BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17 Issue 3 Page 1429–1438
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1429-1438

HIERARCHICAL CLUSTER ANALYSIS OF DISTRICTS/CITIES IN NORTH SUMATRA PROVINCE BASED ON HUMAN DEVELOPMENT INDEX INDICATORS USING PSEUDO-F

Neva Satyahadewi [1*], Steven Jansen Sinaga[2], Hendra Perdana[3]

[1,2,3]*Statistics Study Program, Faculty of Mathematics and Natural Sciences, Tanjungpura University*
Prof. Dr. H. Hadari Nawawi Street, Pontianak, 78124, Indonesia

*Corresponding author's e-mail: * **neva.satya@math.untan.ac.id***

Article History:

Received: 20th February 2023
Revised: 20th July 2023
Accepted: 29th July 2023

Keywords:

Hierarchical cluster;
Human development index indicators;
North Sumatra Province;
Pseudo-F statistics

ABSTRACT

Human development is needed to create prosperity and assist development in a country. In realising this, it is necessary to first look at the quality of human resources in the country, so that its use is more targeted. The measure used as a standard for the success of human development in a country is the Human Development Index (HDI). HDI figure are calculated from the aggregation of three dimensions, namely longevity and healthy living, knowledge, and decent standard of living. The longevity and healthy living dimension is represented by the Life Expectancy. Average Years of Schooling (AYS) and Expected Years of Schooling (EYS) are indicators representing the knowledge dimension. Meanwhile, the decent standard of living dimension is represented by the Expenditure per Capita indicator. The purpose of this study is to explain the characteristics of each cluster obtained from Hierarchical Cluster Analysis of districts/cities in North Sumatra Province based on HDI indicators in 2022 using Pseudo-F. The methods used are Hierarchical Cluster Analysis and Calinski-Harabasz Pseudo-F Statistic. The main concept of this method is to determine the optimum number of groups. This research uses secondary data obtained from BPS. The sample size in this study are 33 districts/cities and the number of variables are 4 variables. The results of the analysis of this study are the formation of 4 clusters with the best method is Ward. Cluster 1 consists of four members, namely Medan City, Pematang Siantar City, Binjai City, and Padang Sidempuan City, where this cluster has a very high HDI level. Meanwhile, Cluster 4 is a cluster that has a very low HDI level with four cluster members, namely Nias District, South Nias District, North Nias District, and West Nias District. Thus, it can be seen that there is a gap between regions in North Sumatra Province.

How to cite this article:

N. Satyahadewi, S. J. Sinaga and H. Perdana., "HIERARCHICAL CLUSTER ANALYSIS OF DISTRICTS/CITIES IN NORTH SUMATRA PROVINCE BASED ON HUMAN DEVELOPMENT INDEX INDICATORS USING PSEUDO-F," *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1429-1438, September, 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Human development is needed to create prosperity and assist development in a country. In realising this, it is necessary to first look at the quality of human resources in the country, so that its use is more targeted. It is hoped that the better quality of human resources can improve the quality of life of the community so that there is no social gap.

Human dignity should be the ultimate goal of a country's development. This is the recommendation of the world institution, the United Nations Development Programme (UNDP), which also states that a dignified human being is one who can enjoy a long, healthy life and live a productive life [1]. Therefore, until now the measure used as a standard for the success of human development in a country is the Human Development Index (HDI). The concept of human development itself basically has a very broad meaning. This concept covers all the basic dimensions of human beings. However, the basic idea of the human development concept at its core is simple: to create positive economic, social, political, cultural, and ecological growth, as well as a change in human well-being. Therefore, human beings must be positioned as the true wealth of the nation [2].

According to the Badan Pusat Statistik (BPS), an essential indicator to gauge success in building the quality of human life is the HDI. HDI is a comprehensive index calculated as a simplified average of three indices or dimensions that illustrate basic human capabilities in enlarging the choices of the people. The three indices are Life Expectancy, education, and decent standard of living [3].

