Vol. 3 No. (1) (June 2025) 48-55

Journal of Data Insights

http://journalnew.unimus.ac.id/index.php/jodi

Implementation of Hierarchical Clustering for Grouping Economic Development Indicators in Central Java Province

Yusrisma Asyfani [1], **Salmaa Fauziah**[2] ,**Indah Manfaati Nur**[3]
[1]Universitas Muhammadiyah Semarang, Indonesia
[2] Universitas Muhammadiyah Semarang, Indonesia
[3] Universitas Muhammadiyah Semarang, Indonesia

DOI: https://doi.org/10.26714/jodi.v3i1.298

Article Information

Article History:
Submitted: 04th Jan 2024
Revised: 23th June 2025
Accepted: 27th June 2025

Keywords:
Average Linkage; Economic Development; Hierarki Clustering.

Abstract

In the midst of global economic shifts, the economy in Indonesia must continue to improve. To help economic recovery after the contraction caused by the COVID-19 pandemic, the Indonesian government has implemented various policies. One way is through the process of increasing per capita income over a long period of time, known as economic development, provided that the number of people living below the absolute poverty line does not increase and income distribution does not decrease. Other efforts can be made by analyzing economic development indicators. One method that can be used is hierarchical cluster analysis to group economic development indicators in Central Java province. Average linkage is used as an approach method after carrying out correlation analysis of the five approaches in hierarchical analysis because the correlation value is the highest. From this analysis two clusters were produced with the first cluster having higher characteristic values compared to the second cluster.

✉ Alamat Korespondensi:
E-mail: salmaafauziah405@gmail.com

e-ISSN: 2988 - 2109

INTRODUCTION

The most frequently used factor to describe a country's success in managing the welfare and prosperity of its people is its economic level. In addition to being one of the developing countries, Indonesia has the largest economy in Southeast Asia. Indonesia's status as a middle-income country and its membership in the G20 makes it one of the newly industrialized countries (Bimawa, 2019). In 2019, the value of Indonesia's digital economy reached US$40 billion, with a projection of reaching US$130 billion by 2025. Due to the majority of central government ownership in 141 State-Owned Enterprises (BUMN), Indonesia is highly dependent on the domestic market and government spending (Tanudi, 2023).

Despite being in the midst of a global economic shift, the Indonesian economy continues to grow. In addition, to help the economy recover after the contraction caused by the COVID-19 pandemic, the Indonesian government has implemented various policies. One of them is through the process of increasing per capita income over a long period, known as economic development, with the note that the number of people living below the absolute poverty line does not increase and income distribution does not decrease (Matdoan and Delsen, 2020).

Central Java is a province in Indonesia located on the island of Java. This province has an area of 32,548.20 km² and a population of around 34.5 million people in 2021. Because it has many large cities and strategic ports, Central Java plays an essential role in the Indonesian economy (Juhro et al., 2022). The economy of Central Java has excellent potential to develop further. The local government has tried to improve several existing sectors of the Central Java economy, such as industry, the Sharia economy, and MSMEs. In addition, infrastructure development and the development of the tourism industry are expected to improve the economy of Central Java (Teja, 2015). The increase in the economy in Central Java can be seen from the development of the GRDP (Gross Regional Domestic Product), HDI (Human Development Index), and TPAK (Labor Force Participation Rate) (Masruri, 2016). GRDP is a statistic that measures the total production of goods and services produced by a country in one year. GRDP can show economic development, indicating whether the economy is developing or not. HDI is a measuring tool used to evaluate the success of a country's human development by referring to three criteria: education, birth rate, and standard of living. HDI provides insight into community welfare and human development. TPAK can show the number of people working and the capacity of an area to generate income and employment resources (Maharani, 2017).

The right policies need to be implemented by the government in order to produce good economic growth, which is in line with the goals of SDG No. 8, namely decent work and economic growth (Bappenas, 2020). Before making policies, an intensive study is needed regarding the indicators of the success of economic development, one of the methods of which is by grouping related to economic development indicators.

