PROC. INTERNAT. CONF. SCI. ENGIN.
Volume 4, February 2021
Page 235-239

ISSN 1504607797
E-ISSN 1505707533

Cluster Analysis Using Hierarchic Method for Classification of District / City of North Kalimantan Province Based on Human Development Indicators (HDI)

Muhammad Rizal[1], Shofia Nur Fadhila[2], Prihantini[3]
[1, 2]Department of Mathematics, Faculty of Science and Technology, UIN Sunan Kalijaga
[3]Department of Mathematics, Bandung Institute of Technology.
[12]Jl. Marsda Adisucipto, No. 1 Yogyakarta 55281, Indonesia. Tel. +62-274-540971, Fax. +62-274-519739.
[3]Jl. Ganesha No.10 Bandung West Java 40132, Indonesia. Tel. (022)251503..
Email: [1]muhdrhizal99@gmail.com , [2]sfadhila19@gmail.com , [3]titinprihantini4@gmail.com

Abstract. HDI is an important indicator in measuring the success of the development of the quality of human life in terms of education, health, and life worthiness. HDI in the province of North Kalimantan (Kaltara) has always experienced a significant increase. However, the increase is always below the national figure. This is due to the development that is not implemented evenly and is not on target. Cluster analysis is a multivariate method that has the aim of grouping, where a group has relatively the same characteristics (homogeneous), while between groups has different characteristics. This study showed that the best method in hierarchical cluster analysis is the ward's method. Therefore, there were 4 clusters in the HDI in North Kalimantan province which included HDI with the high category, namely the city of Tarakan, then the HDI with the medium category, namely Malinau district then the HDI with the low category namely Bulungan district, then the HDI with the very low category namely Tanah Tidung district and Nunukan district.

Keywords: HDI, Cluster, Hierarchy, Kaltara, Ward's method.

Abbreviations: HDI (Human Development Indicators), VIF (Variance Inflation Factor).

Running title: Cluster Analysis Using Hierarchic Method.

INTRODUCTION

UNDP in the Human Development Report 1990 mentions that man is the real wealth of the nation. The main goal of development is to create a viable environment for its people to enjoy longevity, be healthy, and live a productive life. Therefore, a human development index (HDI) was established to emphasize that human development should be a criterion in assessing a country's development.

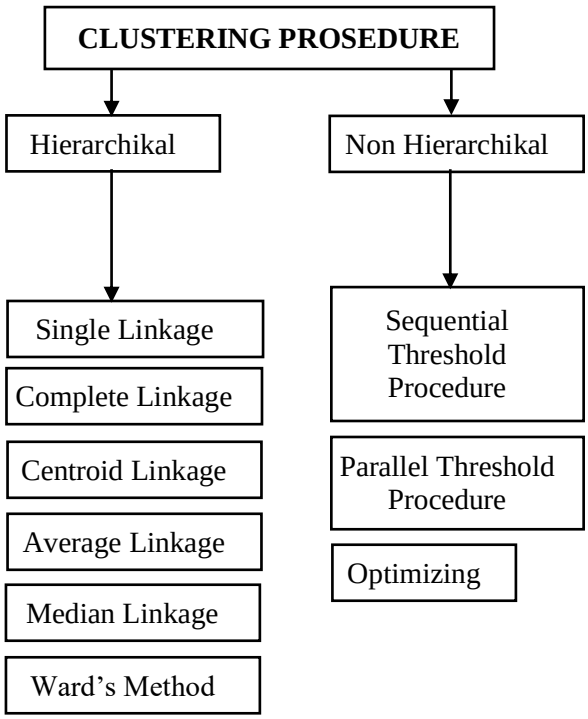
HDI is an important indicator in measuring the success of human quality of life development in terms of education, health, and viability of life. HDI has three elements, namely Health, Education achieved, and living standards or often called economics. The three elements do not stand alone but instead influence each other. It is also influenced by other factors, such as the availability of job opportunities, economic growth, infrastructures, and government policies. So a high HDI indicates the success of economic development in the area.

North Kalimantan province is the 34th province and is considered the youngest province in Indonesia. The trend of the human development index of north Kalimantan province in 2017 was 69.84 then in 2018 was 70.56 and in 2019 at 71.15. It can be said that HDI in North Kalimantan province from year to year has increased (BPS North Kalimantan Province, 2019). But it's always below the national figure. This is because the development carried out by the government is not carried out evenly or in other words not on target. Therefore, one of the solutions that can be applied is to group districts/cities based on HDI indicators, so that the government can take or decide on good policies and strategies and targets in development (Matdon, 2019).

