EULER: Jurnal Ilmiah Matematika, Sains dan Teknologi
Journal Homepage: http://ejurnal.ung.ac.id/index.php/Euler
DOI: https://doi.org/10.37905/euler.v11i2.23113
Desember 2023, Vol. 11, No. 2, pp. 372-382

Determining the Optimum Number of Clusters in Hierarchical Clustering Using Pseudo-F

Steven Jansen Sinaga[1], Neva Satyahadewi[2*], Hendra Perdana[3]
[1,2,3]Statistics Study Program, Universitas Tanjungpura, Pontianak, Indonesia
**Corresponding Author. Email: neva.satya@math.untan.ac.id*

Abstract

Poverty refers to the condition where a person cannot meet the basic necessities based on the minimum living standards. Statistics Indonesia proxied an increase in the poverty rate in North Sumatra Province in 2021 from 8.75% to 9.01%. However, this increase is exclusive to North Sumatra Province, which has Indonesia's 3rd largest number of districts/cities. This study discussed mapping the North Sumatra Province region based on 10 poverty factor variables. The 10 variables are life expectancy, health complaints, poverty line, Gross Regional Domestic Product (GRDP), population growth rate, Expected Years of Schooling (EYS), Human Development Index (HDI), labor force participation rate, open unemployment rate, and district/city minimum wage. The Hierarchical Clustering analysis was employed to compare single, complete, and average linkage methods. The best method was determined based on the pseudo-F statistic value. 4 clusters had complete linkage methods, each of which possessed varied characteristics. Cluster 1 contains cities with the lowest poverty rate, including Medan City and Pematang Siantar City. Cluster 2 consists of cities with low poverty rates, while Cluster 3 consists of cities with high poverty rates. Cities that are included in Cluster 4 have very high poverty rates, including South Nias District and Pakpak Bharat District. The clusters present significant poverty rate gaps among North Sumatra Province regions.

Keywords: *Factors Affecting Poverty; Hierarchical Clustering; Pseudo-F*

1. Introduction

The biggest obstacle in efforts to reduce poverty in the country is currently related to the distribution of economic growth, which has yet to be spread evenly throughout Indonesia [1]. That means there is a disparity in economic conditions between urban and rural areas, where rural regions exhibit a higher poverty rate than urban areas. Urban areas are better organized and facilitated with adequate education, health, and infrastructure.

According to data provided by the Badan Pusat Statistik (BPS), North Sumatra Province has an area of 72,981 km^2 and a population of around 14.936 million people, and the number of individuals living in poverty will be around 1.34 million in 2021. The percentage of poverty in North Sumatra Province remains relatively large. According to BPS, the percentage of poverty in North Sumatra Province in 2020 was 8.75%. In 2021 it will increase to 9.01% [2]. This value certainly does not describe the poverty situation of each region. That means this value only applies to North Sumatra Province, which does not apply to each district/city. Apart from that, North Sumatra is the province in Indonesia with the third largest number of districts/cities.

Each region in North Sumatra Province has its characteristics, one of which is in the economic sector. North Sumatra's GRDP in 2020 is dominated by Medan City with a figure of 29.85%, followed by Deli Serdang District at 13.58%, Langkat District at 5.33%, and Simalungun District at 4.86%.

Submitted: 23 October 2023, Revised: 5 December 2023, Accepted: 16 December 2023

From the education sector, similar to GRDP, the school expectations of the people of North Sumatra Province are still influenced by socio-economic factors. Higher education infrastructure is more widely available in urban areas. Because urban areas are easier to access higher education than rural areas, urban areas with higher economic status are often the choice for continuing higher education. As an illustration, the City of Medan, the City of Pematang Siantar, and the City of Padang Sidempuan have the highest expected years of study for residents who have reached university level.

There has been much research related to poverty. Zuhdiyaty and Kaluge [3] examined the factors influencing poverty in Indonesia during the last five years, namely 2011-2015. The data utilized in this study were sourced from the BPS, which was then analyzed using a quantitative approach and panel data regression. Human Development Index (HDI), economic growth, and Open Unemployment Rate are some variables used. This research indicates that economic growth and Open Unemployment Rate have no impact on poverty in Indonesia, whereas HDI does. Then Aprilia [4], in her research on the influence of economic growth, minimum wages, education, and unemployment rates on poverty levels, found that economic growth, minimum wages, and education had a negative and significant impact on poverty levels. In contrast, the unemployment rate positively and significantly affected poverty levels.