One statistical method for carrying out the grouping process is cluster analysis. Cluster analysis is the classification of objects into several groups based on certain similarities. In its grouping, a measure of distance that can explain the simple group structure of complex data is used to describe the proximity between data (Putri et al., 2021). There are two clustering approaches in cluster analysis. The first is hierarchical, which is used when the desired number of clusters is not known in advance. The second is non-hierarchical, which is used when n objects are grouped into k clusters with a predetermined number of clusters (Hanandya et al., 2021).

Nurissaidah and Rafika (2020) have researched the use of hierarchical cluster analysis to group infectious disease variables in Indonesia, which resulted in 6 groups with the best ward approach method (Ulinnuh and Veriani, 2020). Widodo (2020) also used hierarchical analysis to group districts/cities in Central Java based on the HDI, which resulted in 3 clusters with the average linkage method as the best approach based on high cophenetic correlation values (Widodo et al., 2020).

METHOD

2.1 Economic Development Indicators

Economic development indicators are measures or references used to assess the success of a country in economic development. Economic development indicators refer to a collection of variables or measures used to measure and analyze the economic performance of a country, region, or area (Fuady, 2015). According to Lincoln Arsyad, development economics is an economic science that studies economic problems in developing countries and the policies that need to be implemented to realize economic development. Several indicators of the success of economic development consist of GRDP, the human development index, and the labor force participation rate (Tanudy, 2023)

Gross Regional Domestic Product (GRDP) Based on Market Prices

Gross Regional Domestic Product is the amount of gross value added arising from all economic sectors in a region (Hartono et al., 2018). Added value is the value added from a combination of production factors. GRDP can be calculated using the formula:

$$Y = r + w + i + p$$

Information:

r = rent

w = wages

i = investment

p = profit

Labor Force Participation Rate

According to Bappeda DIY (2020), the Labor Force Participation Rate (TPAK) is the comparison between the number of the workforce (employed and unemployed) and the number of the working-age population and is usually expressed in percent. Labor force participation can also be measured by TPAK, which is helpful in knowing the percentage of the working-age population (in this case, aged 15-64 years) who have the potential to be active as workers in a country. TPAK can be calculated using the formula:

$$TPAK = PK/PUK \times 100\%$$

TPAK = Labor force participation rate
PK = Working population
PUK = Working age population

Human Development Index

According to BPS (2009), the Human Development Index (HDI) is a measure of development achievement based on a number of basic components of quality of life. The Human Development Index is calculated based on data that can describe the four elements, namely life expectancy, which measures success in the health sector; literacy rates and average length of schooling, which measure success in the education sector; and the purchasing power of the community for a number of basic needs as seen from the average amount of expenditure per capita as an income approach that

measures success in the development sector for a decent life (Anggraini and Arum, 2022). The HDI of a region can be calculated using the formula:IPM $= \sqrt{\text{I}}$kesehatan x Ipendidikan x Ipengeluaran

2.2 Cluster Analysis

Cluster Analysis is a multivariate technique that aims to classify objects into different groups, including one group with another. Objects that have a relatively close distance are the same as other objects (Nafisah and Chandra, 2017). The primary purpose of the cluster method is to group a number of data/objects into clusters (groups) so that each cluster will contain data that is as similar as possible (Ramadani and Salma, 2022).

2.3 Clustering Hierarchy

The clustering algorithm must be able to maximize the relative differences of clusters to variations in clusters. The two most common methods in cluster algorithms are the hierarchical Method and the non-hierarchical Method (Haumahu and Nanlohy, 2021). There are five methods for forming hierarchical clusters namely 1. Single Linkage: This Method is based on the minimum distance starting with two objects separated by the shortest distance, then both will be placed in the first cluster, and so on, 2. Complete Linkage, also called the farthest neighbor approach, is basically the maximum distance, 3. Average Linkage is basically the average distance between observations, grouping starting from the pair of observations with the closest distance to the average distance, 4. Ward's Method, in this Method is the distance between two clusters, which is the sum of the squares between the two clusters for all variables. This Method tends to be used to combine clusters with small numbers, 5. In the Centroid Method, the distance between two clusters is the distance between the centroids of the clusters, and the cluster centroid is the middle value of observations on variables in a set of cluster variables (Agnestisia and Masruroh, 2017).

