

## ANALYSIS OF DYNAMIC TIME WARPING IN THE DEVELOPMENT OF GROSS REGIONAL DOMESTIC PRODUCT YOGYAKARTA

Inggrid Septa Narendra<sup>1</sup>, Muhammad Muhajir<sup>2\*</sup>

<sup>1,2</sup>Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Islam Indonesia  
Yogyakarta, Indonesia  
<https://www.uii.ac.id/>

18611117@students.uii.ac.id<sup>1</sup>, mmuhajir@uii.ac.id<sup>2\*</sup>

(\*) Corresponding Author

### Abstract

Poverty in Indonesia has become a common thing that is still difficult to handle due to the presence of the covid virus outbreak attacks are causing the inability to buy and sell transactions, export and import goods and services, then the level of inequality increases. The tool measures the inequality level in an area seen from the Gini Ratio value. The Gini Ratio notes that the DI Yogyakarta province had the highest index value in Indonesia of 0,437 in September 2020. So this study aims to minimize the inequality in the DI Yogyakarta province by using the clustering method and Dynamic Location Quotient (DLQ). The clustering method with a hierarchical algorithm using the Dynamic Time Warping (DTW) distance and the DLQ method to predict regional economic sectors. Based on the result of the clustering analysis, there were 2 clusters, and the DLQ analysis obtained as many as 11 essential and 6 NPN-base sectors. Cluster 1 has 10 GRDP sectors with two industries that will become non-base sectors in the future, while cluster 2 has 7 GRDP sectors with three sectors it will become base sectors in the future.

Keywords: Clustering Hierarki; Dynamic Time Warping; Dynamic Location Quotient

### Abstrak

*Kemiskinan yang terjadi di Indonesia sudah mejadi hal yang biasa yang mana hal tersebut masih sulit untuk diretas ditambah dengan adanya virus wabah covis yang menyerang mengakibatkan akses dalam jual dan beli, ekspor dan impor produk braang dan jasa menjadi terhambat sehingga tingkat kesenjangan meningkat. Alat untuk mengukur tingkat kesenjangan disuatu wilayah dapat dilihat dari nilai gini ratio, gini ratio mencatat bahwa provinsi DI Yogyakarta memiliki nilai indeks tertinggi di Indonesia sebesar 0,437 pada September 2020. Sehingga penelitian ini bertujuan untuk meminimalisir ketimpangan yang terjadi di Provinsi DI Yogyakarta dengan metode clustering dan Dynamic Location Quotient (DLQ). Metode clustering dengan algoritma hirarki menggunakan jarak Dynamic Time Warping (DTW) dan metode DLQ adalah metode untuk memprediksi sector ekonomi disuatu wilayah. Berdasarkan hasil analisis clustering didapatkan 2 cluster dan hasil dari analisis DLQ didapatkan 11 sektor basis dan 6 sektor non basis. Cluster 1 terdapat 10 sektor PDRB dengan 2 sektor didalamnya akan menjadi sector non basis dimasa mendatang, sedangkan cluster 2 terdapat 7 sektor PDRB dengan 3 sektor didalamnya akan menjadi sector basis dimasa mendatang.*

*Kata kunci: Clustering Hierarki; Dynamic Time Warping; Dynamic Location Quotient*

### INTRODUCTION

Poverty is a critical problem in every country, especially in developing countries like Indonesia. Indonesia's situation is vital, and there is no poverty treatment yet. The BPS website explained that the number of poor people in Indonesia in March 2020 was 26,42 million or 9.78%, an increase of 0.56% or 1.63 million people in September 2019 and an increase of 1.28 million

or 0.37% against March 2019. Poverty appears due to income inequality and unequal income distribution. Allowing the problem to drag on will further complicate the situation and often lead to negative consequences for social and political conditions (Kurniawan, 2009). Before 2020, a fresh outbreak of the covid virus affected almost the globe, including Indonesia. This issue severely impacts the health, social and economic fields (Majidah & Paramartha, 2021). This Pandemic worsens



economic conditions, which causes a decrease in demand for goods and services due to restrictions on community activities (Majidah & Paramartha, 2021), even to the closure of production sites, places of buying and selling, and even recreation. The Pandemic has increased the number of poor people and income inequality (Irawan & Sulisty, 2022).

Income inequality is the disparity between the relative income of high and low-income citizens. The impact of influence affects development, especially economic growth. At least in the form of reduction, burden, and misery eliminated by economic growth.

On the other hand, income inequality will lead to economic inefficiency, social weakness, and solidarity. To measure income inequality among the population, used based on the size distribution of income. However, because income data is challenging to obtain, the measurement of inequality of income distribution is approached using expenditure data. Four measures reflect income distribution inequality: the Gini Coefficient (Gini Ratio), World Bank size, Theil Index, and L-Index.