Lisnawati [5] researched the influence of health, education level and investment on poverty in Padang City. The results of the data analysis show that health, education level and investment have a negative and significant effect on poverty in Padang City. Meanwhile, Mirah, et al. [6] researched the influence of labour force participation rates on economic growth and poverty in North Sulawesi Province. It was found that the level of labour force participation of men and women was able to have a positive and significant influence on the development of economic growth and reducing poverty rates in North Sulawesi Province, while economic growth was unable to have an impact on reducing poverty in North Sulawesi Province.

Kevin, et al. [7] in his research stated that simultaneously the variables inflation and population growth rate had a significant effect on poverty. Partially, the inflation variable has no significant effect on poverty, while the population growth rate variable has a significant effect on poverty. Then, Leasiwal [8] said that poverty in Maluku is dominated by people who live in rural areas. The variables that significantly influence poverty are people's purchasing power, inflation, Average Years of Schooling (AYS), literacy rate, gross enrollment rate, life expectancy rate, and the number of senior high schools.

Regarding cluster analysis research, Wahyuni and Jatmiko [9] grouped districts/cities on the island of Java based on poverty factors. These factors include the percentage of households working in agriculture, Average Years of Schooling (AYS), household expenditure per capita, and Open Unemployment Rate. The method approach used is average linkage. As a result, there are 2 clusters, namely districts/cities with low poverty levels, represented by Cluster 1, and Cluster 2 representing districts/cities with high poverty levels.

So far, numerous research aimed at categorizing districts/cities in North Sumatra Province based on poverty factors, utilizing methods such as single linkage, complete linkage, average linkage, K-Means, K-Medoids, etc. However, no one has compared these methods in cases of poverty in North Sumatra Province. Recognizing the necessity, researchers aim to categorize districts/cities in North Sumatra Province based on poverty factors, employing a comparative analysis of several methods using Pseudo-F statistics values so that the best method is obtained and the most optimum grouping (cluster) is obtained.

Finding the characteristics of North Sumatra Province, grouping was carried out based on the diversity of these characteristic. This study employs Hierarchical Clustering to categorize districts/cities in North Sumatra Province. The concept behind this method is to unite two districts/cities with the most comparable characteristics, followed by the combination of these pairs with one or more additional districts/cities that share the highest similarity, forming a hierarchy (sequence) of districts/cities within the cluster [10]. This research aims to produce a categorization of districts/cities in North Sumatra Province using a Hierarchical Clustering analysis approach based on poverty factors. The research findings can be used as a reference and consideration by the government in formulating poverty reduction policies to reduce poverty further.

2. Research Methods

This research relies on secondary data obtained from the BPS, specifically focusing on poverty factors in districts/cities in North Sumatra Province in 2021. The variables for each poverty indicator are as follows.

Table 1. The Research Variable

Indicator	Variable	Description	Unit
Health	X_1	Life Expectancy	Year
	X_2	Health Complaints	%
Isolation	X_3	Poverty Line	IDR
	X_4	Gross Regional Domestic Product (GRDP)	IDR
	X_5	Population Growth Rate	%
Education	X_6	Expected Years of Schooling (EYS)	Year
	X_7	Human Development Index (HDI)	%
Employment	X_8	Labour Force Participation Rate	%
	X_9	Open Unemployment Rate	%
	X_{10}	District/City Minimum Wage	IDR

The study employed districts and cities within North Sumatra Province as its sample units, specifically comprising 25 districts and 8 cities. The data analysis was conducted utilizing the R software. The utilized package for this research is 'library(factoextra)' [11].

2.1 Cluster Analysis

Cluster analysis is a multivariate method employed to categorize objects/cases (respondents) into several groups, with each group comprising objects/cases that share similarities with one another [12]. The basic understanding of this analysis is that there are shared similarities between members within a given set of data. Thus, members with the same characteristics can be grouped into one or more groups/clusters [13].

Talakua et al. [14] asserts that the features of an exemplary cluster are those that possess the following:

1) Cluster members have significant homogeneity (similarity) to each other.
2) Significant heterogeneity between different clusters.

From the two points above, it can be deduced that the best cluster is characterized by a high degree of similarity in features among its constituent objects, while exhibiting distinct differences from other clusters.

