hierarchical Agglomerative Clustering (HAC)

Hierarchical clustering is an unsupervised algorithm that groups data with two mechanisms; agglomerative or divisive. In this study, agglomerative mechanism is used as analysis technique to analyse GRDP data. An agglomerative algorithm is grouping all data with bottom-up approach. After all initial steps in data mining model were fulfilled (from data cleaning until transformed data), HAC was then utilized to group the GRDP, generating two outputs. Each instance of data is represented by a single cluster, which is then merged until all clusters have been merged into a single cluster containing all data. Here, the stages of data cluster using HAC:

1. Preparing the data
2. Compute distance matrix. i.e. using Euclidean Distance, Equation (1) as expressed:

$$d_{euc}(x,y) = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/2} \quad (1)$$

Where, x and y are two vectors of length n.

3. Grouping two closeness of a single cluster. We can use some technique, i.e. single, complete, average or ward linkage. Here, the form of linkage function:
   - Single linkage – Equation (2):
     $$f = \min(d(x,y)) \quad (2)$$
   - Complete linkage – Equation (3):
     $$f = \max(d(x,y)) \quad (3)$$
   - Average linkage – Equation (4):
     $$f = \text{average}(d(x,y)) \quad (4)$$
   - Ward linkage – Equation (5):
     $$f = ESS(XY) - [ESS(X) + ESS(Y)] \quad (5)$$

   Where the ESS is an Error Sum Square described by the following Equation (6):

   $$ESS(X) = \sum_{i=1}^{Nx}\left|x_i - \frac{1}{Nx}\sum_{j=1}^{Nx}x_j\right|^2 \quad (6)$$

   Where |.| is an absolute value of a vector

4. Updating distance matrix
5. Repeat step 2 and 3 until one cluster left
6. Visualize the dendrogram

2.3. Data

The data used in this study is gross regional domestic income (GRDP) at constant prices for eight districts/cities in Banten Province from 2005 to 2019. The GRDP data consists of 17 sub-sectors that affect the level of regional development achievement. The GRDP sub-sectors are agricultural, forestry and fishery; mining and digging; industrial processing; gas and electricity; water supply; construction; big and small trading, and vehicles reparation; transportation and warehousing; accommodation and food supply; information and communication; financial services; real estate; company services; government administration, defense and mandatory social security; educational; health and social activities; and another services.

The data was taken from the official website of the Indonesia Statistics (BPS) of Banten Province and each city or district. In this study, we grouped data into periodic groups and analyzed them to identify additional patterns of regional inequality development. i.e. the data year 2005 and 2006 were grouped into 1st periodic data, namely 2005-2006.

Kolmogorov-Smirnov is an implement to test the normality of the data. Generally, data is normally distributing, except one data (specifically 2005) that has a Asymp.Sig. (2-tailed) value, lower than 0.05. We hypothesize that the non-availability of GRDP data for some regions in 2005 explains this anomalous circumstance. We also use Skewness and Kurtosis value to test the normality of the data. Both technique shows that all the data are normally distributing (with value Skewness and Kurtosis between -1 until +1).

2.4. Analysis Tool

This research uses ORANGE software as an analysis tool. GRDP data used in this study are regional development data of Banten Province. The data grouping stage using ORANGE begins with designing workflows in the work window. There are seven workflow components used for the clustering stage. The first one is the File component as the data source. The dataset is saved in the *.xls file extension format and then configured to be retrieved with the Files component. The contents of the dataset are represented by the Data Table component. Distance measurement is carried out using the Distance component, then presented using the Distance Matrix component by applying the Euclidean Distance technique. The data normalization stage aims to overcome the significant differences in the range of GRDP values between regions.

The next stage is the grouping of development data of which distances have been calculated. In this study, the clustering technique uses the Hierarchical Clustering component provided by ORANGE. Several configurations are done to get the expected result. The configuration in question is in the form of selecting the hierarchical cluster method used, namely single linkage, average linkage, complete linkage, and ward linkage. The grouping results are presented using two components. The first component is the Box Plot, which functions to see the value of the cluster results in terms of their quartile values. The second component is a Data Table, which presents clustered data.

3. Results

There are four scenarios for grouping development inequality data. They are single, complete, average, and ward linkage methods. The following explains each data grouping scenario that has been carried out.

3.1. Single Linkage Clustering Results

Theoretically, the single-linkage method groups data based on the shortest distance of two data objects whose distance has been measured. Each object/data point is seen as a single cluster measured by its distance from one another.