Table 1. Gini Ratio for Indonesia from September 2017 to September 2020

Year	Rural+Urban (%)	
	Semester 1 (March)	Semester 2 (September)
2017	-	0,391
2018	0,389	0,384
2019	0,382	0,380
2020	0,381	0,385

Source : Berita Resmi Statistik (Badan Pusat Statistik, 2021a)

Table 1 shows that the Gini ratio in Indonesia from 2017 to 2019 experienced a significant yearly decline. Then based on the BPS report, the economic inequality rate as measured by the Gini Ratio increased by 0.385 in September 2020, an increase of 0.004 points when compared to March 2020 of 0.381 and an increase to September 2019 of 0.380.

Table 2. Gini Ratio by Province and Region for Indonesia in 2020

Province	Rural+Urban (%)	
	Semester 1 (March)	Semester 2 (September)
Aceh	0,323	0,319
North Sumatra	0,316	0,314
West Sumatra	0,305	0,301
Riau	0,329	0,321
Jambi	0,320	0,316
South Sumatra	0,339	0,338
Bengkulu	0,334	0,323

Province	Rural+Urban (%)	
	Semester 1 (March)	Semester 2 (September)
Lampung	0,327	0,320
Kep. Bangka Belitung	0,262	0,257
Riau Islands	0,339	0,334
DKI Jakarta	0,339	0,400
West Java	0,403	0,398
Central Java	0,362	0,359
DI Yogyakarta	0,434	0,437
East Java	0,366	0,364
Banten	0,363	0,365
Bali	0,369	0,369
West Nusa Tenggara	0,376	0,386
East Nusa Tenggara	0,354	0,356
West Kalimantan	0,317	0,325
Central Kalimantan	0,329	0,320
South Kalimantan	0,332	0,351
East Kalimantan	0,328	0,335
North Kalimantan	0,292	0,300
North Sulawesi	0,370	0,368
Central Sulawesi	0,326	0,321
South Sulawesi	0,389	0,382
Southeast Sulawesi	0,389	0,388
Sulawesi	0,364	0,356
Gorontalo	0,408	0,406
West Sulawesi	0,318	0,326
Maluku	0,308	0,290
North Maluku	0,308	0,290
West Papua	0,382	0,376
Papua	0,392	0,395

Source : Badan Pusat Statistik Pusat 2020 (Badan Pusat Statistik, 2021b)

Based on table 2, the Gini Ratio by the province in Indonesia in 2020, also known as the Gini Ratio province, has increased from the previous year. The highest Gini Ratio was recorded in the DI Yogyakarta province at 0.434 and 0.437 in March 2020 and September 2020. One indicator to see the economic condition in a region or region is by looking at the Gross Regional Domestic Product (GRDP) value. GRDP is the total value added of final goods and services produced by all economic units in an area. The GRDP sector contributes to developing a regional economy with 17 business sectors. Characteristics of different sectors can affect the growth of GRDP in the province of DI Yogyakarta so that the features of other sectors identified by performing cluster analysis, as said earlier, that GRDP represents the economic condition of a region.

Therefore it is necessary to research to determine the development of the next sector or to encourage economic growth by using DLQ analysis, which is an analysis to predict the development of the financial sector in the future. The existence of cluster analysis and DLQ analysis may benefit from a government strategy based on the clusters' features and the economic sector's future projections.

### The Significance of The Research

The benefits of this research are to provide an overview of how well the economic sectors are in terms of their performance and provide recommendations on which sectors have high and low potential so that they can determine steps to improve sector performance in order to improve the economy in the province of DI Yogyakarta.

### Research Contribution

This research can contribute to the provincial government of DI Yogyakarta as advice, consideration, or reference in dealing with the performance of leader sectors to be developed as well as seeing the prospective leading sectors to develop in the future in order to reduce income inequality in the Province of DI Yogyakarta and find out how the characteristics of every each GRDP sector by business field in DI Yogyakarta Province.

## RESEARCH METHODS

### Types of research

This study uses quantitative data form of figures from the GRDP of the DI Yogyakarta province and the National GDP appropriated from the website of the Central Statistics Agency.

### Research Target / Subject

The population is the entire object under study, so the population of this research is data on all Gross Regional Domestic Product (GRDP) DI Yogyakarta Province and National Gross Domestic Product. Therefore, while the sample is part of the number and characteristics owned by the population, the sample used in the population is the Gross Regional Domestic Product (GRDP) of DI Yogyakarta Province and the National Gross Domestic Product from 2015 to 2020.

### Data, Instruments, and Data Collection Techniques

This study uses quantitative data in the form of numbers. This secondary data was obtained from the Badan Pusat Statistik DI Yogyakarta Province website and Badan Pusat Statistik

Nasional, assisted by the tidying process using Microsoft Excel software.

### Data analysis technique

This research uses Microsoft Excel 2016 software and R-Studio software. This research employs several methods, including descriptive analysis to determine the general description of each variable, DLQ to predict future competent GRDP sectors, clustering to group objects with different properties so that members of the same group share similar characteristics, and dynamic time warping for a path that generates the shortest distance between two-time s.