Prior to initiating the clustering process, data standardization is implemented when deemed necessary. This is important to address cases where the data units exhibit significant differences. For instance, suppose the poverty line variable is measured in hundreds of thousands, whereas the human life expectancy is measured in tens, such a pronounced significant disparity could result in an invalid distance calculation in that case. Z-Score should be used to standardize data if significant unit differences occur. Two data sets with significant unit disparities are automatically narrowed due to the standardization process [15]. The standardization value can be determined using the following equation [16].

$$Z_{i,j} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{1}$$

where:

$Z_{i,j}$: standardization for the *i*-th data of the *j*-th variable
x_{ij} : data from the *i*-th object in the *j*-th variable
μ_j : means of the *j*-th variable
σ_j : standard deviation of the *j*-th variable.

Following the standardization of data with distinct units, the subsequent step involves determining a distance measure. Three methods are available for determining the distance between data: assessing the correlation between a pair of objects across multiple variables, evaluating the association between objects, and measuring the distance between two objects. In this study, the chosen method involves measuring the distance between two objects.

The method of measuring distance known as Euclidean distance is extensively used. This method puts objects into specific clusters that measure the object's distance from the cluster's center. Objects can be included in the cluster within a specific range. The Euclidean distance equation is formulated as follows [17].

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

where:

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

After determining the distance, the next step involves form groupings. Two methods are available for grouping data: the Hierarchical Method and Non-Hierarchical Method. This research adopts the Hierarchical Method. Dendrograms are employed to elucidate hierarchical processes [15]. This study utilizes three methods of Hierarchical Clustering analysis, including the following.

2.1.1 Single Linkage

The proximity or nearest neighbor principle is employed to calculate the distance between clusters through the application of a single linkage [18]. The calculation formula for determining the distance is as follows [19]:

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

where:

$d_{(UV)W}$: distance between *cluster* (*UV*) dan *cluster W*

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

2.1.2 Complete Linkage

The complete linkage method employs the distance between the farthest neighbors in separate clusters to establish the distance between the clusters. Can be formulated with [20]:

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

where:

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

2.1.3 Average Linkage

This approach clusters objects by considering the average distance of all objects within one cluster to the average of all objects in another cluster. In various situations, this approach is deemed more reliable compared to the previous two approaches, and it can be expressed as follows [12]:

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

where:

$\text{average}(d_{UW}, d_{VW})$: the average between *cluster U* and *W* with *cluster V* and *W*.

2.2 Calinski-Harabasz Pseudo-F Statistic

The approach commonly used for identifying the optimum number of clusters is Pseudo-F statistics. Pseudo-F statistics, also referred to as Pseudo-F, shows superior performance among 30 methods and are generally considered applicable, according to research conducted by Milligan and Cooper [21]. Calinski and Harabasz formulated Pseudo-F in the following equation [22].

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

$$R^2 = \frac{SST - SSW}{SST} \tag{7}$$

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

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

where:

$SST\ (Sum\ Square\ Total)$: the total sum of the squares of the sample distance to the overall average

$SSW\ (Sum\ Square\ Within)$: the total sum of the squares of the sample distance to the cluster average

n : the number of samples in each cluster

c : the number of variables

p : the number of clusters

x_{ijk} : *i*-th sample in the *j*-th variable of the *k*-th cluster

$\bar{x}_j$: average of all samples on the *j*-th variable

$\bar{x}_{jk}$: sample average of the *j*-th variable and *k*-th cluster

N : the number of samples.

The most optimum number of clusters for partitioning the data is revealed by the highest value of the Pseudo-F statistic [23].

3. Results and Discussion

In this segment, we delve into the outcomes derived from examining the district/city classifications grounded in poverty factors within North Sumatra Province, employing the Hierarchical Clustering methodology. The discourse initiates by offering a comprehensive overview to comprehend the poverty characteristics in North Sumatra Province. Subsequently, followed by the application of the Hierarchical Clustering analysis approach for grouping.

3.1 General Description of District/City Poverty in North Sumatera Province

The data characteristics of poverty factors in North Sumatra Province are outlined, presented in Table 2. Analyzing poverty indicators related to health factors shown by life expectancy (X_1) and health complaints (X_2), the average life expectancy (X_1) in North Sumatra Province is 69.1 years, with a diversity of 2.51. Mandailing Natal District records the most minor life expectancy, namely 62.65 years, while Pematang Siantar City exhibits for the highest, which is 73.77 years. Then, the percentage of health complaints in North Sumatra Province, shown by X_2, has an average of 21.35% with a diversity of 6.05. The smallest percentage was found in the people of Binjai City at 10.51%, and then the most significant percentage was experienced by the people of Padang Sidempuan City at 32.02%.