Figure 2 shows the results of the hierarchical grouping of data using the single-linkage method in the form of a dendrogram. In the dendrogram, it can be seen that all regions are grouped into four cluster groups. The first cluster group (C1) contains three regions, namely Pandeglang Regency, Lebak Regency, and Serang City. In the initial grouping, Lebak Regency was grouped with Serang City into one new cluster group, then the results of the cluster group of the two regions were regrouped into one new cluster group, namely C1.

The next cluster group is C2, which contains three regions, namely Serang Regency and South Tangerang City, which are then grouped into one cluster group, and the cluster group formed is regrouped with Cilegon City into a new cluster group, namely C2. The next cluster groups are C3 and C4, each of which only consists of one area, namely Tangerang City for C3 and Tangerang Regency for C4. The C3 and C4 cluster groups then form a new cluster group that is combined with the C2 cluster group. In the end, the C2 cluster group is then combined with C1 into a single cluster group.

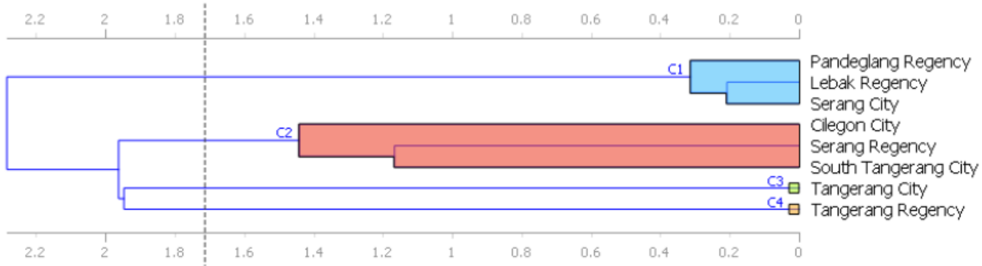

Figure 2. Results of Single Linkage Hierarchical Clustering

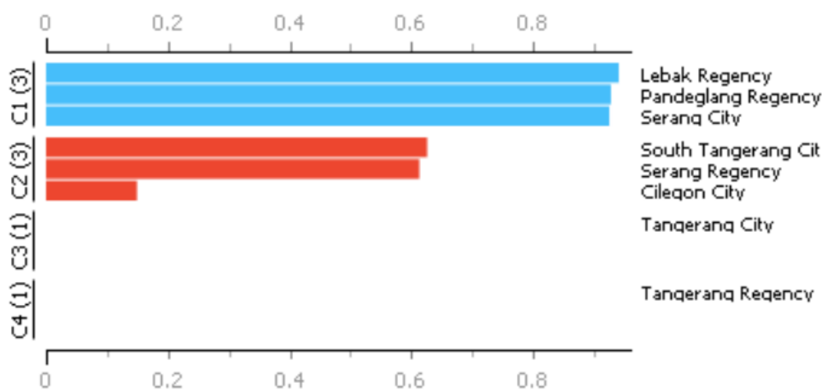

Figure 3. Silhouette Score Visualization of Single Linkage Clustering Results

Visualization of clusters in the form of a dendrogram, as shown in Figure 2, shows areas with the proximity of development inequalities. For example, based on the exploration of the GRDP dataset from 2005 to 2019, Lebak Regency and Serang City turned out to have development achievements with GRDP values that were close to each other. Therefore, from a visual perspective and the results of measuring the distance between data objects, it is very reasonable for the two regions to be grouped into one adjacent cluster group even though Serang City and Lebak Regency are geographically not very close. It means that the position of Serang City is separated by Serang Regency. Lebak Regency is geographically actually side by side with Pandeglang Regency, so the development achievement data have almost the same characteristics. Therefore, based on the GRDP dataset obtained and the results of distance measurements, Lebak and Pandeglang regencies are grouped into the same cluster group. This condition also applies to other regions. The results of the cluster formed represent the proximity of the development achievements of one region to another. In this case, geographic location does not significantly affect cluster results. However, in certain areas, the geographical location also has a close position to one another, if explored further. For Serang Regency, Cilegon City and South Tangerang City are grouped into the same cluster group. Meanwhile, Tangerang City and Tangerang Regency are a single cluster, which is then combined into one large cluster group. When viewed from the GRDP dataset, the two regions have GRDP values far above the other six regions.

To test the consistency of cluster results, this study also performed measurements using the Silhouette method. The Silhouette plot component is added to the workflow to test the consistency of each data object when it is a member of a particular cluster. A high silhouette score indicates that the cluster is very good. Table 2 shows the silhouette scores of each cluster group, while Figure 3 shows the visualization of the silhouette scores.