### Dynamic Location Quotient

Dynamic Location Quotient analysis is a modification of Location Quotient (L.Q.) analysis by accommodating the magnitude of GRDP from sector or sub-sector production values from time to time because L.Q. analysis has a weakness, namely static criteria that only provide an overview at one point in time. Ts means that the sector that has excelled this year will not necessarily be superior in the coming year. On the contrary, the sector that has not excelled this year will excel in the future. So, as an alternative, Dynamic Location Quotient (DLQ) analysis is used, which can predict which sectors will be competent in the future by using the following formula (Nazipawati, 2007):

$$DLQ = \left( \frac{(1+g_{ik})/(1+g_k)}{(1+g_{tp})/(1+g_p)} \right)^t \dots\dots\dots (1)$$

From the results of the DLQ calculation, it can be categorized into three, namely (Simamora & Kifli, 2017):

1.  $DLQ > 1$  means the potential for sector i development in the DI Yogyakarta province is faster than the potential of sector i at the national level. However, the sector can still be expected to become the primary sector
2.  $DLQ = 1$  means the potential development of sector i in the province of DI Yogyakarta proportional to the development potential of sector i at the national level.
3.  $DLQ < 1$  means the potential development of sector i in the province of DI Yogyakarta is lower than the potential development of sector i at the level of, and the sector cannot be expected to be the base sector in the future.

### Cluster Assumption Test

In the cluster assumption, there are two assumptions, according to Hair et al., namely (Hair et al., 2006):



1. Representative Sample, take sample taken can represent or represent the existing population. There is no provision for a representative sample size. However, a large enough sample size is required to carry out the clustering process correctly. Testing a representative sample can be done with the Kaiser-Mayer-Olkin test. Kaiser-Mayer-Olkin (KMO) is widely used to see the adequacy of a sample. This KMO test measures the adequacy of overall sampling and sampling adequacy for each indicator. The KMO test has a value of 0 to 1. If the KMO ranges from 0.5 to 1, the sample can be said to represent the population or a representative sample (Ningrat et al., 2016).
2. Multicollinearity is an almost perfect linear relationship between some or all of the independent variables (Gujarati & Porter, 2015). In cluster analysis, the variables involved must be free from multicollinearity problems. One way to detect multicollinearity problems is to use a correlation matrix (R. E. Sihombing et al., 2019). If the correlation coefficient between variables is from 0.8 to 1.0 (0.8 r 1.0), multicollinearity occurs (Prihastuti, 2015. The correlation coefficient formula is as follows (Hidayat, 2012):

$$r_{xy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{\{n \sum X^2 - (\sum X)^2\} \{n \sum Y^2 - (\sum Y)^2\}}} \dots \dots \dots (2)$$

**Principal Component Analysis (PCA)**

PCA is a data reduction technique multivariate (a lot of data) that seeks to transform (transform) an initial or original data matrix into a set of more linear combinations small. Still, it absorbs some of the amounts of variance from the initial data. The goal of PCA is to explain as much as possible the amount of variance in the original data with as little, and perhaps the primary component called the factor (Nafisah & Chandra, 2017). PCA will form a new set of dimensions, then be ranked based on data variants. Finally, PCA will generate the Principal Component obtained from the eigenvalue and eigenvector decomposition of the covariance matrix. The PCA algorithm, in general, is as follows (Hediyati & Suartana, 2021):

1. Calculate the mean ( $\bar{X}$ ) of the data on each dimension.  

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \dots \dots \dots (3)$$
2. Calculate the covariance matrix using the equation  

$$C_x = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T \dots \dots \dots (4)$$
3. Calculate the covariance matrix eigenvector ( $vm$ ) and eigenvalue ( $\lambda m$ ).

4. Sort the eigenvalues in descending order. The principal component (P.C.) is a series of eigenvectors according to the order of eigenvalues at stage 3.
5. Generate a new dataset.

Three ways to determine the number of principal components used for further analysis are based on the value of eigenvalues, percentage variance, and scree plots. The first is by looking at the eigenvalues, which show the number of variances related to a factor. Factors with eigenvalues greater than one will be retained. The second is based on the presentation of variance. The third with scree plot, showing the relationship between factors and eigenvalues (Nafisah & Chandra, 2017). The number of factors taken is determined based on the cumulative amount of variance achieved, which is usually the first order of 80%.

**Cluster Validity Test**

The cluster validity test is used to see the goodness or quality of the cluster analysis results. The cophenetic correlation coefficient is used to test the validity of the clustering result in this study. The cophenetic correlation coefficient is the correlation coefficient between the original elements of the dissimilarity matrix (Euclidean distance matrix) and the elements generated by the dendrogram (cophenetic atrix) (Dias & Santos, 2013). The formula for calculating the ciphonetic correlation coefficient is as follows (Fadliana & Rozi, 2015):

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{cik} - \bar{d}_c)}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})^2][\sum_{i < k} (d_{cik} - \bar{d}_c)^2]}} \dots \dots \dots (6)$$

The value ranges between -1 and 1; a value close to 1 means that the solution resulting from the clustering process is quite good.