Cluster analysis is a technique for placing several objects into groups based on the similarity of certain characteristics. The advantage of this method in grouping data is that it can group large amounts of observation data and relatively large variables and the data reduced by the group will be easily analyzed. Also, cluster analysis can be used in ordinal data scales, intervals, and ratios. In general, an object is inserted into a cluster or group so it is more likely to relate to other objects in its cluster than to objects from other clusters. Cluster formation is based on the strong err relationship between objects (Lina, 2011).

In general, there are two methods of grouping data in cluster analysis namely hierarchical methods and non-hierarchical methods (Epha Supandi, 2020). Hierarchical cluster analysis has several methods namely single linkage method, Complete Linkage method, Centroid Linkage, Average Linkage method, median method (Median Method), within groups method, and ward method (Ward's Method) while the non-hierarchy method has several methods namely Sequential threshold procedure method, Parallel threshold procedure, and Optimizing.

According to Epha Supanda in Simamora (2005), cluster analysis classification can be displayed in the following chart form.



In this study will classify districts/cities in North Kalimantan province based on the human development index (HDI).

MATERIALS AND METHODS

A. Data Source

The data used in this study is secondary data obtained from north Kalimantan's central statistics agency (BPS) in 2019.

B. Research variables

In this study, the data were grouped based on 5 Districts / Cities in North Kalimantan Province, namely Tarakan District, Malinau District, Bulungan District, Tanah Tidung District, and Nunukan District.

Research Variable Table.

Variable	Description
X_1	Number of poor people
X_2	Number of health facilities and infrastructure
X_3	Number of educational facilities
X_4	Population dependency load

C. Data Analysis Techniques

In this study, data classification used hierarchical cluster analysis with Ward's method and its data processing using IBM SPSS 22 applications. The data analysis steps in this study are as follows.

1. Collect data.
2. Calculate descriptive statistics based on research variables.
3. Testing cluster analysis assumptions,
   a. The sample must represent the population

The test statistics used in this study are Kaiser-Meyer-Olkin (KMO) values.

b. Multicholinearity

To find out if the data indicated a case of multicollinearity or not then it is seen from the value of Tolerance and Variance Inflation Factor (VIF).

4. Choose a cluster analysis procedure

This study used hierarchical cluster analysis using Ward's method. The steps of testing cluster analysis using SPSS software are as follows.

a. Input variable in variable view sheet.
b. Insert data in datasheet view.
c. Click analyze – classify – Hierarchical Cluster.
d. Insert all variables into the variable field.
e. On the "method" menu select ward's method.
f. On the checklist "plots" menu for dendrogram options.
g. Next click continues, and then Ok.
h. Determine the number of clusters.
i. Interpretation.

After getting the results of the cluster analysis, the next step is to interpret the results of the cluster formed based on the data obtained.

RESULTS AND DISCUSSIONS

A. HDI Overview in North Kalimantan Province

Descriptive analysis based on the human development index (HDI) of districts/cities in North Kalimantan province in 2019.

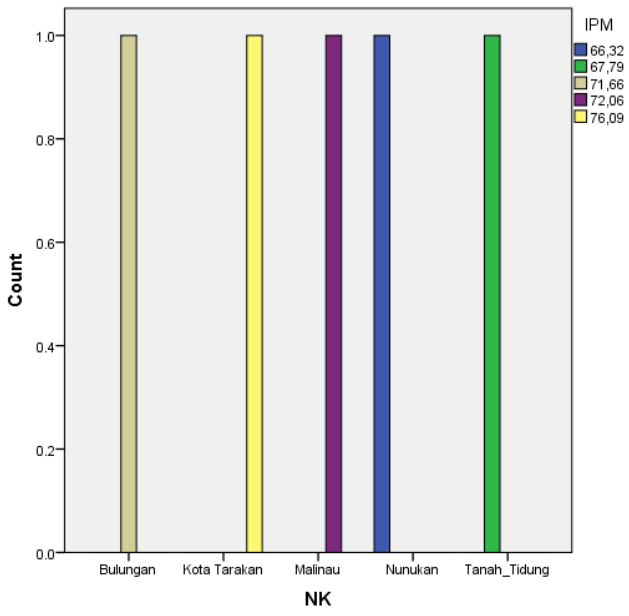


Figure 1. Descriptive HDI Data of North Kalimantan Province.

Based on the above image, it can be seen that the highest HDI in North Kalimantan province is Tarakan City with HDI of 76.09 then Malinau district with HDI 72.06 and the shortest HDI namely Nunukan district with HDI 66.32.