Table 2 shows that most C1 cluster members have a silhouette score close to 1, so this cluster is very good. Meanwhile, the C2 cluster has a silhouette score ranging from 0.6, while one member has a score of 0.15. A very significant difference is shown in clusters C3 and C4, with a silhouette score of zero.

Table 2. Silhouette Score of Clustering Results with Single Linkage Method

No	Name of Regency	Cluster	Silhouette (Cluster)
1	Pandeglang Regency	C1	0.947481
2	Lebak Regency	C1	0.960455
3	Cilegon City	C2	0.15237
4	Serang City	C1	0.944494
5	Serang Regency	C2	0.626007
6	South Tangerang City	C2	0.639075
7	Tangerang City	C3	0
8	Tangerang Regency	C4	0

3.2. *Complete Linkage Clustering Results*

Complete linkage groups data objects into a new cluster group based on the furthest distance between two single clusters (data objects), and this is, of course, contrary to the single linkage method, wherein single linkage, two data objects are grouped based on the shortest distance between them. That is, visually, the results of the cluster dendrogram are the opposite of a single linkage. Figure 4 shows a dendrogram of cluster results with complete linkage; the grouping results are divided into two main cluster groups, C1 and C2; the C1 cluster group has members in the form of Tangerang Regency and Tangerang City. C2 contains the results of five times the grouping process from the previous single and group clusters. For example, a single cluster of Lebak Regency and Serang City then forms a cluster group. The combined results of the two are then combined into one cluster group with Pandeglang Regency (to make it easier to pronounce, the result of combining these three areas is called C2.1).

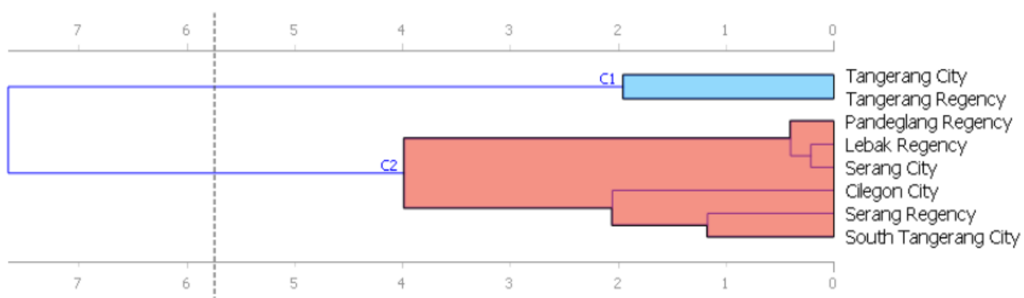

Figure 4. Clustering Results with Complete Linkage Method

South Tangerang City and Serang Regency were combined into one cluster group, then combined with Cilegon City to form a new cluster group (C2.2). The results from C2.1 and C2.2 are then combined into one large cluster group, namely C2. The main cluster in this complete linkage is formed from C1 and C2. Visually as in Figure 3, it can be seen that the position of the cluster group formed for each region has not changed; what is different is the location of the cluster group formed, and this is very reasonable considering that complete linkage groups data based on the furthest distance from the two data objects.

As with the single linkage method, the consistency of the cluster results is also checked by looking at the silhouette score in this complete linkage. Table 3 shows that cluster C1 has a silhouette value that is higher than C2 or tends to be close to the value 1. So it can be said that the cluster formed is very good. As for the results of cluster C2, most of them are dominated by scores that tend to approach 1, although there is one member of the cluster, namely Cilegon City, which has a negative silhouette score (-0.21576). Figure 5 shows the silhouette scores for all cluster members using the complete linkage method.

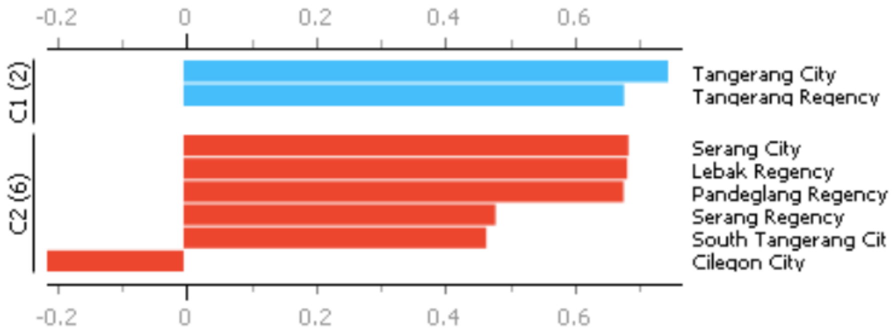

Figure 5. Visualization of Silhouette Score of Clustering Results with Complete Linkage Method

3.3. Average Linkage Clustering Results

The concept of grouping with the average linkage method emphasizes the merging of two data objects based on the average distance between them.