**Silhouette Method**

The silhouette coefficient is one measure of accuracy that can be used in determining the accuracy if time series grouping. In addition, the silhouette coefficient is used to determine the grouping quality. With the following formula (Ayundari & Sutikno, 2019):

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \dots \dots \dots (7)$$

**Cluster Analysis Hierarchical Method**

Clustering is the process of making groupings so that all members of each partition have similarities based on a particular matrix. The hierarchical Method is a structured and gradual



grouping method based on the similarity of properties between objects. In this grouping analysis, the number of clusters is unknown or has not been determined. Generally, the hierarchical Method has two ways of grouping it: merging (agglomerative) and separating (devising) (Alwi, 2018). There are several methods for cluster analysis using agglomerative hierarchical methods based on Linkage as follows (Widodo et al., 2018):

1. A single Linkage Cluster is an agglomerative grouping procedure based on the minimum or closest distance between objects.
2. Complete Linkage is the same as single Linkage, but by grouping objects with the furthest distance or few similarities.
3. Average Linkage is a grouping formed based on the average distance value of all individuals in one group with the average distance of all individuals in other groups.
4. Ward's Method is the distance between two groups. The ward's Method is the sum of the squares between the two groups for all variables. This Method tries to minimize variables within the group and tends to be used to combine groups with small numbers.
5. The Centroid Method defines the similarity between clusters and the distance between the two existing cluster centroids, and the centroid is the average distance in a cluster obtained by averaging all members of a particular cluster. With this Method, every time a new cluster occurs, the centroid will be recalculated until a fixed cluster is formed.

### Dynamic Time Warping

Dynamic Time Warping (DTW) uses dynamic programming techniques to find all possible paths and select them to produce the minimum distance between two-time series using a distance matrix where each element in the matrix is the cumulative distance of the minimum value from three surrounding neighbors. When given two-time series data  $Q = q_1, q_2, q_3, \dots, q_m$  where with size  $m$  and  $C = c_1, c_2, c_3, \dots, c_n$  With size  $n$ , the matrix is formed. Each element  $i, j$  is the cumulative distance of the distance ( $i$ ) and the minimum value of three parts adjacent to the component  $i, j$ , where  $0 < i$  and  $0 < j \leq m$  so that the component ( $i, j$ ) can be defined as follow (Ayundari & Sutikno, 2019):

$$e_{ij} = d_{ij} + \min\{e_{(i-j)(i-j)}, e_{(i-1)j}, e_{i(i-1)}\} \dots \dots \dots (8)$$

The calculation of the value can be written as follows:

$$d_{ij} = (q_i - c_j)^2 \dots \dots \dots (9)$$

The smallest cumulative distance or FTW distance on  $(m, n)$  is defined as follows:

$$d_{DTW}(Q, C) = \min_{\forall W \in P} \left\{ \sqrt{\sum_{k=1}^K d_{wk}} \right\} \dots \dots \dots (10)$$

## RESULTS AND DISCUSSION

### Descriptive Analysis

Descriptive analysis determines the description or distribution of sample or population data (Sugiyono, 2006). So this descriptive analysis assesses the condition of each GRDP sector according to business fields in the DI Yogyakarta Province from 2015 to 2020.

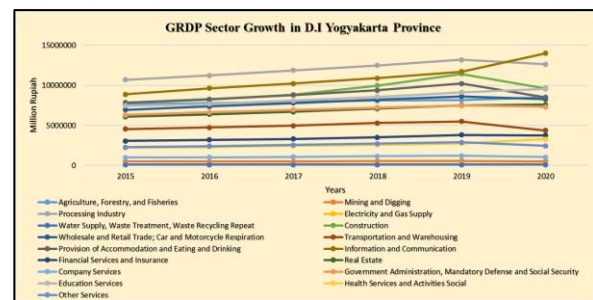


Figure 1. Visualization of DI Province's GRDP Sector Growth Yogyakarta

Figure 1 demonstrates that the growth of the GRDP in the information and communication sector experienced a significant increase in 2020 as a result of the virus that affected all activities; consequently, the internet played a crucial role in enabling people to remain connected while engaging in online activities, such as work-related matters, online education, and online shopping (cashless). On the other hand, GRDP sectors, such as providing food and drink accommodation, transportation, warehousing, and construction, experienced a deep decline resulting in increased unemployment.