BPS recorded the HDI of North Sumatra Province in 2022 at 72.71. The HDI figure is calculated from the aggregation of three dimensions, namely longevity and healthy living, knowledge, and a decent standard of living. Each dimension is being represented by indicators. The longevity and healthy living dimension is represented by the Life Expectancy at the birth indicator. Meanwhile, the Average Years of Schooling (AYS) and Expected Years of Schooling (EYS) are indicators representing the knowledge dimension. Finally, the decent standard of living dimension is represented by the adjusted Expenditure per Capita indicator. Based on data in 2022, Life Expectancy at birth in North Sumatra Province has reached 69.61 years. This figure shows that a newborn baby can survive until the age of 69.61 years. In terms of education, the population of North Sumatra Province aged 25 years and over has averaged 9.71 years of schooling or has completed grade IX. In addition, the average child aged 7 years who starts school is expected to receive up to 13.31 years of education or the equivalent of a Diploma I. On the economic aspect, Expenditure per Capita has reached Rp10,848,000 per capita per year.

BPS also noted that since 2020, human development in North Sumatra Province has shown positive developments. Every year, the HDI of North Sumatra Province increases by an average of 0.47% per year. In a period of two years, there has been an increase in HDI of up to 0.94 points. This development shows the general improvement of human development in North Sumatra Province.

Although the HDI of North Sumatra Province has successfully increased, the condition of human development in each district/city is still varied and uneven. Therefore, the clustering of districts/cities in North Sumatra Province is necessary for planning and evaluating government programme targets. The statistical analysis used is Hierarchical Cluster analysis which is then further analysed using Pseudo-F to determine the most optimum cluster.

There have been many previous studies related to HDI using Hierarchical Cluster analysis. Sharon Syalomitha R., et al. [4] grouped districts/cities in West Java Province according to HDI indicators in 2020. In this study, Hierarchical Cluster analysis was used, where of the five Hierarchical Cluster analysis methods, the optimum cluster with a cophenetic correlation value closest to 1 was obtained in the Average Linkage Method. The results of the 4 cluster grouping show that cluster 1 tends to have low HDI status, cluster 2 with high HDI status, cluster 3 with medium HDI status, and cluster 4 with very high HDI status.

Research on HDI using Hierarchical Cluster analysis was also conducted by Riyana Putri and Edy Widodo [5], namely Hierarchical Cluster analysis for clustering districts/cities in Central Java based on HDI indicators in 2015. Root Mean Square Standard Deviation (RMSSTD) is the validity index used to calculate the ideal number of groups. Based on the RMSSTD index value, the best methods are Average Lingkage, Complete Lingkage, and Ward with a total of 4 clusters. Cluster 1 consists of 19 districts/cities, cluster 2 consists of 3 districts/cities, cluster 3 consists of 10 districts/cities and cluster 4 consists of 3 districts/cities.

In previous studies, no one has discussed the clustering of districts/cities in North Sumatra Province based on HDI indicators using the Pseudo-F method. Where based on research conducted by Milligan and Cooper, it is found that the Pseudo-F method provides the best results among 30 methods and is a method that can be used globally [6].

Hierarchical Cluster analysis is an analysis in which the clustering of data is done by measuring the proximity distance of each object which then forms a dendrogram. There are several methods of hierarchical cluster analysis, including Single Linkage, Complete Linkage, Average Linkage, Ward, and Centroid [7]. Pseudo-F statistic is one of the commonly used methods to determine the optimum number of groups. Large values of Pseudo-F indicate that the number of clusters used to partition the data is optimum [8].

2. RESEARCH METHODS

2.1 Materials and Data

The data used in this study are secondary data obtained from BPS, namely data on HDI constituent indicators in districts/cities in North Sumatra Province in 2022. The variables of each HDI indicator include the following.

Table 1. Research Variables

Indicator	Variable	Description	Unit
Longevity and Healthy Living	X_1	Life Expectancy	Year
Education	X_2	Average Years of Schooling (AYS)	Year
	X_3	Expected Years of Schooling (EYS)	Year
Decent Standard of Living	X_4	Expenditure per Capita	Rp

***Data source**: BPS North Sumatra*

The sample unit in this study used districts/cities in North Sumatra Province, namely 25 districts and 8 cities. Data analysis procedures were carried out using R software. The package used in this study is 'library(factoextra)' [9].