The results of regional development data grouping with average linkage are shown in Figure 6.

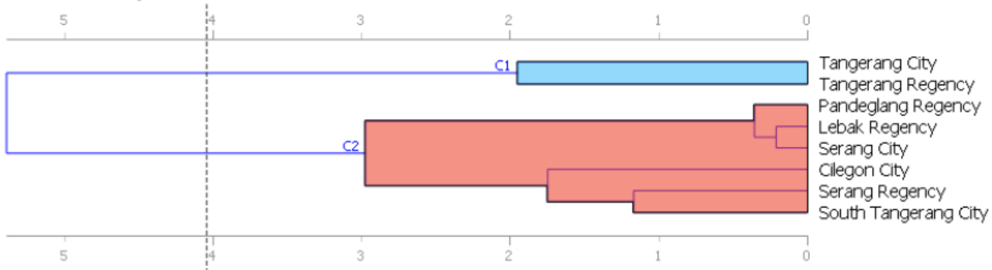

Figure 6. Clustering Results with Average Linkage Method

Table 3. Silhouette Score of Clustering Results with Average Linkage Method

No	Name of Regency	Cluster	Silhouette (Cluster)
1	Pandeglang Regency	C2	0.695003
2	Lebak Regency	C2	0.700629
3	Cilegon City	C2	-0.21576
4	Serang City	C2	0.702694
5	Serang Regency	C2	0.492703
6	South Tangerang City	C2	0.478032
7	Tangerang City	C1	0.765408
8	Tangerang Regency	C1	0.695779

The grouping results using the average linkage method are not different from that of complete linkage. The number of main group clusters and the cluster group's position for each region have also remained unchanged. It should also be noted that the configuration of distance and other measurements is no different from the single or complete linkage method. In other words, in the case of this study, it can be said that the cluster results between average linkage and complete have the same results.

At the same time, the silhouette score for the average linkage is shown in Table 3.

3.4. Ward Linkage Cluster Results

The main procedure of the ward linkage method is group two data objects (single cluster) based on the total sum of squares in each variable. This method is indeed more different than single, complete, and average linkage. The results of the cluster with ward linkage are shown in Figure 7.

The cluster results for the three HAC techniques (i.e. complete, average, and ward linkage) have the same outcomes. Visually, the dendrogram is not different, even though it has different silhouette scores from one another.

In clusters using the ward linkage technique, Tangerang City and Tangerang Regency are grouped into one cluster group (C1). In contrast, other regencies/cities are grouped into another cluster group (C2). The two cluster groups were then regrouped into a new single cluster forming a hierarchy (see Figure 7).

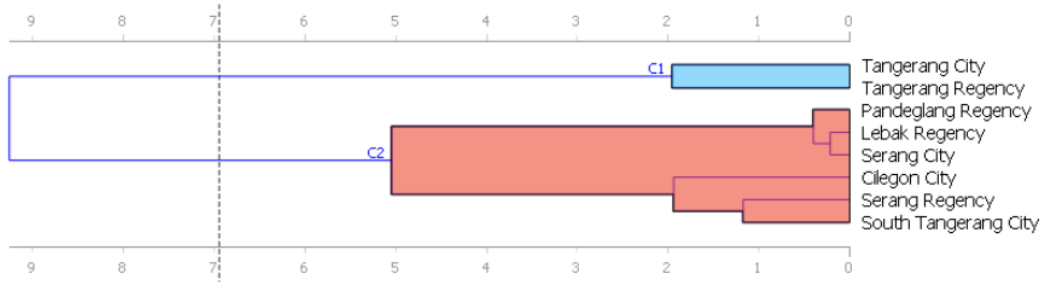

Figure 7. Clustering Results with Ward Linkage

4. Discussion

In the discussion section, we discuss two aspects of the research findings. First, a discussion of variations in regional development inequality patterns based on periodic cluster results. Second, an analysis of cluster outcomes regarding regional conditions and development accomplishments.

4.1. Changes in the Pattern of Proximity of Development Inequality Clustering Results

Data clustering scenarios are applied to four different HAC techniques. To find out more about the extent to which the patterns of the proximity of inequality between regions have changed, this study also carried out an additional scenario, namely the grouping of GRDP data every two years. For example, based on data from 2005-2006, 2006-2007, and so on until 2018-2019.