### Dynamic Location Quotient Analysis

Dynamic Location Quotient (DLQ) Analysis from this study predicts competent sectors in the future to reduce economic inequality (F. N. Sihombing, 2018) in the DI Yogyakarta Province by looking at the DLQ value of each industry. If the DLQ value  $> 1$ , then sector  $i$  in the DI Yogyakarta province is growing faster than the sector at the national level and will become the base sector in the future, whereas if the DLQ value is  $< 1$ , sector  $i$  in the DI Yogyakarta province is lower (slower) developing than the national level and cannot become the base sector in the future (Simamora & Kifli, 2017)



Table 3. Results of DLQ Analysis of the GRDP Sector by Business Fields of DIY Province 2015-2020

No.	GRDP Sector by Business Field	Average GRDP Growth by DIY Business Field in 2015-2020	Dynamic Location Quotient Results
1	Agriculture, Forestry, and Fisheries	2,145	0,329078
2	Mining and digging	1,455	213,2543
3	Processing industry	3,233	2,279655
4	Electricity and Gas Supply	4,628	18,54063
5	Water Supply, Waste Treatment, Waste Recycling Repeat	3,877	0,432079
6	Construction	4,752	2,166185
7	Wholesale and Retail Trade; Car and Motorcycle Respiration	4,027	6,861865
8	Transportation and Warehousing	0,505	0,007148
9	Provision of Accommodation and Eating and Drinking	2,707	1,763402
10	Information and Communication	8,867	1,456642
11	Financial Services and Insurance	4,94	0,673607
12	Real Estate	4,805	3,661204
13	Company Services	2,348	0,03865
14	Government Administration, Mandatory Defense, and Social Security	3,465	1,521023
15	Education Services	5,487	2,857265
16	Health Services and Activities Social	8,225	2,326564
17	Other Services	2,693	0,043584

Table 3 shows the results of the DLQ analysis coefficient values of the GRDP sectors according to the business fields of the DI Yogyakarta province or the economic sector in the DI Yogyakarta province. The results of the DLQ analysis show that there are 11 sectors with a value  $> 1$ , namely the mining and quarrying sector; processing industry sector; electricity and gas procurement sector; construction sector; wholesale and retail trade, car and motorcycle respiration; the sector of providing accommodation and food and drink; information and communication sector; real estate sector; government administration, defense, and mandatory social security sectors; education services sector; and the health services sector and social activities. Sectors with a value of  $> 1$  are included in the base sector and provide more income for the region because the base sector does not only meet the needs of the region's goods and services, but goods and services from the base sector are exported outside the region. The ability of these deposited sectors to become the primary sector is due to the presence of natural resources (SDA), which accommodate abundant resources such as soil fertility, natural wealth, minerals, climate, and water resources. In addition to good natural resources, human resources are a factor in these sectors becoming the leading sector or the basis because sufficient and intelligent human resources will encourage the performance of these sectors. The area of land or land is also a forum for producing goods and services from critical sectors.

Meanwhile, for the other six sectors, the coefficient of DLQ analysis is  $< 1$ , namely the agriculture, forestry, and fishery sectors; water supply, waste management, waste and recycling sectors; transportation and warehousing sector; financial services and insurance sector; corporate service sector; and other service sectors. Sectors with a value of  $< 1$  are included in the non-base sector, where the sector can only meet its region's goods and services needs.

#### Multicollinearity Test

The multicollinearity test shows a linear correlation or relationship between some or all independent variables. In cluster analysis, each variable must be free from multicollinearity. One of the multicollinearity tests is to use the correlation coefficient ( $r$ ). If the correlation coefficient ( $r$ ) is  $> 0.8$ , then multicollinearity occurs. Otherwise, if the correlation coefficient ( $r$ ) is  $< 0.8$ , then multicollinearity is free.

Table 4. Results of DLQ Analysis of the GRDP Sector by Business Fields of DI Y Province 2015-2020

	X1	X2
X1	10000000	0,9995417
X2	0,9995417	1,0000000
X3	0,9988504	0,9997580
X4	0,9966838	0,9980541
X5	0,9911028	0,9931214
X6	0,9767921	0,9803359

	X3	X4
X1	0,9988504	0,9966838
X2	0,9997580	0,9980541
X3	1,0000000	0,9990074
X4	0,9990074	1,0000000
X5	0,9950122	0,9984134
X6	0,9811782	0,9805875

	X5	X6
X1	0,9911028	0,9767921
X2	0,9931214	0,9803359
X3	0,9950122	0,9811782
X4	0,9984134	0,9805875
X5	1,0000000	0,9777661
X6	0,9777661	1,0000000

Table 4 shows the correlation matrix output for each variable. It seems that the GRDP in Yogyakarta in 2015 ( $X_1$ ), the GRDP in Yogyakarta in 2016 ( $X_2$ ), in the GRDP in Yogyakarta in 2017 ( $X_3$ ), in the GRDP in Yogyakarta in 2018 ( $X_4$ ), in the GRDP in Yogyakarta in 2019 ( $X_5$ ), and the GRDP variable in Yogyakarta in 2020 ( $X_6$ ) is  $> 0.8$ , meaning that these variables contain multicollinearity.