2.2 Methods

2.2.1 Cluster Analysis

Cluster analysis is a multivariate technique used to group objects/cases (respondents) into groups where each group contains objects/cases that are similar to each other [10]. This analysis begins with the understanding that a certain amount of data actually has similarities among its members. Therefore, it is possible to group members who have similar characteristics in one or more than one group/cluster [11].

Talakua, Leleury, and Taluta [12] state that the characteristics of a good cluster are those that have:

1) High homogeneity (similarity) between members in one cluster (within cluster).

2) High heterogeneity (difference) between one cluster and another (between-cluster).

It means, that the variation members within cluster should be low, while the variation between clusters should be high. From these two things, it can be concluded that a successful cluster is a cluster that has a high level of similarity in characteristics between one object and another, but is not very similar to other clusters.

Before the clustering step, data standardisation is carried out if necessary. This needs to be done if the data units have a large difference. For example, if the Expenditure per Capita variable has units of millions, while a person's Life Expectancy only has units of tens, then this striking difference will make the distance calculation invalid. If there is a significant difference, then the data must be standardised by transforming the data to Z-Score. The standardisation process makes two data with a large unit difference automatically become narrower [13]. The way to define the standardisation score is by using the following equation [14].

$$Z_{i,j} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

Description:

$Z_{i,j}$: standardisation for *i*-th data on *j*-th variable
x_{ij} : data from the *i*-th object on the *j*-th variable
$\bar{x}_j$: mean of the *j*-th variable
s_j : standard deviation of the *j*-th variable

After the data that is deemed to have very distinct units is uniformed, the following step is to determine the distance measure. There are three methods that can be chosen to determine the distance measure between data, namely measuring the correlation between a pair of objects on several variables, measuring the association between objects, and measuring the distance between two objects. In this research, the method used is measuring the distance between two objects.

There are various distance measurements, the most popular being the euclidean distance method. Basically, this method will put an object into a certain cluster that measures the distance of the object to the cluster centre. If the object is within a certain distance, it can be included in the cluster. Here is the euclidean distance equation [15].

$$d_{hi} = \sqrt{\sum_{j=1}^{c}(x_{hj} - x_{ij})^2} \quad (2)$$

Description:

d_{hi} : distance between *h*-th object and *i*-th object
c : number of variables
x_{hj} : data from the *h*-th object on the *j*-th variable
x_{ij} : data from the *i*-th object on the *j*-th variable

Once the way to determine the distance is established, the next step is to create a grouping. There are two methods in clustering data, namely Hierarchical Method and Non-Hierarchical Method. In this research, the method used is the Hierarchical Method.

This method starts clustering with two or more objects that have the closest similarity. The process is then continued to other objects that have the second proximity. So on the cluster will create a kind of 'tree' in which there is a clear hierarchy (level) between objects, from the most similar characteristics to the most dissimilar. Logically all objects will ultimately only form one cluster. Dendrograms are commonly used to help make the hierarchical process clearer [13]. There are 5 methods contained in the Hierarchical Cluster analysis, including the following.

2.2.1.1 Single Linkage

In determining the distance between clusters using a Single Linkage, the closest distance or nearest neighbour rule is chosen [16]. The distance calculation is formulated as follows [17].

$$d_{(UV)W} = min(d_{UW}, d_{VW}) \quad (3)$$

Description:

$d_{(UV)W}$: distance between cluster (*UV*) and cluster *W*
$min(d_{UW}, d_{VW})$: nearest neighbour distance between clusters *U* and *W* or between clusters *V* and *W*

2.2.1.2 Complete Linkage

In the Complete Linkage method, the distance between clusters is defined by the farthest-neighbour distance between two objects in different clusters. Where it can be formulated as follows [18].