Table 4. Patterns of Changes in the Position of Development Inequality

Name of Regency	Periods						
	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012
Lebak Regency	C7	C7	C5	C5	C2	C2	C3
Pandeglang Regency	C5	C5	C4	C4	C2	C1	C2
Tangerang Regency	C2	C2	C6	C8	C8	C5	C5
Tangerang City	C1	C1	C1	C6	C6	C4	C4
South Tangerang City	C6	C8	C3	C3	C5	C8	C8
Cilegon City	C3	C3	C7	C7	C7	C6	C6
Serang City	C8	C6	C2	C2	C1	C3	C1
Serang Regency	C4	C4	C8	C1	C4	C7	C7
Name of Regency	**Periods**						
	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019
Lebak Regency	C3	C3	C3	C3	C3	C3	C3
Pandeglang Regency	C2	C2	C2	C2	C2	C2	C2
Tangerang Regency	C5	C5	C5	C5	C5	C5	C5
Tangerang City	C4	C4	C4	C4	C4	C4	C4
South Tangerang City	C8	C8	C8	C8	C8	C8	C8
Cilegon City	C6	C6	C6	C6	C6	C6	C6
Serang City	C1	C1	C1	C1	C1	C1	C1
Serang Regency	C7	C7	C7	C7	C7	C7	C7

The grouping results show that the pattern of changes in the position of proximity to development inequality occurred very significantly from 2005 to 2011. From 2011 to 2019, the pattern of changes in the position of proximity to development inequality tended not to happen.

Table 4 shows the position of the cluster for each region based on the data period every two years. C1 notation represents the first single cluster, and so on for C2 to C8. The presentation of cluster results into a single cluster is intended to see changes in patterns more clearly. For example, in Table 4, Lebak Regency experienced a change in its cluster position from 2006-2007 to 2007-2008 and from 2008-2009 to 2009-2010. The rest did not experience a change in cluster position. The reading of the table matrix of the pattern of the proximity of inequality between regions is read with two approaches. The row Approach represents the cluster change for each data period. The column approach represents the pattern of the proximity of development inequality between regions. For example, Tangerang City (C1) and Tangerang Regency (C2) have a pattern of unequal proximity to each other. Tangerang Regency (C2) and Cilegon City (C3) are also indicated to have a close pattern of development inequality between the two. And so on for other areas.

4.2. Analysis of Clustering Results on Regional Conditions

The next stage of this research is the interpretation of cluster results on the conditions of each region. The interpretation stage is done by looking at two conditions. The first one is based on the cluster results obtained, and the second is determined by the worth of development accomplishments and a bit of historical development of the area that is being interpreted. Based on the first point of view, it does not differ from one technique to another. A significant difference is seen in the single linkage cluster results. However, the areas that are the outputs of single linkage clusters are no different from other regions. Single linkage clusters the region into three cluster groups. The other three techniques (complete, average, and ward) cluster the region into two groups. In this interpretation stage, we will interpret the cluster results based on the ward linkage approach to facilitate the discussion. Although it is possible to use the results from other techniques because all of them give the same cluster group results, the researcher chooses one technique to discuss.

As shown in Figure 6, Tangerang City and Tangerang Regency are grouped into the same cluster group.

Geographically, these two regions have close positions separated by certain administrative boundaries. So it is very reasonable if the two regions are grouped in the same cluster group. However, the interpretation of the cluster output by looking at the geographical location alone is certainly not enough. Pay attention to the City of South Tangerang. Geographically, South Tangerang City is also close to Tangerang City and Tangerang Regency. However, why is South Tangerang clustered into a different cluster group (not together with Tangerang City and Tangerang Regency)? To answer this, we can look at development achievement data based on the GRDP value of each regency/city.

Based on the GRDP data for each regency/city, it is known that the development achievements of South Tangerang City tend to approach the development achievements of other regencies/cities which are geographically not close to the Greater Tangerang area, and this is why South Tangerang City is not grouped into the same cluster group as Tangerang City and Tangerang Regency. Historically, South Tangerang City is a new area in Tangerang Raya and was only established in 2008. In contrast to Tangerang City and Tangerang Regency, which have been around for a long time, it is natural to have development achievements far from their neighboring areas (South Tangerang City). There is a possibility in the next five to ten years that the City of South Tangerang will accelerate away from the achievements of other regional developments that are one cluster group at the same time. It can be seen in increasing development achievement (GRDP value), which is significant every year.

Therefore, based on the GRDP value data, we can answer why South Tangerang City is not grouped with the other two regions in the same cluster group even though they are geographically located. The visualization of the Dendrogram ward linkage, as shown in Figure 6, also shows that Serang City and Lebak Regency are grouped into the same cluster group. Then the two are clustered into one new cluster group. Geographically, Lebak Regency is closer to Pandeglang Regency, while Serang City is close to Serang Regency. Serang City is geographically located far from both Lebak Regency and Pandeglang Regency. However, Serang City and Lebak Regency were clustered into the same cluster group for the first time. Then, Pandeglang Regency was combined into one cluster group.