Table 2. Descriptive Statistics Human Development Index Indicator (HDI).

	N	Minimum	Maximum
X1	5	47.20	87.80
X2	5	3.00	12.00
X3	5	6.80	31.20
X4	5	46.24	53.17
Valid N (listwise)	5		

Based on table 2, it can be seen that variable X1 gets a minimum value of 47.20 while the maximum value is 87.80, furthermore, the X2 variable gets a minimum value of 3.00 while for a maximum value of 12.00, the next X3 variable gets a minimum value of 6.80 while for a maximum value of 31.20, the next for variable X4 gets a minimum value of 46.24 while for a maximum value of 53.17.

B. Cluster Analysis Assumptions

According to Hair, et al (1998 and Prayudho (2009) in Princess (2014), there are two assumptions in the cluster analysis, namely:

a. Sample Assumptions representing.

The data used in cluster analysis is representative sample data. The results of the assumption test with the Kaiser-Meyer-Olkin Measure (KMO) are as follows.

Table 3. KMO and Bartlett's Test.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.703
Bartlett's Test of Sphericity	Approx. Chi-Square	2.173
	df	3
	Sig.	.537

Based on table 3, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy value is 0.703. THE KMO value of 0.73 ranges from 0.5 to 1, so it can be concluded that the sample can represent the population and the variables can be used for further analysis.

b. Assumptions of Multicollinearity.

Multicollinearity is a state that indicates the presence of a correlation or strong relationship between two or more independent variables in multiple regression models or other words multicollinearity is a test conducted to determine whether or not there is a correlation between independent variables or free variables. According to Santoso (2000), testing for the absence of Symptoms of Multicholinearity is carried out about the value of variance Inflation factor (VIF) and Tolerance value. If the VIF value is below 10 and the tolerance value is close to 1 then it can be concluded that there is no Multicolineclineity.

To determine whether or not multicollinearity exists in data is to look at the Value of Tolerance and Variance Inflating Factor (VIF).

Table 4. Multicolineity Test.

Model	Collinearity Statistics	
	Tolerance	VIF
1 (Constant)		
X1	,508	1,967
X3	,555	1,803
X4	,620	1,612

Based on table 4 above, it appears that the predictor variables are X1, X3, and X4. The tolerance value for all three predictor variables is more than 0.10 and the VIF value for all three variables is less than 10. It can then be concluded that there is no indication of multicollinearity in all three predictor variables.

Model	Collinearity Statistics		
	Tolerance	VIF	Minimum Tolerance
1 x2	5,465E-005	18299,905	5,465E-005

From the table. It can be seen that the X2 variable is not included in the predictor variable because the data is insignificant. This can be seen from its tolerance and VIF value.

C. Cluster Analysis by Hierarchical Method.

In this study, the analysis used is cluster analysis using the Average Linkage, Single Linkage, Complete Linkage, Centroid Linkage, Median Linkage, and Ward's Linkage methods. Of the methods tested, the best and most appropriate method used to view group composition is Ward's method because it has the least amount of RMSSTD (Root Mean Square Standart Deviation) value index as the determination of the number of selected groups. The RMSSTD index value of the entire method can be seen in the following table.

RMSSTD Index Value Table Group Analysis with 6 Methods

Jumlah Kelompok	Single Linkage	Median Linkage	Complete Linkage	Average Linkage	Ward's	Centroid Linkage
1	35,700	35,700	35,700	35,700	17,850	35,700
2	105,586	129,655	171,574	138,580	104,287	129,655
3	367,481	523,623	790,702	624,839	546,844	590,076
4	823,856	1439,314	2293,922	1226,178	1418,418	1089,467

Based on the table above it can be seen that from the whole test using hierarchical cluster analysis, the number of optimum groups with the smallest RMSSTD index is obtained. The number of groups obtained is 4 groups with an RMSSTD value of 17,850.

The grouping of districts/cities in North Kalimantan province using the ward's method can be seen in the following table.