### Principal Component Analysis

When multicollinearity occurs, there are several ways to overcome it: Principal Component Analysis (PCA). PCA is a multivariate data reduction technique that seeks to convert an initial or original data matrix into a set of fewer linear combinations without compromising the available information. PCA reduces  $n$  variables into  $k$  new variables or principal components called factors. In clustering, there are several ways to determine the main components, namely by looking at the eigenvalue (eigenvalue  $> 1$ ) or the total variance explained by more than 80%.

Table 5. Importance of Component

PCA	PC1	PC2
Standard Deviation	9,639	7,687
Proportion of Variance	0,9917	0,00631
Cumulative Proportion	0,9917	0,9980

PCA	PC3	PC4
Standard Deviation	4,263	8,250
Proportion of Variance	0,1940	0,00007
Cumulative Proportion	0,9999	1

PCA	PC5	PC6
Standard Deviation	2,369	2,0299
Proportion of Variance	0,00001	0
Cumulative Proportion	1	1

Table 5 shows the importance of components. There are three essential components: standard deviation, variance proportion, and cumulative proportion. To determine the principal component, we can use the cumulative proportion. The formation of components is said to be good if the cumulative proportion is high. Table 5 shows that the variance of the six variables can be explained by the first component (PC1) of 9.917e x or 0.9917 or 99.17%, meaning that the formation of the main component is suitable for further analysis because it has a cumulative proportion value above 80%.

### Cluster Validity Test

In the hierarchical cluster analysis, there are five agglomerative hierarchical methods: single Linkage, complete Linkage, average Linkage, wards, and centroid. In addition, the cluster validity test can be used by looking at the cophenetic correlation coefficient. The value of the cophenetic correlation coefficient ranges from -1 to 1. Therefore, if the value of the cophenetic correlation coefficient is close to 1, the solution resulting from the clustering process is quite good.

Table 6. Cophonetic Correlation Coefficient

Method	Cophonetic Correlation Coefficient
Single Linkage	0,740803
Complete Linkage	0,8034567
Average Linkage	0,8136157
Ward's	0,7797824
Centroid	0,8097512

Table 6 explains that the cluster validity test with the cophenetic correlation coefficient shows that the average Method provides the best solution of the other four methods because it has an enormous cophenetic correlation coefficient value of 0.8136157.

### Silhouette Method

Before looking for cluster analysis, it would be nice to determine the optimal number of clusters. Forming the optimal number of clusters fulfils several ways, one of which is the silhouette method. The silhouette method combines separation and cohesion methods that are useful for seeing the quality and strength of clusters, whether objects are placed in a cluster or not.

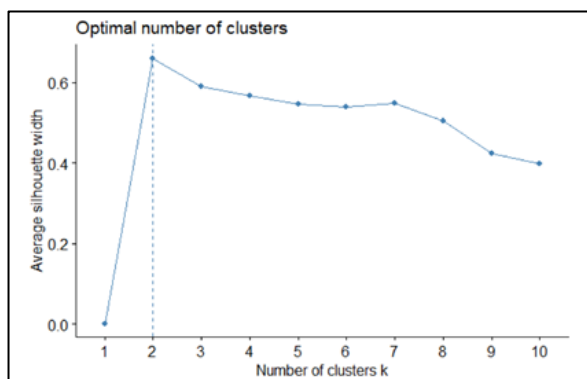


Figure 2. Output Silhouette Method

Figure 2 shows the output of the formation of the number of clusters using the silhouette method and recommending the optimal  $k$  or determining the optimal number of clusters when  $k = 2$ , as seen from the vertical line on the x-axis when  $k = 2$ .

### Hierarchical Cluster Analysis

Hierarchical cluster analysis is a grouping method in which the number of groups to be created is unknown. Data clustering is done by measuring each object's proximity, forming a dendrogram. In this study, clusters were obtained by sequential merging (agglomerative) techniques, namely the average linkage method. Average Linkage is a hierarchical clustering method based on the distance average of all objects in one cluster with all objects in other clusters. The distance used in this study is the Dynamic Time Warping Distance. The distance in this Method has accurate measurement and is suitable for hierarchical clusters. From the results of cluster analysis using the hierarchical algorithm of the average linkage method, the dendrogram results are obtained as follows:

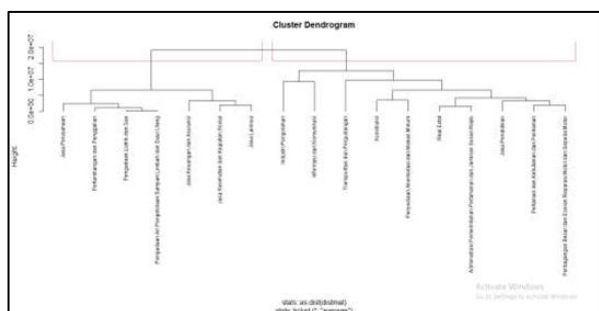


Figure 3. Output Dendrogram Average Linkage

The clustering carried out by this study aims to group the GRDP sectors according to the business fields of the DI Yogyakarta province based on the GRDP in DI Yogyakarta in 2015-2020. The dendrogram results in Figure 3 explain that the

grouping using the average linkage method obtained 2 clusters. Cluster 1 has 10 GRDP sectors, and Cluster 2 has 7 GRDP sectors.