Geographically, it does not make sense, but by looking at the GRDP data owned by Serang City and Lebak Regency, it is seen the reason Serang City and Lebak Regency clustered into one cluster group. Likewise, premises other regions.

Serang Regency is clustered with South Tangerang City to form a new cluster group, then merged with Cilegon City to create a larger cluster group. They were combined with the combined cluster group of Serang City and Lebak Regency, and Pandeglang Regency to form a larger cluster group, namely C2. Historically and geographically, some are close together.

However, each region has different development achievements, so adjacent regions are not clustered in one group and vice versa.

The results of the formed clusters also show how the pattern of regional inequality is close to one another. C1 cluster group, for example, contains three areas that have development inequalities that are close to each other. Tangerang City and Tangerang Regency have a rapid development gap. The inequality range is quite far compared to the regencies'/cities members of the C2 cluster group. The C2 cluster group is divided into two sub-cluster groups representing the closeness of development inequality between regions. The first sub-cluster consists of Serang City and Lebak Regency, which have the closest inequality level. These two regions are clustered based on the closest inequality level to other regions, namely Pandeglang Regency. The second sub-cluster of C2 shows South Tangerang City and Serang Regency, which have close development inequality, which forms a sub-cluster and then merges with Cilegon City to form a sub-cluster sub-group which will eventually form the C2 cluster group. This pattern of proximity to development inequality can only be seen with the concept of hierarchical grouping. Each region with a close level of inequality will be clustered according to its development achievements.

5. Conclusion

The clustering results with HAC using four cluster techniques show that Tangerang City and Tangerang Regency are clustered into the same cluster group (C1). In contrast, other areas, namely South Tangerang City, Serang City, Cilegon City, Pandeglang, and Lebak Regencies, are also clustered into different cluster groups (C2). The grouping of the eight regions in Banten Province using the cluster technique, indicates the closeness of development inequality between regions. Cluster members in C1 and C2 contain areas with development inequality that are close to each other. This can be confirmed from the regional GRDP value of each member of the C1 cluster. This also shows that the value of GRDP has a fairly close relationship to determine the grouping of development inequality in a region.

Geographical and historical analyses were also carried out to interpret the grouping results, although they did not have a significant relationship.

Some areas are geographically close together, such as Lebak Regency with Tangerang Regency, Pandeglang Regency with Serang Regency and South Tangerang City with Tangerang City. However, geographical location does not affect the results of the clusters formed.

For example, although Lebak Regency is geographically close to Tangerang Regency, the two are not in the same cluster. Likewise, between the City of Tangerang and the City of South Tangerang, geographically the two are close together, but in fact they are classified in different groups of development inequalities. Regional groupings do not consider the historical and geographical location of regencies/cities. However, they are based on data on the value of regional development achievements in the form of GRDP. With this hierarchical grouping, it is seen the closeness of development inequality between one region and another so that it can be taken into consideration by local governments to determine priorities for future development directions: which regencies/cities need top priority, and which ones should be lowered so that the distribution of regional development in Banten Province can be realized.

Acknowledgment

This research was supported by a grant of the the Institute for Research, Community Service and Publications, funded by Universitas Bhayangkara Jakarta Raya.

References

[1]. Lipshitz, G. (1993). The main approaches to measuring regional development and welfare. *Social Indicators research*, 163-181.

[2]. Baransano, M. A., Putri, E. I. K., Achsani, N. A., & Kolopaking, L. M. (2016). Analysis of factors affecting regional development disparity in the province of West Papua. *Journal of Economics and Development Studies*, *4*(2), 115-128.,

[3]. Sipayung, H.M. (2013). *Pemanfaatan Data Statistik Belum Optimal*, Kantor Berita ANTARA. Retrieved from: http://www.antarabengkulu.com/berita/17829/pemanfaatan-data-statistik-belum-optimal [accessed: 21 January 2023]

[4]. Redaksi. (2014). *Rendah, IPM Banten Urutan 23 dari 33 Provinsi*, Radarbanten. Retrieved from: https://www.radarbanten.co.id/rendah-ipm-banten-urutan-23-dari-33-provinsi/ [accessed: 03 February 2023]

[5]. Wawengkang, Y. (2014). Darurat Demokrasi Akibat Kepentingan Bisnis. Retrieved from: https://www.kompasiana.com/joewawengkang/54f85e0ea33311315e8b4b4a/darurat-demokrasi-akibat-kepentingan-bisnis [accessed: 15 January 2023].