$$d_{(UV)W} = max(d_{UW}, d_{VW}) \quad (4)$$

Description:

$max(d_{UW}, d_{VW})$: the longest distance between cluster *U* and *W* or between cluster *V* and *W*

2.2.1.3 Average Linkage

This method groups objects based on the distance between two clusters which is considered as the mean distance between all members in a cluster with all members of another cluster. Under various circumstances, this method is considered more stable than the previous two methods, which can be formulated as follows [10].

$$d_{(UV)W} = average(d_{UW}, d_{VW}) \quad (5)$$

Description:

$average(d_{UW}, d_{VW})$: the average between clusters *U* and *W* and clusters *V* and *W*

2.2.1.4 Ward

This method seeks to minimise the variation between objects in a cluster. Where it can be formulated as follows [19].

$$d_{(UV)W} = \frac{[(n_W+n_U)d_{(UW)}+(n_W+n_V)d_{(VW)}]-n_W d_{(UV)}}{n_W+n_{(UV)}} \quad (6)$$

Description:

$d_{(UW)}$: distance between cluster *U* and *W*
$d_{(VW)}$: distance between cluster *V* and *W*
$d_{(UV)}$: distance between cluster *U* and *V*
$n_{(UV)}$: the number of objects in cluster *UV*
n_W : the number of objects in cluster *W*

2.2.1.5 Centroid

In this method, the distance between two clusters is the distance between the two centroid cluster. The centroid is the average distance in a cluster, which is derived by averaging all members of a given cluster. The centroid is only calculated when objects are merged, so every time the members increase, the centroid will change [20].

2.2.2 Calinski-Harabasz Pseudo-F Statistic

Pseudo-F statistic is one of the commonly used methods to determine the optimum number of groups. Research by Milligan and Cooper shows that the Pseudo-F statistic, hereafter referred to as Pseudo-F, gives the best results among 30 methods and is a method that can be used globally [6]. Pseudo-F was formulated by Calinski and Harabasz as well as expressed by the following equation [21].

$$Pseudo\ F = \frac{\frac{R^2}{p-1}}{\frac{1-R^2}{N-p}} \quad (6)$$

$$R^2 = \frac{SST-SSW}{SST} \quad (7)$$

$$SST = \sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{k=1}^{p}\left(x_{ijk} - \bar{x}_j\right)^2 \quad (8)$$

$$SSW = \sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{k=1}^{p}\left(x_{ijk} - \bar{x}_{jk}\right)^2 \quad (9)$$

Description:

SST : the total sum of the squares of the sample distances to the overall mean
SSW : the total sum of the squares of the sample distances to the cluster mean
n : number of samples in each cluster
c : number of variables
p : number of clusters
x_{ijk} : *i*-th sample in *j*-th variable of *k*-th cluster
$\bar{x}_j$: average of all samples in the *j*-th variable
$\bar{x}_{jk}$: average of samples in *j*-th variable and *k*-th cluster
N : number of samples

The highest value of Pseudo-F statistic indicates that the number of clusters used to partition the data is optimum.

3. RESULTS AND DISCUSSION

This section discusses the results of the analysis of grouping districts/cities based on HDI indicators in North Sumatra Province using Hierarchical Cluster. The discussion begins by presenting an overview to determine the characteristics of HDI indicators in North Sumatra. Then continued grouping with the Hierarchical Cluster analysis approach.

3.1 Overview of HDI in Districts/Cities at North Sumatra Province

The following are the data characteristics of the HDI constituent indicators in North Sumatra Province presented in **Table 2**. Based on HDI indicators from the dimension of longevity and healthy living indicated by Life Expectancy (X_1), it appears that the average Life Expectancy (X_1) in North Sumatra is 74.25 years with a diversity of 2.51. The Life Expectancy of the people in Mandailing Natal District is the smallest at 63.05 years, while the largest is the people in Pematang Siantar City at 74.25 years.

