

Regional Clustering Based on HDI Components in Indonesia

Azlam Nas*, Sri Mulatsih, Muhammad Findi

Department of Economics, IPB University, Bogor, Indonesia

ABSTRACT

Article Info

Volume 8 Issue 1

Page Number: 21-25

Publication Issue :

January-February-2021

Article History

Accepted : 05 Jan 2021

Published : 13 Jan 2021

This research aims to see the distribution of HDI in Indonesia as seen from its grouping based on its constituent components. The panel data regression model is formed from the results of cluster analysis using the K-means method using the Manhattan distance. The data used for cluster analysis is the panel data of the HDI components by province in 2010-2019. The results of the cluster analysis form 5 optimum groups, with the characteristics of group 1, HDI is relatively moderate, group 2 is also relatively moderate but not better than group 1, group 3 is relatively high, group 4 is relatively low, and group 5 is relatively very high.

Keywords : HDI, Cluster Analysis, K-Means, Manhattan Distance

I. INTRODUCTION

Development is a tool used to achieve the goals of the nation. In the Preamble to the 1945 Constitution, the objectives of the Indonesian nation are stated, namely, to promote general welfare and to educate the nation's life. The achievement of these goals can be measured through the Human Development Index (HDI) (Mirza, 2011).

High economic growth is the main target of development in developing countries, because it is closely related to the increase in goods and services produced in society. The more goods and services produced, will absorb a lot of labor, which in turn can increase welfare. As a production factor, quality human resources will determine the success of

economic development. Humans are not only an object of development but also a subject, which contributes to the progress of an area which at the macro level becomes the progress of a country.

The government plays an important role as a formulator and determinant of policies in achieving quality and prosperous human development. In the modern economy, the role of government is grouped into three, namely roles in allocation, distribution and stabilization (Musgrave, 1989). In implementing its role, the Indonesian government has adopted the SDGs (Sustainable Development Goals) as outlined in *Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024*.

Within the SDGs, there are several targets related to human development, namely the third goal to ensure a healthy life and support the welfare of all for all ages, the fourth goal is ensuring quality education and supporting lifelong learning opportunities for all, and the eighth goal is economic growth and a decent job. Supporting sustainable and inclusive economic development, full and productive employment and decent work for all (Rita, 2020)

Regional clustering needs to be done as material for planning and evaluating government program targets. The goal is to increase the human development rate based on the HDI components. Therefore, a statistical analysis is used, namely cluster analysis using the K-Means method with Manhattan distance. The analysis was carried out by grouping the Provinces based on the HDI components. Existing components are stated as variables and HDI is stated as objects that are grouped based on similar characteristics. Every other object is in the same cluster (Talakua, 2017).

This study classifies provinces in Indonesia based on similarity in regional characteristics in terms of 4 HDI components. These components include life expectancy (AHH), average length of schooling (RLS), length of schooling expectations (HLS), and per capita income. The results of the clustering from this study can be used as a reference for grouping based on the condition of the HDI profile in Indonesia.

II. METHODS AND MATERIALS

The data used in this research is the life expectancy (AHH) generated from the population census. Meanwhile, the expected length of schooling (HLS), average length of schooling (RLS), and per capita income are generated from the national socio-economic survey (Susenas). Type of data is panel data with 33 provinces between 2010 and 2019.

Cluster analysis is an analysis technique that uses data variables to classify objects with similar

characteristics. In cluster analysis, there is no response variable, so the existing data only has indicator variables that will be used to classify existing objects. Objects with similar characteristics tend to be included in one group and vice versa with different characteristics in other groups (Mattjik et al., 2011).

Clustering Analysis in Panel Data is different from cluster analysis in cross-section data, in this clustering the algorithm that used must be able to support the structure of panel data. K-Means method can be used in panel data clustering by modifying the size of the distance adjusted to the panel data structure. However, there are several approaches that can be used for clustering panel data (Cahyani, 2020).

Based on the results of Cahyani's (2020) research, K-means cluster analysis with panel data, the best is to use the Manhattan distance. The Manhattan distance by setting $p = 1$, the definition of the Manhattan distance is given by :

$$Dist(y_{1..}, y_{2..}) = \sum_{j=1}^T \sum_X^M [y_{1jX} - y_{2jX}]$$

The maximum distance is obtained by passing to the limit $p \rightarrow +\infty$, it takes the form :

$$Dist(y_{1..}, y_{2..}) = \left\{ \sum_{j=1}^T \sum_X^M [y_{1jX} - y_{2jX}]^p \right\}^{\frac{1}{p}}, p \rightarrow \infty$$

In this research the number of clusters was not determined. To see the optimum number of clusters, criteria of the Calinski and Harabasz values were used. Criteria Calinski and Harabasz (Genolini, 2016) is given in :

$$C(k) = \frac{Trace(B) n - k}{Trace(W) k - 1}$$

Determine the optimum number of groups based on the largest Calinski and Harabasz criteria value. Thus, draw a conclusion from the result of evaluated clustering with the best distance measurement.

III.RESULTS AND DISCUSSION

A. The Optimum Number of Clusters

The results of data processing with software R showed that the optimum value of the Calinski and Harabasz (CH) criteria was 23.8791 (Table I) with the number of groups is 5. The number of clusters tried was $k = 2$ to $k = 10$.