[6]. Lessmann, C. (2012). Regional inequality and decentralization: an empirical analysis. *Environment and Planning*, *44*(6), 1363-1388.

[7]. Swastyari, D. (2008). *Regional Inequality in Indonesia. Is the General allocation Fund (DAU) Likely to Have an Impact?* [Masters Dissertation, Institute of Social Studies] The Netherlands: The Hague.

[8]. Prahara, I., Nugraha, M. R. S., Astabella, R. D., Pramesti, R. A., Hayat, D. M., & Putri, R. F. (2021). Analysis of Economic Inequality and Human Development: A Case Study of West Sumatra Province in 2015-2019. In *E3S Web of Conferences, 325*, 06001. EDP Sciences. Doi: 10.1051/e3sconf/202132506001

[9]. Carvalho, R. D. S., Diniz, A. S., Lacerda, F. M., & Mello, P. A. D. A. (2012). Gross domestic product (GDP) per capita and geographical distribution of ophthalmologists in Brazil. *Arquivos Brasileiros de Oftalmologia*, *75*, 407-411. Doi: 10.1590/S0004-27492012000600007

[10]. Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining*, *4*. John Wiley & Sons.

[11]. Cios, K.J., Pedrycz, W., Swiniarski, R.W., dan Kurgan, L.A. (2007). *Data Mining : A Knowledge Discovery Approach*, Springer.

[12]. Krizanic, S. (2020). Educational data mining using cluster analysis and decision tree technique: A case study, *International Journal of Engineering Business Management*,12. Doi: 10.1177/1847979020908675

[13]. Hooshyar, D., Pedaste, M., & Yang, Y. (2019). Mining educational data to predict students' performance through procrastination behavior. *Entropy*, *22*(1), 12. Doi: 10.3390/e22010012

[14]. Asri, H., Mousannif, H., & Al Moatassime, H. (2019). Reality mining and predictive analytics for building smart applications. *Journal of Big Data*, *6*, 1-25. Doi: 10.1186/s40537-019-0227-y

[15]. Mirmozaffari, M., Alinezhad, A., & Gilanpour, A. (2017). Heart disease prediction with data mining clustering algorithms. *Int'l Journal of Computing, Communications & Instrumentation Engg*, *4*(1), 16-19.

[16]. Miao, F., Xie, X., Wu, Y., & Zhao, F. (2022). Data Mining and deep learning for predicting the displacement of "Step-like" landslides. *Sensors*, *22*(2), 481. Doi: 10.3390/s22020481

[17]. Oo, H.Z. & Kham, N.S.M. (2018). Pattern Discovery Using Association Rule Mining on Clustered Data, *International Journal of New Technology and Research (IJNTR), 4*(2), 7-11.

[18]. Nugraha, J. A. M. (2019). Drug Inventory Grouping using Clustering Data Mining. *Indonesian Journal of Information Systems (IJIS)*, *2*(1), 1-12.

[19]. Gorbatiuk, K., Mantalyuk, O., Proskurovych, O., & Valkov, O. (2019). Analysis of regional development disparities in Ukraine with fuzzy clustering technique, *SHS Web of Conferences 65*, 04008. Doi: 10.1051/shsconf/20196504008

[20]. Soares, J.O., Marques, M.M.L & Monteiro, C.M.F. (2003). A Multivariate Methodology to Uncover Regional Disparities: A Contribution to Improve European Union and Governmental Decisions, *European Journal of Operational Research, 145*, 121-135.

[21]. Rašić Bakarić, I. (2005). Uncovering regional disparities–the use of factor and cluster analysis. *Privredna kretanja i ekonomska politika*, *15*(105), 52-76.

[22]. Lukovics, M. (2009). Measuring regional disparities on competitiveness basis. *Regional Competitiveness, Innovation and Environment*, 39-53.

[23]. Poledníková, Eva. (2014). Comparing regions ranking by MCDM methods: The case of Visegrad countries. *WSEAS Transactions on Business and Economics, 11*, 496-507.

[24]. Vydrová, H. V., & Novotná, Z. (2012). Evaluation of Disparities in Living Standards of Regions of the Czech Republic. *Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensis*, *60*(4), 407-414.

[25]. Ramzan, S.., Khan, M.I, & Zahid F.M. (2013). Regional Development Assessment Based on Socioeconomic Factors in Pakistan Using Cluster Analysis. *World Applied Sciences Journal, 21*(2), 284-292.