Table 2. Statistics Descriptive

	N	Minimum	Maximum	Mean	Std. Deviation
X_1 [Life Expectancy]	33	62.65	73.77	69.15	2.51
X_2 [Health Complaints]	33	10.51	32.02	21.35	6.05
X_3 [Poverty Line]	33	329,308	583,588	442,529.64	58,563.93
X_4 [Gross Regional Domestic Product (GRDP)]	33	11,832,505	64,078,946	30,451,109	13,392,033.83
X_5 [Population Growth Rate]	33	0.52	2.46	1.30	0.42
X_6 [Expected Years of Schooling (EYS)]	33	12.27	14.75	13.25	0.58
X_7 [Human Development Index (HDI)]	33	61.99	81.21	71.06	4.52
X_8 [Labour Force Participation Rate]	33	61.84	87.70	72.84	7.85
X_9 [Open Unemployment Rate]	33	0.70	11.00	4.91	2.88
X_{10} [District/City Minimum Wage]	33	2,499,423	3,329,867	2,748,150	224,641.57

The next indicator of poverty is the isolation factor shown by variables X_3, X_4, and X_5. The variable is the poverty line, where in Table 2, the average poverty line in North Sumatra Province is IDR 442,529.64/capita/month, with a diversity of 58,563.93. The most minor nominal is IDR 329,308 per capita/month in the South Nias District area, and the largest nominal is in the Pematang Siantar City area, worth IDR 583,588 per capita/month. The following variable is X_4, namely Gross Regional Domestic Product (GRDP). The mean GRDP in North Sumatra Province is IDR 30,451,109.33, with a diversity of 13,392,033.83. South Nias District is the district with the smallest GRDP, IDR 11,832,505, while the largest GRDP is located in Medan City, with IDR 64,078,946. Then, variable X_5 is the population growth rate. The mean population growth rate in North Sumatra Province is 1.3%, with a standard deviation of 0.42. Sibolga City recorded the lowest population growth rate at 0.52%, while the highest was in Pakpak Bharat District, with a value of 2.46%.

The next factor is education, which includes the expected length of schooling variable (X_6) and the human development index (X_7). Based on Table 2, the mean Expected Years of Schooling (EYS) in North Sumatra Province is 13.25 years with a variation of 0.58. South Nias District records the lowest EYS value, namely 12.27 years, while the highest EYS is observed in Medan City, 14.75 years. The mean Human Development Index (HDI) in North Sumatra Province is 71.06, exhibiting a standard deviation of 4.52. West Nias District has the lowest HDI value at 61.99, while Medan City boasts the highest HDI at 81.21.

The final factor is employment, which includes the labour force participation rate (X_8), the open unemployment rate (X_9), and the district/city minimum wage (X_{10}). Table 2 indicates that the average Labour Force Participation Rate in North Sumatra Province is 72.8%, with a standard deviation of 7.85. Labuhan Batu District has the lowest Labour Force Participation Rate at 61.84%, while the largest Labour Force Participation Rate is in Pakpak Bharat District at 87.7%. Furthermore, the average Open Unemployment Rate in North Sumatra Province is 4.9%, with a diversity of 2.88. Samosir District records the smallest Open Unemployment Rate at 0.7%, while the most significant percentage is in Pematang Siantar City, 11%. Finally, there is the District/City Minimum Wage variable (X_{10}), where the average District/City Minimum Wage in North Sumatra Province is IDR 2,748,150 with a variation of 224,641.57. The smallest District/City Minimum Wage is in the Pakpak Bharat District, IDR 2,499,423, whereas the highest District/City Minimum Wage is in Medan City at IDR 3,329,867.

3.2 Cluster Hierarchical Analysis

The district/city classification in this research employs Hierarchical Clustering analysis, a technique that categorizes data into distinct clusters. The selection of the hierarchical approach is based on the assumption that many of the clusters formed are not predetermined. The chosen method for this cluster analysis is the Hierarchical Cluster method, selected for its ability to yield the most accurate results. Three Hierarchical Clustering analysis methods are used in this research: single linkage, complete linkage, and average linkage. Next, the outcomes of these three methods are compared, and the method that produces the most optimum clusters is chosen.