The next HDI indicator is from the knowledge dimension shown by variables X_2 and X_3. Based on **Table 2**, the Average Years of Schooling (AYS) in North Sumatra is 9.29 years with a diversity of 1.37. The people of Nias District have the smallest AYS value at 5.88 years, while the largest AYS is found in Medan City at 11.50 years. Then the average Expected Years of Schooling (EYS) in North Sumatra is 13.35 years with a diversity of 0.55. The people of South Nias District have the lowest EYS value of 12.48 years, while the highest EYS is in Medan City, which is 14.77 years.

Table 2. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
X_1	33	63.05	74.25	69.53	2.51
X_2	33	5.88	11.50	9.29	1.37
X_3	33	12.48	14.77	13.35	0.55
X_4	33	6,152.00	15,503.00	10,716.09	2,092.46

The last dimension is a decent standard of living as indicated by the Expenditure per Capita variable (X_4). Based on **Table 2**, it can be seen that in North Sumatra the average Expenditure per Capita is Rp10,716,090.91 with a variation of 2,092,463.13. The region with the smallest Expenditure per Capita is the people in West Nias District at Rp6,152,000, while the largest Expenditure per Capita is in Medan City at Rp15,503,000.

3.2 Hierarchical Cluster Analysis

The grouping of districts/cities in this study uses Hierarchical Cluster analysis which divides a number of data into several specific clusters. The approach chosen in conducting this cluster analysis is hierarchical because it is assumed that the number of clusters formed is unknown. The method chosen is the Hierarchical Cluster method which can provide results with the best accuracy. There are 5 methods in Hierarchical Cluster analysis, namely Single Linkage, Complete Linkage, Average Linkage, Ward, and Centroid. Furthermore, the clustering results of the five methods will be compared and the method with the most optimal cluster will be selected.

Given that the data units of the variables collected have a large scale of difference, it is necessary to standardise the data into Zscore form. The Zscore results are used as the basis for cluster analysis.

The next step was to measure the similarity between districts/cities in North Sumatra Province. The distance measure chosen in this study is the Euclidean distance. The lower the Euclidean distance between a couple of districts/cities, the more the characteristics of the two districts/cities are similar or almost the same.

The following stage is to decide the number of clusters in clustering districts/cities in North Sumatra Province. In this study, the number of clusters used is between 2 to 5 clusters. When conducting cluster analysis, what can be shown is only the cluster members at a certain number of clusters, but not to decide

how many clusters are formed. Therefore, to determine the optimum number of clusters, one way is to use the Pseudo-F value, the results of which have been presented in **Table 3**.

Table 3. Pseudo-F Statistics

Number of Clusters	Single	Complete	Average	Ward	Centroid
2	13.4573	15.6605	13.5962	13.4573	-
3	8.0124	-	18.6980	18.6980	8.7208
4	12.0721	13.2707	20.3423	21.5832	14.0940
5	11.2529	20.4382	-	-	11.2529

The Pseudo-F value shown in **Table 3** is a statistical tool used to determine the most optimum cluster. The greater the Pseudo-F value, the more optimum the clustering results. Based on this table, it can be observed that the highest Pseudo-F value is 21.5832, so the best cluster in classifying districts/cities in North Sumatra Province based on the indicators that make up HDI is the Ward method with 4 clusters. By using the Ward method, the members of each of the clusters are presented in **Table 4**.

Table 4. Cluster Member Details

Cluster Number	Cluster	Number of Members	Members in Cluster
1	Very High	4	Medan, Pematang Siantar, Binjai, and Padang Sidempuan
2	High	18	Sibolga, Tebing Tinggi, Gunungsitoli, Tapanuli Utara, Toba Samosir, Labuhan Batu, Asahan, Simalungun, Dairi, Karo, Deli Serdang, Langkat, Humbang Hasundutan, Samosir, Serdang Bedagai, Batu Bara, Labuhanbatu Selatan, and Labuhanbatu Utara
3	Low	7	Tanjung Balai, Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Pakpak Bharat, Padang Lawas Utara, and Padang Lawas
4	Very Low	4	Nias, Nias Selatan, Nias Utara, and Nias Barat

The characteristics of each cluster formed by the Hierarchical Method (Ward) are in **Table 5**. Cluster 1 is districts/cities with a very high HDI. All variables in cluster 1 have the highest values among the other clusters.