Table 7. Average Linkage Cluster Member

Cluster	DI Yogyakarta Province GRDP Sectors	Number of Sectors
1	Agriculture, Forestry, and Fisheries; Processing industry; Construction; Wholesale and Retail Trade (Car and Motorcycle Respiration); Transportation and Trade; Accommodation and Food and Drink Providers; Information and Communication; Real Estate; Government Administration of Defense and Social Security; Education Services	10
2	Mining and excavation; Electricity and Gas Procurement; Water Procurement for Waste Management and Recycling; Financial Services and Insurance; Health Services and Social Activities; Service Other	7

Each sector included in the cluster has the same characteristics. The sectors incorporated in cluster 1 contribute generously to the economy in the DI Yogyakarta province. Meanwhile, suppose it is associated with the DLQ analysis under the agriculture, forestry, and fishery sectors; the transportation and warehousing sectors for the future need to be considered again in their production processes or performance because these two sectors predict to become non-base sectors in the future. In that case, this is also reinforced by looking at the growth rates of the two sectors, which have decreased significantly due to the Pandemic that hit so the impact will carry over in the future. Meanwhile, the sectors belonging to cluster 2 experienced a profound decline, so the contribution to the economy in the DI Yogyakarta province was not good, but if it is associated with the DLQ analysis, several sectors in cluster 2, such as the mining and quarrying sector, the electricity and gas procurement sector, and the mining and quarrying sector. Health services and social activities will contribute well because they predict to become the primary sector in the future.



## CONCLUSIONS AND SUGGESTIONS

### Conclusion

From the results of the analysis that researchers have done, several conclusions are obtained, namely:

General description of the growth of the GRDP sector in the province of Daerah Istimewa. From 2015-to 2020, Yogyakarta experienced a decline in its economic sectors, such as the transportation and warehousing sector, the food and drink accommodation provider sector, and the construction sector. However, the information and communication sector has become a reliable sector.

Then, based on the results of the calculation of the Dynamic Location Quotient (DLQ) analysis, there are 11 GRDP sectors in Yogyakarta which are predicted to become the base sector in the future, where these sectors can meet the needs of goods and services for their region and even be able to export goods and services to the other areas. These sectors consist of the mining and quarrying sector; processing industry; procurement of electricity and gas; construction; wholesale and retail trade in respiration of automobiles and motorcycles; provision of accommodation and food and drink; information and communication; real estate; mandatory government administration, defence, and social security; education services; health services and social activities.

And then, based on the results of the cluster with the hierarchical algorithm of the average linkage method, it produces 2 clusters where cluster 1 has the characteristics of sectors with good GRDP contributions consisting of the agricultural, forestry, and fishery sectors; processing industry; construction; wholesale and retail trade in respiration of automobiles and motorcycles; transportation and warehousing; provision of accommodation and food and drink; information and communication; real estate; government administration of defense and social security; education services. While cluster 2 has the characteristics of sectors with a slow GRDP contribution due to the decline in the Pandemic consisting of the mining and quarrying sector; electricity supply and gases; water supply, waste management, waste, and recycling; financial and insurance services; company services; health services and social activities; other services.

### Suggestion

We suggest that further research can use other clustering algorithms to produce more optimal clusters and other DLQ methods to predict leading sectors to get the best results. Moreover, this

research can be used as a reference, especially in the economic field by considering the GRDP sector's growth rate.