Considering the substantial scale disparity among the collected variable data units, it is essential to standardize the data into Z_{score} form. The outcomes of Z_{score} standardization serve as the foundation for cluster analysis.

Then, the following stage involved assessing the similarities among districts/cities in North Sumatra Province. The distance measurement adopted in this investigation is the Euclidean distance. A smaller Euclidean distance between two districts/cities indicates a higher degree of similarity or nearly identical characteristics between them.

The subsequent step involves determine the quantity of clusters within the district/city groupings in North Sumatra Province. This research used many clusters, namely between 2 and 5 clusters. The only thing that can be shown in cluster analysis is the membership of a specific number of clusters, not the total number formed. Consequently, one way for determining the optimum number of clusters is by employing the Pseudo-F value, and the outcomes are presented in Table 3.

Table 3. *Pseudo-F Statistics* for Three Cluster Methods

The Number of Cluster	*Pseudo-F* Value		
	Single	Complete	Average
2	4.6138	6.8574	4.6138
3	3.7099	5.6136	4.3510
4	3.2605	**7.0978**	5.0960
5	3.4555	6.7712	4.2597

The Pseudo-F value, as indicated in Table 3, serves as a statistical approach to identify the optimum cluster. The grouping outcomes are considered more favorable when the Pseudo-F value is higher. Based on this table, the highest Pseudo-F value is 7.0978, indicating that the complete linkage method with 4 clusters is the most effective approach for grouping districts/cities in North Sumatra

Province based on poverty factors. The members in each cluster are presented in Table 4 using the complete linkage method.

Table 4. Cluster Member Details

Cluster Number	Cluster	Many Members	Members in The Cluster
1	Very Low	2	Medan and Pematang Siantar
2	Low	20	Sibolga, Tanjung Balai, Tebing Tinggi, Binjai, Padang Sidempuan, Gunungsitoli, Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Labuhan Batu, Asahan, Simalungun, Deli Serdang, Langkat, Serdang Bedagai, Batu Bara, Padang Lawas Utara, Padang Lawas, Labuhanbatu Selatan, and Labuhanbatu Utara
3	High	9	Nias, Tapanuli Utara, Toba Samosir, Dairi, Karo, Humbang Hasundutan, Samosir, Nias Utara, and Nias Barat
4	Very High	2	Nias Selatan and Pakpak Bharat

Table 5 displays each cluster's characteristics using the Hierarchical Method (complete linkage). Cluster 1 is a district/city cluster where poverty is very low. Within Cluster 1, the values of X_1, X_3, X_4, X_6, X_7, X_9, and X_{10} are the most elevated compared to the other clusters. Additionally, this cluster exhibits the lowest values for two poverty indicators: health complaints (X_2) and Labour Force Participation Rate (X_8).

Table 5. The Average Poverty Variable for Each Cluster

Variable	Unit	Average			
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
X_1	Year	73.50	68.54	69.91	67.41
X_2	%	15.53	22.97	19.17	20.83
X_3	IDR	580,357.00	442,896.55	435,875.67	330,976.00
X_4	IDR	49,670,245.50	34,077,006.55	21,756,700.78	14,097,839.50
X_5	%	1.33	1.23	1.24	2.28
X_6	Year	14.66	13.16	13.18	13.07
X_7	%	80.19	71.54	69.29	65.15
X_8	%	65.48	68.64	82.22	79.98
X_9	%	10.91	5.98	1.70	2.64
X_{10}	IDR	2,915,693.00	2,813,294.25	2,614,659.56	2,529,879.50

Cluster 2 represents districts/cities characterized by low poverty levels. Cluster 2 exhibits a notable distinction in one poverty indicator variable when compared to the other clusters, specifically the population growth rate (X_5). Cluster 2 has the lowest population growth rate compared to the other clusters.

Cluster 3 encompasses districts/cities characterized by high poverty levels. Within Cluster 3, there are variable characteristics whose values vary. Cluster 3 has the highest poverty indicator variable compared to the other clusters, namely the labour force participation rate (X_8). This cluster also has one of the lowest poverty indicator variables: the open unemployment rate (X_9).