[26]. Kosova, O. (2013). The Innovation Development in Ukraine: Problems and Development Perspectives. *International Journal of Innovation and Business Strategy* (IJIBS), 2(1).

Plagiarism Check TEM V12N2

ORIGINALITY REPORT

14%	12%	9%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Tb. AiMunandar, Edi Winarko. "Regional Development Classification Model using Decision Tree Approach", 'Foundation of Computer Science', 2015 Internet Source	1%
2	ejournal.undiksha.ac.id Internet Source	1%
3	lib.iitta.gov.ua Internet Source	1%
4	ojs.unm.ac.id Internet Source	1%
5	Submitted to Polytechnic Institute Australia Student Paper	1%
6	dspace.cuni.cz Internet Source	1%
7	W. K. Wong. "A hybrid approach for packing irregular patterns using evolutionary strategies and neural network", International Journal of Production Research, 11/17/2009 Publication	<1%

8	ijrrjournal.com Internet Source	<1 %
9	kb.psu.ac.th Internet Source	<1 %
10	www.frontiersin.org Internet Source	<1 %
11	scholar.unair.ac.id Internet Source	<1 %
12	Submitted to American Public University System Student Paper	<1 %
13	Peter Bajorski. "Statistics for Imaging, Optics, and Photonics", Wiley, 2011 Publication	<1 %
14	hal.archives-ouvertes.fr Internet Source	<1 %
15	www.igi-global.com Internet Source	<1 %
16	Submitted to Kocaeli Üniversitesi Student Paper	<1 %
17	Submitted to School of Business and Management ITB Student Paper	<1 %
18	Submitted to Global Banking Training Student Paper	<1 %

19	hrcak.srce.hr Internet Source	<1 %
20	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
21	pure.tue.nl Internet Source	<1 %
22	Submitted to Erasmus University of Rotterdam Student Paper	<1 %
23	Submitted to Technical University of Crete Student Paper	<1 %
24	repository.upiyptk.ac.id Internet Source	<1 %
25	journalofscience.ou.edu.vn Internet Source	<1 %
26	Submitted to Indiana University Student Paper	<1 %
27	Submitted to University of Sydney Student Paper	<1 %
28	Submitted to University of North Texas Student Paper	<1 %
29	ageconsearch.umn.edu Internet Source	<1 %

30	garuda.kemdikbud.go.id Internet Source	<1 %
31	Mariam Melhem, Bouchra Ananou, Mustapha Ouladsine, Michel Combal, Jacques Pinaton. "Product quality prediction using alarm data : Application to the semiconductor manufacturing process", 2017 25th Mediterranean Conference on Control and Automation (MED), 2017 Publication	<1 %
32	repec.mnje.com Internet Source	<1 %
33	www.research.unipd.it Internet Source	<1 %
34	Abraham Cardenas Tristan, Ashilah Putri Mandiri, Hani Verdiant, Yosi Yoseva, Ratih Fitria Putri. "Analysis of Carrying Capacity of Crop Agricultural Land in Banten Province", E3S Web of Conferences, 2023 Publication	<1 %
35	ijict.iaescore.com Internet Source	<1 %
36	ejurnal.ubharajaya.ac.id Internet Source	<1 %
37	www.diva-portal.org Internet Source	<1 %

38	Gabriel Lipshitz. "The main approaches to measuring regional development and welfare", Social Indicators Research, 1993 Publication	<1 %
39	Quan Thanh Tho. "Web Mining for Identifying Research Trends", Lecture Notes in Computer Science, 2003 Publication	<1 %
40	R S Septianingrum, B Aditya, A P Wijaya, M R Irshabdillah, N A Husna, R F Putri. "Economic inequality and its impact on human development: insight of Banten Province 2011-2015", IOP Conference Series: Earth and Environmental Science, 2020 Publication	<1 %
41	Wang, Y.Z.. "Using genetic algorithm methods to solve course scheduling problems", Expert Systems With Applications, 200307 Publication	<1 %
42	jp.feb.unsoed.ac.id Internet Source	<1 %
43	jurnal.untan.ac.id Internet Source	<1 %
44	"Information Systems and Technologies to Support Learning", Springer Science and Business Media LLC, 2019 Publication	<1 %

45	sc.hec.gov.pk Internet Source	<1 %
46	Liu, Huadong. "CLUSTERING: HIERARCHICAL ALGORITHMS", Lecture Notes in Data Mining, 2006. Publication	<1 %
47	W Meilvidiri, A S Ulita, Asrudi, Wa Ode Alzarliani. "Comparison of competitive model advantage tools in the economic potential of north Kolaka Regency", IOP Conference Series: Earth and Environmental Science, 2019 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		