2.4 Data Collection Techniques

The type of data used in this study uses secondary data from the website of the Central Statistics Agency of Central Java Province (https://jateng.bps.go.id), namely X1 (PDRB), X2 (IPM), and X3 (TPAK) per district/city during 2022.

Table 1. Research Variables

Variable	Description
X1	Constant Gross Regional Domestic Product
X2	Human Development Index
X3	Labor Force Participation Rate

2.5 Data Analysis Stages

The research objects used were all regencies/cities in Central Java, totaling 35 regencies/cities. The analysis process was carried out using R Studio software with the following steps:

1. Inputting data and descriptive statistics
2. Conducting Assumption Tests, including KMO and Non-Multicollinearity tests
3. Data standardization and distance calculations
4. Calculation of cophenetic correlations
5. Cluster validation and dendrogram formation
6. Drawing conclusions

RESULTS AND DISCUSSION

1. Input and Descriptive Analysis

Table 2. Variable Descriptive Statistics

Kabupaten.Kota	X1	X2	X3
Length:35	Min. : 6889453	Min. :67.03	Min. :64.75
Class :character	1st Qu.: 16532047	1st Qu.:70.78	1st Qu.:68.63
Mode :character	Median : 22447731	Median :73.15	Median :70.99
	Mean : 29980939	Mean :73.50	Mean :71.12
	3rd Qu.: 31763682	3rd Qu.:75.89	3rd Qu.:73.72
	Max. :152999374	Max. :84.35	Max. :79.57

The distribution of GRDP, HDI, and TPAK data in Central Java in 2022 can be seen in the table above. The smallest GRDP figure is in Magelang City at 6889453, the largest at 152999374 in Semarang City, and the average GRDP of Central Java is 29980939; for HDI the largest value is at 84.35 in Salatiga City, the smallest at 67.03 in Brebes Regency with an average of 73.50, and for TPAK the highest figure is 79.57 in Magelang Regency, the lowest 64.75 in Banyumas Regency, the average is at 71.12 .

2. Assumption Test

Table 3. KMO Test Results

Kaiser-Meyer-Olkin factor adequacy			
Call: KMO(r = data[, 2:4])			
Overall MSA = 0.59			
MSA for each item =			
X1	X2	X3	
0.58	0.58	0.65	

In cluster analysis, assumption tests are carried out in the form of KMO and non-multicollinearity tests to see whether the sample grouping used is representative of the population and the variables used are not interrelated. From Table 3, it can be seen that each variable has a KMO value of more than 0.5, which means that the sample is sufficiently representative or represents the population. In the non-multicollinearity test, all variables have a VIF value <10, which means that there is no multicollinearity in the variables.

3. Choosing the Best Method

Table 4. Correlation Calculation

cors	corave	corcomp	corcen	corward	
1	0.9591234	0.9660774	0.8783251	0.9641566	0.8366233

The selection of the best method can be known by looking at the Cophenetic correlation value for each method. From Table 4. it can be seen that the Correlation Coefficient value with Single Linkage is 0.959, the Correlation Coefficient with Average Linkage is 0.966, the Correlation

Coefficient with Complete Linkage is 0.878, the Correlation Coefficient with Centroid Linkage is 0.964, the Correlation Coefficient with Ward's Method is 0.836 The correlation value closest to 1 is with the Average Linkage method, so it is selected as the best method.

4. Cluster Validity

Table 4. Cluster and Dendogram Validation

Score	Method	Clusters	
Connectivity	2.9289683	hierarchical	2
Dunn	0.7327652	hierarchical	2
Silhouette	0.6142752	hierarchical	2

Based on the connectivity index, Dunn and Silhouette selected 2 clusters as the optimal cluster in the Average Linkage grouping method based on indicators that form economic development.