Table 5. Average HDI Indicator Variables for Each Cluster

Variable	Unit	Ward			
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
X_1	Year	72.62	70.24	65.88	69.65
X_2	Year	11.28	9.49	9.24	6.47
X_3	Year	14.45	13.18	13.41	12.89
X_4	Rp	12,735.50	11,304.83	10,297.71	6,779.50

Cluster 2 are districts/cities with relatively high HDI. In cluster 2, there are three HDI indicator variables that are classified as high, namely Life Expectancy, AYS, and Expenditure per Capita. Cluster 2 also has one HDI indicator variable that is classified as low, namely EYS.

Cluster 3 consists of districts/cities with relatively low HDI. Cluster 3 has one HDI indicator variable that is classified as high, namely EYS, two HDI indicator variables that are classified as low, namely AYS and Expenditure per Capita, as well as one very low HDI indicator variable, namely Life Expectancy.

Cluster 4 are districts/cities with a very low HDI. This is because in cluster 4, there are three HDI indicator variables that are very low, namely AYS, EYS, and Expenditure per Capita, and as well as one HDI indicator variable that is classified as low, namely Life Expectancy.

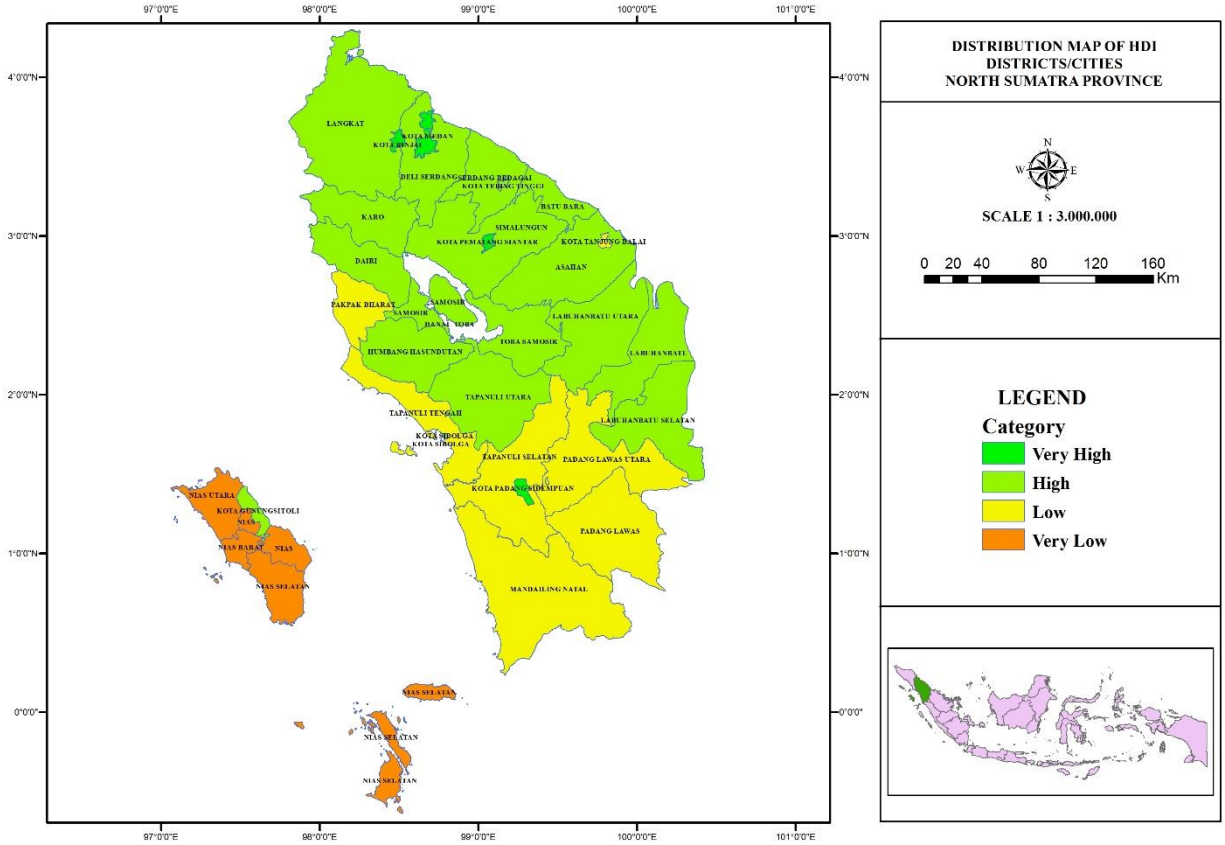

Figure 1. Map of Regional Grouping Results

Based on the map in **Figure 1**, it can be observed that Nias District, South Nias District, North Nias District and West Nias District have very low HDI levels. The western and southern regions of North Sumatra Province mostly have low HDI levels. Whereas in the eastern and northern regions, the majority have HDI levels that are classified as high. Meanwhile, Medan City, Pematang Siantar City, Binjai City, and Padang Sidempuan City are the areas with HDI levels in the very high category. The results of this study are in line with the research of Reinaldo Arifsantoso Purba et al. **[22]** which was published in the