Cluster 4 consists of districts/cities with very high poverty levels. Cluster 4 has the two highest poverty indicator variables: health complaints (X_2) and population growth rate (X_5). This cluster also has the six lowest poverty indicator variable values, namely X_1, X_3, X_4, X_6, X_7, dan X_{10}.

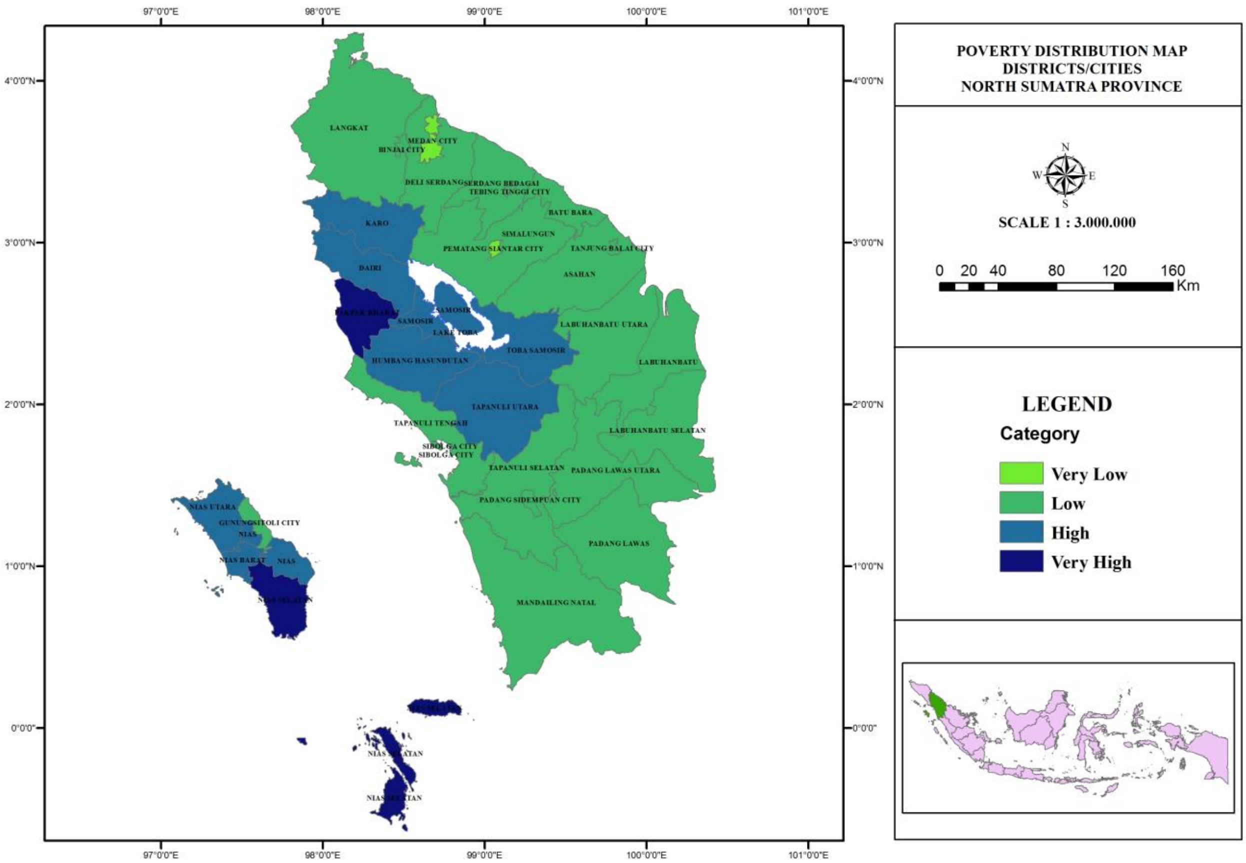

Figure 1. Map of Regional Grouping Results

Examining the map depicted in Figure 1 reveals that South Nias District and Pakpak Bharat District have very high levels of poverty. Most of the western part of North Sumatra Province has a high poverty level. Meanwhile, in the eastern region, the majority have relatively low poverty levels. Meanwhile, Medan City and Pematang Siantar City are areas where the poverty level is in the very low category.