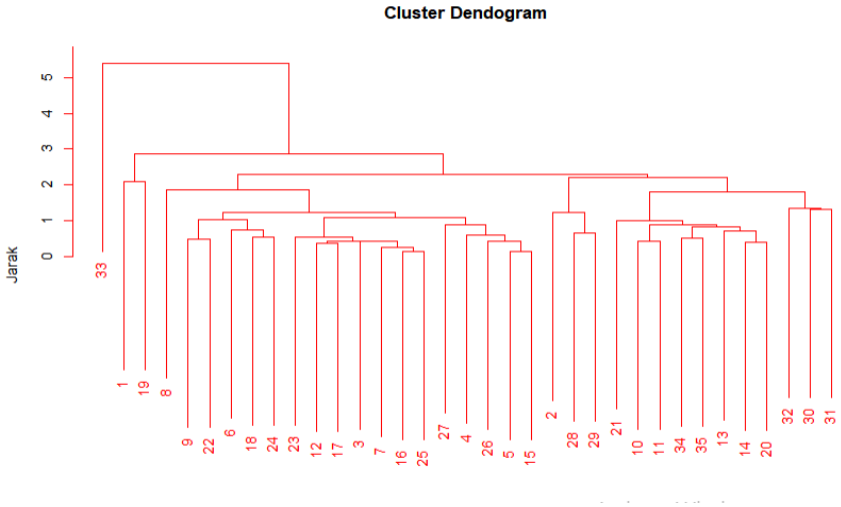

Figure 1. Dendogram of the division of district/city areas in Central Java

Cluster 1 on the dendrogram consists of 22 regencies/cities, namely Cilacap Regency, Magelang Regency, Salatiga City, Kudus Regency, Pemalang Regency, Demak Regency, Klaten Regency, Sukoharjo Regency, Karanganyar Regency, Boyolali Regency, Semarang Regency, Purworejo Regency, Pati Regency, Kendal Regency, Temanggung Regency, Wonogiri Regency, Purbalingga Regency, Wonosobo Regency, Blora Regency, Batang Regency, Banjarnegara Regency, Pekalongan Regency, Kebumen Regency, Grobogan Regency, Banyuman Regency, Tegal Regency, and Brebes Regency.

Cluster 2 on the dendrogram consists of 8 regencies/cities, namely Semarang City, Magelang City, Surakarta City, Sragen Regency, Jepara Regency, Pekalongan City, Tegal City, and Rembang Regency..

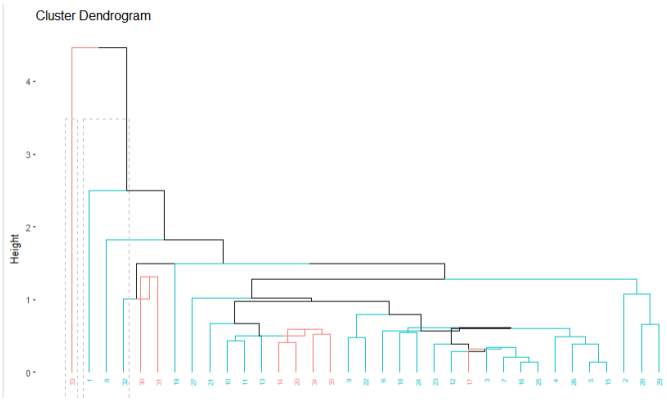

Figure 2. Dendogram of division of districts/cities according to clusters

The characteristics of the GRDP, HDI, and TPAK variables in Cluster 1 are low. Compared to Cluster 1, the characteristics of the GRDP, HDI, and TPAK variables in Cluster 2 are high.

Based on this, it can be concluded that Cluster 1 has low economic development indicators, and Cluster 2 has high indicators. The government or related agencies can use this information to make efforts to improve economic development indicators in Regencies/Cities in Central Java, especially those included in cluster 1, to improve the quality of human resources in Central Java province.

CONCLUSION

Cluster 1 on the dendrogram consists of 22 regencies/cities, namely Cilacap Regency, Magelang Regency, Salatiga City, Kudus Regency, Pemalang Regency, Demak Regency, Klaten Regency, Sukoharjo Regency, Karanganyar Regency, Boyolali Regency, Semarang Regency, Purworejo Regency, Pati Regency, Kendal Regency, Temanggung Regency, Wonogiri Regency, Purbalingga Regency, Wonosobo Regency, Blora Regency, Batang Regency, Banjarnegara Regency, Pekalongan Regency, Kebumen Regency, Grobogan Regency, Banyuman Regency, Tegal Regency, and Brebes Regency. Cluster 2 on the dendrogram consists of 8 regencies/cities, namely Semarang City, Magelang City, Surakarta City, Sragen Regency, Jepara Regency, Pekalongan City, Tegal City, and Rembang Regency.