## REFERENCES

- Alwi, W. (2018). Analisis Kluster untuk Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, 6(1), 35. doi:10.24252/msa.v6i1.4782
- Ayundari, I., & Sutikno. (2019). Penentuan Zona Musim di Mojokerto Menurut Karakteristik Curah Hujan Dengan Metode Time Series Based Clustering. *Inferensi*, 2(1), 63-70. dx.doi.org/10.12962/j27213862.v2i2.6819
- Badan Pusat Statistik. (2021a). *Gini Ratio Maret 2021 tercatat sebesar 0,384*. Badan Pusat Statistik. <https://www.bps.go.id/pressrelease/2021/07/15/1845/gini-ratio-maret-2021-tercatat-sebesar-0-384-.html>
- Badan Pusat Statistik. (2021b). *Gini Ratio Menurut Provinsi dan Daerah 2019-2020*. Badan Pusat Statistik. <https://www.bps.go.id/indicator/23/98/2/gini-ratio-menurut-provinsi-dan-daerah.html>
- Dias, A. R. da S., & Santos, C. T. dos. (2013). A cophenetic correlation coefficient for Tocher's Method. *Pesq. Agropec. Bras., Brasilia*, 48(6), 589-596. doi.org/10.1590/S0100-204X2013000600003
- Fadliana, A., & Rozi, F. (2015). Penerapan Metode Agglomerative Hierarchical Clustering Untuk Klarifikasi Kabupaten/Kota Di Propinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana. *Jurnal Matematika Murni Dan Aplikasi*, 4(1), 35-40. doi.org/10.18860/ca.v4i1.3172
- Gujarati, D. N., & Porter, C. D. (2015). *Dasar-dasar ekonometrika* (Edisi 5). Salemba Empat.
- Hair, J. F., Anderson, R. E., Tatham, R. L., Babin, B. J., & Black, W. C. (2006). *Multivariate Data Analysis Sixth Edition*. Pearson Education, Inc.
- Hediyati, D., & Suartana, I. M. (2021). Penerapan Principal Component Analysis(PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro. *Journal Information Engineering and Educational Technology*, 5(2), 49-54. doi.org/10.26740/jieet.v5n2.p49-54
- Hidayat, A. (2012). *Klasik, Uji Pearson Product Moment dan Asumsi*. Statistikian.Com. <https://www.statistikian.com/2012/07/pearson-dan-asumsi-klasik.html>

- Irawan, A. D., & Sulisty, A. Q. P. (2022). Pengaruh Pandemi Dalam Menciptakan Ketimpangan Sosial Ekonomi Antara Pejabat Negara Dan Masyarakat. *Jurnal Citizenship Virtues*, 2(1), 251–262. doi.org/10.37640/jcv.v2i1.1184
- Kurniawan, D. (2009). *Kemiskinan Di Indonesia dan Solusinya*.
- Majidah, A. M., & Paramartha, D. Y. (2021). Analisis Dampak dan Siklus Pertumbuhan Lapangan Usaha: Kajian Lapangan Usaha pada Masa Pandemi COVID-19. *Seminar Nasional Official Statistics*, 2021(1), 409–418. doi.org/10.34123/semnasoffstat.v2021i1.897
- Nafisah, Q., & Chandra, N. E. (2017). Analisis Cluster Average Linkage Berdasarkan Faktor-Faktor Kemiskinan di Provinsi Jawa Timur. *Zeta – Math Journal*, 3(2), 31–36. doi.org/10.31102/zeta.2017.3.2.31-36
- Nazipawati. (2007). Aplikasi Model Static dan Dynamic Location Quotients dan Shift-Share Dalam Perencanaan Ekonomi Regional (Studi Kasus Kabupaten Ogan Komering Ulu Propinsi Sumatera Selatan). *Jurnal Pembangunan Ekonomi Wilayah*, 2(2), 82–86. doi.org/10.20884/1.erjpe.2007.2.2.393
- Ningrat, D. R., Maruddani, D. A. I., & Wuryandari, T. (2016). Analisis Cluster Dengan Algoritma K-Means dan Fuzzy C-Means Clustering Untuk Pengelompokan Data Obligasi Korporasi. *Jurnal Gaussian*, 5(4), 641–650. doi.org/10.14710/j.gauss.v5i4.14721
- Prihastuti, D. (2015). *Analisis Generalized Two Stages Ridge Regression (GTSRR) untuk Mengatasi Multikolinearitas dan Autokorelasi beserta Aplikasinya*. Universitas Negeri Yogyakarta.
- Sihombing, F. N. (2018). Identifikasi Pangan Unggulan di Kota Medan : Location Quotient dan Dynamic Location Quotient. *Jurnal Pembangunan Perkotaan*, 6(2), 91–94. ejpp.balitbang.pemkomedan.go.id/index.php/JPP/article/view/45%0A
- Sihombing, R. E., Rachmatin, D., & Dahlan, J. A. (2019). Program Aplikasi Bahasa R Untuk Pengelompokan Objek Menggunakan Metode K-Medoids Clustering. *Jurnal EurekaMatika*, 7(1), 58–79. doi.org/10.17509/jem.v7i1.17888
- Simamora, P. B., & Kifli, F. W. M. (2017). Analisis DLQ (Dynamic Location Quotient) Terhadap Sektor Ekonomi Di Propinsi Sumatera Utara. *Jurnal Masepi*, 2(1).
- Sugiyono, P. D. (2006). *Statistik untuk Penelitian*. CV ALFABETA.
- Widodo, E., Sari, N. N., Hidayati, I., Yubinas, F., Yuniarti, M., & Novyantika, R. D. (2018). Analisis Cluster Penderita Disabilitas Mental di Provinsi Daerah Istimewa Yogyakarta Tahun 2016. *Prosiding Konferensi Nasional Penelitian Matematika Dan Pembelajarannya (KNPMP) III 2018*, 2502–6526, 577–586. <http://hdl.handle.net/11617/10140>