4. Conclusion

Conclusively, the research analysis yielded the formation of four clusters, with the complete linkage method identified as the most effective. Each cluster exhibits distinct characteristics. Cluster 1, comprising Medan City and Pematang Siantar City, demonstrates a remarkably low poverty level. Cluster 2 includes twenty members characterized by a low poverty level. In contrast, cluster 3 encompasses nine members with a high poverty level. Cluster 4 represents regions with a very high poverty level, featuring South Nias District and Pakpak Bharat District as its two members. Consequently, a notable disparity exists among the regions in North Sumatra Province.

References

[1] B. Tisniwati, "Analisis Faktor-Faktor yang Mempengaruhi Tingkat Kemiskinan di Indonesia," *Jurnal Ekonomi Pembangunan*, vol. 10, no. 1, 2012, doi: 10.22219/jep.v10i1.3714.

[2] BPS, "Provinsi Sumatera Utara dalam Angka 2022," *Book*, 2022.

[3] N. Zuhdiyaty and D. Kaluge, "Analisis Faktor-Faktor yang Mempengaruhi Kemiskinan di Indonesia Selama Lima Tahun Terakhir," *Jurnal Ilmiah Bisnis dan Ekonomi Asia*, vol. 11, no. 2, 2017, doi: 10.32812/jibeka.v11i2.42.

[4] R. D. Aprilia, "Pengaruh Pertumbuhan Ekonomi, Upah Minimum, Pendidikan, dan Tingkat

Pengangguran Terhadap Tingkat Kemiskinan (Studi Kasus Kabupaten/Kota di Jawa Timur Tahun 2008-2013)," *Jurnal Ilmiah Mahasiswa FEB*, vol. 4, no. 2, 2016.

[5] D. Lisnawati, "Pengaruh Tingkat Kesehatan, Tingkat Pendidikan dan Investasi Terhadap Kemiskinan di Kota Padang," *Journal of Chemical Information and Modeling*, vol. 53, no. 9, 2017.

[6] M. R. Mirah, P. Kindangen, and I. P. F. Rorong, "Pengaruh Tingkat Partisipasi Angkatan Kerja Terhadap Pertumbuhan Ekonomi dan Kemiskinan di Provinsi Sulawesi Utara," *Jurnal Pembanguan Ekonomi dan Keuangan Daerah*, vol. 20, no. 04, 2020.

[7] K. Kevin, A. K. Putri, and A. Nasrun, "Pengaruh Inflasi dan Laju Pertumbuhan Penduduk Terhadap Kemiskinan di Sumatera Bagian Selatan Tahun 2011-2018," *SOROT*, vol. 15, no. 1, 2020, doi: 10.31258/sorot.15.1.33-42.

[8] T. C. Leasiwal, "Determinan dan Karakteristik Kemiskinan di Provinsi Maluku," *Cita Ekonomika Jurnal Ekonomi*, vol. VII, no. 2, 2013.

[9] S. Wahyuni and Y. A. Jatmiko, "Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Faktor-Faktor Kemiskinan dengan Pendekatan Average Linkage Hierarchical Clustering," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 10, no. 1, 2018, doi: 10.34123/jurnalasks.v10i1.197.

[10] A. Yulianto, Roni, R. Setiadi, and A. Indriyani, "Analisis Positioning Operator Seluler dan Strategi Pemasaran untuk Memenangkan Pasar di Wilayah Yogyakarta," *Jurnal Economic and Management (JECMA)*, vol. 1, no. 1, 2020, doi: 10.46772/jecma.v1i01.223.

[11] 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.

[12] 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.

[13] 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.

[14] M. W. Talakua, Z. A. Leleury, and A. W. Talluta, "Analisis Cluster dengan Menggunakan Metode K-Means untuk Pengelompokan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 11, no. 2, 2017, doi: 10.30598/barekengvol11iss2pp119-128.

[15] 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.

[16] 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.

[17] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis: Pearson Prentice Hall*. 2007.

[18] 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: Journal of Statistics and Its Application on Teaching and Research*, vol. 3, no. 1, 2021, doi: 10.35580/variansiunm14626.

[19] 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.

[20] M. Ackerman and S. Ben-David, “A Characterization of Linkage-based Hierarchical Clustering,” *Journal of Machine Learning Research*, vol. 17, 2016.

[21] 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, doi: 10.12962/j23373520.v1i1.582.

[22] 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.

[23] Y. Yang, F. Wang, and F. Jiang, “A Topic Model for Hierarchical Documents,” 2017. doi: 10.1109/DSC.2016.97.