REFERENCES

Anggraini, L., & Rismawati Arum, P. 2022. Analisis Cluster Menggunakan Algoritma K-Means Pada Provinsi Sumatera Barat Berdasarkan Indeks Pembangunan Manusia Tahun 2021.

Masruri. 2016. Analisis Pengaruh Pertumbuhan Ekonomi,Ipm,Tpakdanpengangguran Terbukaterhadapketimpangan Pendapatan Antardaerahdiprovinsi Jawatengah Tahun 2011-2014.

Rachmat Fuady. 2015. Kajian Teori Dan Implementasi Pembangunanterhadaptolokukur Keberhasilan Pembangunan.

Riska Ramadani dan Admi Salma. 2022. Metode Average Linkage dan Warddalam Pengelompokan Kesejahteran Sumatera Barat Tahun 2021.

Ayub Tanudi. 2023. Buku Ajar Perekonomian Indonesia. (2023). www.penerbitwidina.com

Ivandora Agnestisia dan Nisa Masruroh. 2017. Clustering Kondisi Laboratorium Berdasarkan Penilaian Mahasiswa Dengan Metode Ward Di Jurusan Teknik Industri Upn "Veteran" Jawa Timur.

Nofiantika Rosulya Putri, C., Kurniawati, E., Manfaati Nur, I., & Wasono, R. 2021. Pengelompokkan dan Pemetaan Wilayah Kecamatan di Kabupaten Kendal Berdasarkan Potensi Sektor Pertanian dengan Pendekatan Complete Linkage Hierarchical Clustering.

Hartono, R., Busari, A., dan Awaluddin, M. 2018. Pengaruh Produk Domestik Regional Bruto (PDRB) Dan Upah Minimum Kota (UMK) Terhadap Penyerapan Tenaga Kerja. *INOVASI*, *14*(1), 36–43. http://journal.feb.unmul.ac.id/index.php/INOVASI

Haumahu, G., dan Nanlohy, Y. W. A. 2021. Penerapan Analisis Klaster Hierarki Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Konsumsi Kalori Penduduk. *VARIANCE: Journal of Statistics and Its Applications*, *2*(2), 75–79. https://doi.org/10.30598/variancevol2iss2page75-79

Nurissaidah Ulinnuh, Rafika Veriani. 2020. Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward. *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan. (n.d.). https://doi.org/10.30743/infotekjar.v5i1.2464*

Matdoan, M. Y., dan Noya Van Delsen, M. S. 2020. Penerapan Analisis Cluster Dengan Metode Hierarki Untuk Klasifikasi Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia. *STATMAT (Jurnal Statistika dan Matematika)*

Nafisah, Q., & Chandra, N. E. 2017. Analisis Cluster Average Linkage Berdasarkan Faktor-Faktor Kemiskinan di Provinsi Jawa Timur.

Solikin M. Juhro. 2022. Bank Indonesia Dalam Dinamika Perekonomian Jawa Tengah Semarang Sebagai Simpul Ekonomi.

Arizka Sofiyana Maharani. (2017). Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Partisipasi Angkatan Kerja Di Pulau Kalimantan.

Mohamad Teja. 2015. Untuk Kesejahteraan Masyarakat Di Kawasan Pesisir.

Widodo, E., Mashita, S. N., dan Prasetyowati, Y. G. (2020). Perbandingan Metode Average Linkage, Complete Linkage, dan Ward'S pada Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia. *Faktor Exacta*, *13*(2), 81. https://doi.org/10.30998/faktorexacta.v13i2.6581

Hanandya, S., Nur A, M., Nur, I. M., & Haris, A. (2021.). Perbandingan Metode Single Linkagedan Average LinkagePada Pengelompokan Wilayah Kecamatan di Kabupaten Kendal Berdasarkan Potensi Sektor Pertanian.