SMART AI journal where it was stated that there are 4 districts/cities in North Sumatra Province whose Human Development Index is classified as low, including Nias, South Nias, North Nias, and West Nias. This shows that the research results that have been obtained can be a reference material for the government in determining policies so that districts/cities with low HDI levels receive more attention, so that they can build the quality of life of the community evenly.

4. CONCLUSIONS

According to the analysis that has been carried out, the conclusion of this research is that the results of the clustering formed there are 4 clusters and the best method is Ward. This is because the Pseudo-F value of the Ward method is the largest compared to other methods, namely with a value of 21.5832. Each cluster has its own characteristics. Cluster 1 consists of four members, namely Medan City, Pematang Siantar City, Binjai City, and Padang Sidempuan City, where this cluster has a very high HDI level. All variables in cluster 1, namely Life Expectancy, AYS, EYS, and Expenditure per Capita have the highest value among other clusters. Cluster 2 consists of eighteen members whose HDI levels are included in the high category. Cluster 2 has high Life Expectancy, AYS, and Expenditure per Capita as well as low EYS. While cluster 3 consists of seven members whose HDI levels are included into the low category. Cluster 3 has high EYS, low AYS and Expenditure per Capita, as well as very low Life Expectancy. Finally, cluster 4 is a cluster that has a very low HDI level with four cluster members, namely Nias District, South Nias District, North Nias District, and West Nias District. Cluster 4 has very low AYS, EYS, and Expenditure per Capita, as well as low Life Expectancy. Thus, it can be seen that there is a gap between regions in North Sumatra Province.