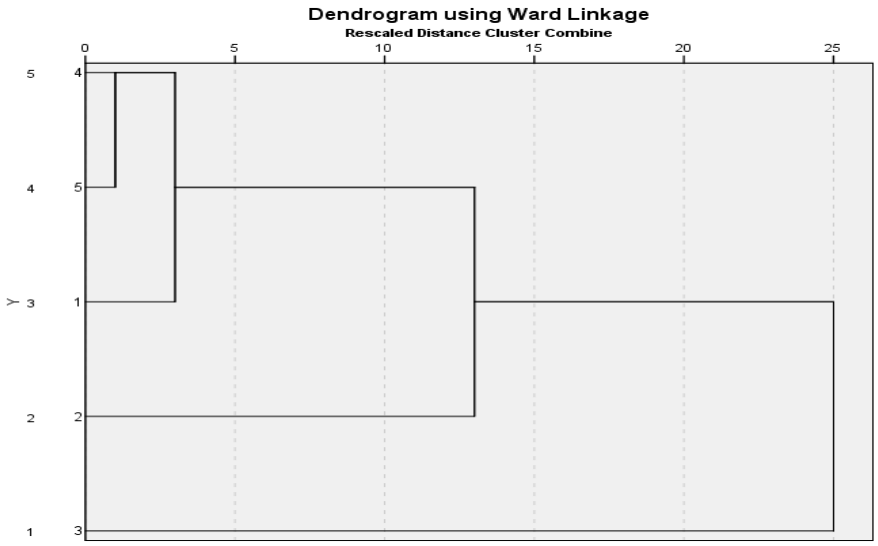
District/City Grouping Table

Case	4 Clusters	3 Clusters	2 Clusters
1	1	1	1
2	2	2	1
3	3	3	2
4	4	1	1
5	4	1	1

Based on the table above, you can see the grouping of districts/cities based on HDI indicators in North Kalimantan Province with case 1 = Malinau District, case 2= Bulungan district, case 3= Tana Tidung District, case 4= Nunukan District, and case 5= Tarakan district are as follows.

1. If 4 groups are formed, then cluster member 1 is Malinau District, cluster member 2 is Bulungan District, Cluster member 3 is Tana Tidung District, while cluster member 4 is Nunukan District and Tarakan District.
2. If 3 groups are formed, then obtained cluster member 1 is Malinau District, Nunukan District, Tarakan District, cluster member 2 is Bulungan District, and cluster member 3 is Tana Tidung District.
3. If 2 groups are formed, it is obtained that the members of cluster 1 are Malinau, Bulungan District, Nunukan District, and Tarakan District, while the members of cluster 2 are Tana Tidung District.

While the grouping of Districts / Cities is based on the HDI Prov indicator. North Kalimantan uses dendrograms as follows:



CONCLUSION

Based on the results and discussion, it can be concluded as follows:

1. The application of cluster analysis with hierarchical methods to classify districts/cities in North Kalimantan province based on indicator human development index (HDI) is carried out with six methods namely Average Linkage, Single Linkage, Complete Linkage, Centroid Linkage, Median Linkage, and Ward's Linkage methods. Obtained the best method is ward's method.
2. Analysis of clusters using hierarchical methods in districts/cities in North Kalimantan province based on the HDI indicator is as follows.
   a. If formed 4 groups, then cluster member 1 is Malinau District, cluster member 2 is Bulungan District, Cluster member 3 is Tana Tidung District, while Cluster 4 member is Nunukan District and Tarakan District.
   b. If formed 3 groups, then obtained cluster member 1 is Malinau District, Nunukan District, Tarakan District, cluster member 2 is Bulungan District, and cluster member 3 is Tana Tidung District.
   c. If formed 2 groups, it is obtained that the members of cluster 1 are Malinau, Bulungan District, Nunukan District, and Tarakan District, while the members of cluster 2 are Tana Tidung District.

REFERENCES

BPS Kalimantan Utara. *Indeks pembangunan manusia kalimantan utara 2019*. www.kaltarabps.go.id.

Matdoan, M.Y., Noya Van Delsen. 2020. Penerapan Analisis Cluster Dengan Metode Hierarki Untuk Klasifikasi Kabupaten/kota di Provinsi Kalimantan Utara Berdasarkan Indeks Pembangunan Manusia. Dalam jurnal statistika matematika, Vol. 2, No.2 : 123-130.

Romesburg, Charles. 2004. Cluster Analysis For Researches. Lulu Press: United States of America.

Safitri, Diah., Widiharih., Tatik. dkk. 2012. Analisis Cluster Pada Kabupaten/Kota Di Jawa Tengah. Berdasarkan Produksi Palawija. Dalam jurnal media statistika, Vol. 5, No. 1:11-16.

Supandi, Diana, Epha. 2020. Bahan Ajar Analisi Multivariate. Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.

Yulianto, Safa'at., Hidayatullah, Hilya, Kishera. 2014. Analisis Klaster Untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. Dalam jurnal Statistika, Vol. 2, No. 1: 56-63.