REFERENCES

[1] United Nations Development Programme (UNDP), *Human Development Report 1990. Published for the United Nations Development Program (UNDP).* New York, Oxford: Oxford University Press. 1990.
[2] Badan Pusat Statistik, *Indeks Pembangunan Manusia 2018*.
[3] Badan Pusat Statistik, *Indeks Pembangunan Manusia 2006-2007*.
[4] S. Syalomitha R., S. Nadhira F., G. Anistya A., R. Raihana Z., J. Dominique P. S., C. Aliyyah N., A. Putri P., and T. Toharudin, "Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Menurut Indikator Indeks Pembangunan Manusia Tahun 2020," *Seminar Nasional Statistika X*, 2021.
[5] R. Putri and E. Widodo, "Analisis Klaster Hierarki untuk Pengelompokan Kabupaten/Kota di Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia (IPM) Tahun 2015," *Seminar Nasional Integrasi Matematika dan Nilai Islami*, vol. 1, no. 1, 2017.
[6] A. Mauludina and Setiawan, "Pengelompokan Kecamatan di Pulau Madura Berdasarkan Sektor Pertanian Sebelum dan Setelah Berdiri Jembatan Suramadu," *Sains dan Seni ITS*, vol. 1, no. 1, 2012.
[7] L. Ferreira and D. B. Hitchcock, "A Comparison of Hierarchical Methods for Clustering Functional Data," *Communications in Statistics: Simulation and Computation*, vol. 38, no. 9, 2009, doi: 10.1080/03610910903168603.
[8] Y. Yang, F. Wang, and F. Jiang, "A Topic Model for Hierarchical Documents," 2017. doi: 10.1109/DSC.2016.97.
[9] A. Kassambara and F. Mundt, "factoextra: Extract and Visualize the Results of Multivariate Data Analyses," *R Package Version 1.0.7.*, vol. 1, no. 3, 2020.
[10] N. Asiska, N. Satyahadewi, and H. Perdana, "Pencarian Cluster Optimum Pada Single Linkage, Complete Linkage dan Average Linkage," *Bimaster: Buletin Ilmiah Matematika, Statistika, dan Terapannya*, vol. 8, no. 3, 2019, doi: 10.26418/bbimst.v8i3.33173.
[11] A. Alfira, F. Hermin, and E. D. Wiraningsih, "Analisis Hybrid Mutual Clustering Menggunakan Jarak Square Euclidean," *JMT: Jurnal Matematika dan Terapan*, vol. 1, no. 1, 2017, doi: 10.21009/jmt.1.1.3.
[12] M. W. Talakua, Z. A. Leleury, and A. W. Taluta, "Analisis Cluster dengan Menggunakan Metode K-Means untuk Pengelompokan Kabupaten/Kota di Povinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 11, no. 2, 2017, doi: 10.30598/barekengvol11iss2pp119-128.
[13] H. Wicaksono, "Penilaian Hasil Kegiatan Belajar Mahasiswa Menggunakan Metode Cluster Non-Hierarki," *Infoman's*, vol. 11, no. 1, 2017, doi: 10.33481/infomans.v11i1.16.
[14] N. A. A. Lestari, M. N. Hayati, and F. D. T. Amijaya, "Analisis Cluster Pada Data Kategorik dan Numerik dengan Pendekatan Cluster Ensemble (Studi Kasus: Puskesmas di Provinsi Kalimantan Timur Kondisi Desember 2017)," *Jurnal Eksponensial*, vol. 11, 2020.
[15] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis: Pearson Prentice Hall*. 2007.
[16] A. A. Yusfar, M. A. Tiro, and Sudarmin, "Analisis Cluster Ensemble dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kinerja Pembangunan Ekonomi Daerah," *VARIANSI: Jurnal Statistics and Its Application on Teaching and Research*, vol. 3, no. 1, 2020, doi: 10.35580/variansiunm14626.
[17] M. Śmieja, D. Warszycki, J. Tabor, and A. J. Bojarski, "Asymmetric Clustering Index in a Case Study of 5-HT1A Receptor Ligands," *PLoS One*, vol. 9, no. 7, 2014, doi: 10.1371/journal.pone.0102069.
[18] M. Ackerman and S. Ben-David, "A Characterization of Linkage-based Hierarchical Clustering," *Journal of Machine Learning Research*, vol. 17, 2016.
[19] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura Journal of Mathematics*, vol. 1, no. 2, 2019, doi: 10.34312/jjom.v1i2.2354.
[20] A. C. Rencher and W. F. Christensen, *Methods of Multivariate Analysis: Third Edition*. 2012. doi: 10.1002/9781118391686.
[21] A. Ruda, "Exploring Tourism Possibilities using GIS-based Spatial Association Methods," *Geographia Technica*, vol. 11, no. 2, 2016, doi: 10.21163/GT_2016.112.09.
[22] R. A. Purba, S. Suhada, F. Anggraini, D. Hartama, and S. Retno, "Penerapan K-Means Clustering Pada Pengelompokan Kabupaten/Kota Sumatera Utara Berdasarkan Indeks Pembangunan Manusia," *SmartAI Journal*, vol. 1, no. 2, 2022.