TABLE I
CALINSKI & HARABASZ CRITERION SCORES

Cluster (k)	CH
2	19,1134
3	17,1325
4	19,4149
5	23,8791
6	22,2879
7	20,8619
8	21,2565
9	20,5289
10	18,7590

B. Member of Clusters

The results of regional clustering based on the achievement of its HDI-forming components with analysis of the K-Means cluster with Manhattan distances (Table II) show that group 1 has the most members, with 10 members. Unlike the case with group 5 which only has 1 member is the province of DKI Jakarta. Group 2 has 8 members, group 3 has 7 members and group 4 has 7 members.

TABLE III

THE RESULTS OF CLUSTER ANALYSIS ARE BASED ON THE ACHIEVEMENTS OF THE COMPONENTS OF THE HDI, 2010-2019

Cluster	Province	Number of Members
1	Jambi, South Sumatra,	10

	Lampung, Bangka Belitung Islands, West Java, Central Java, East Java, Central Kalimantan, South Kalimantan, South Sulawesi	
2	Aceh, North Sumatra, West Sumatra, Bengkulu, Central Sulawesi, Southeast Sulawesi, Maluku, North Maluku	8
3	Riau, Riau Islands, DI Yogyakarta, Banten, Bali, East Kalimantan, North Sulawesi	7
4	West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Gorontalo, West Sulawesi, West Papua, Papua	7
5	DKI Jakarta	1

C. Characteristics of Cluters

The clusters formed from processing have inter-group characteristics, namely as follows:

- 1) Cluster 1: the HDI value of this group is relatively moderate, as indicated by the HDI average value of 68.27 and the growth rate of 0.92%. AHH is relatively high with an average value of 70.19 years and an average growth rate of 0.22%. The HLS figure is also relatively moderate when compared to other clusters, this can be seen from the average value of 12.03 years. Meanwhile, the RLS-figure is relatively low, namely 7.60 years. Seen from per capita income, this group is relatively high with an average of Rp. 10 million, the average growth rate was the highest compared to other clusters, namely 2.38%, as shown in Table III.
- 2) Cluster 2: The average HDI score of 68.01 indicates that group 2 has a relatively moderate HDI score, but when compared to group 1, the HDI for group 2 is not better. AHH value is also relatively moderate with an average of 68.05 years. The HLS

rate for group 2 was the highest among other clusters, namely 13.01 years. The RLS figure is relatively high, namely 8.51 years. Group 2 income per capita is relatively low with an average value of Rp. 8.79 million.

3) Cluster 3: Group 3 HDI is relatively high with an average of 72.67. AHH is also relatively high with an average value of 71.36 years. The education index for group 3 is relatively high, this can be seen from the HLS and RLS figures, which have an average value of 12.84 years and 8.81 years, respectively. In terms of per capita income, group 3 is relatively high with an average of Rp. 11.72 million, but the growth rate was slow compared to other clusters, namely 1.79%.

4) Cluster 4: HDI in this group is the lowest among other clusters. The HDI average value was 62.76. AHH is also relatively low with an average of 65.96 years. The HLS is relatively low with an average of 11.87 years, as well as the relatively low RLS number with an average of 6.79 years. Seen from the economic index, per capita income is also relatively low with an average of Rp. 7.9 million.

5) Cluster 5: The HDI of this group is relatively very high, with an average of 78.72. AHH value is relatively high with an average of 72.30 years. The HLS rate is relatively high with an average of 12.45 years, while the RLS rate is the highest among other clusters with an average of 10.69 years. Meanwhile, the average per capita income reached Rp. 17 million, relatively very high compared to other clusters.

TABLE IIIII

MEAN AND GROWTH OF HDI AND ITS CONSTITUENT COMPONENTS FROM CLUSTERING RESULTS

Cluster		1	2	3	4	5
IPM	M	68,27	68,01	72,67	62,76	78,72
	G	0,92%	0,88%	0,75%	1,07%	0,63%
AHH	M	70,19	68,05	71,36	65,96	72,30
	G	0,22%	0,24%	0,19%	0,29%	0,17%
HLS	M	12,03	13,01	12,84	11,87	12,45
	G	1,54%	1,44%	1,37%	1,96%	1,00%

RLS	M	7,60	8,51	8,81	6,79	10,69
	G	1,32%	1,35%	1,11%	1,73%	0,72%
Income per Capita (000)	M	10.093	8.798,23	11.724	7.997	17.029
	G	2,38%	2,00%	1,79%	1,99%	2,30%

Note, M : Mean, G : Growth

The phenomenon that occurs from the results of clustering shows that the HDI of clusters that are relatively medium and low has a higher growth rate than clusters with high HDI. The same is seen in the AHH, HLS, and RLS components. The closer to the optimum position, the growth rate slows down. However, in terms of per capita income, the cluster growth rate is high in line with their status.

IV.CONCLUSION

From the results of the cluster analysis, 5 clusters were formed based on the HDI-forming index with characteristics that were homogeneous between members and heterogeneous between clusters. Cluster 1 province with a relatively moderate HDI, relatively high AHH, relatively moderate HLS, relatively low RLS, and relatively high per capita income. Cluster 2 provinces with relatively moderate HDI and AHH, relatively high HLS and RLS, and relatively low per capita income. Cluster 3 provinces with relatively high HDI, AHH, HLS, RLS and income per capita. Cluster 4 provinces with relatively low HDI, AHH, HLS, RLS and income per capita. Cluster 5 consists of only 1 province, namely DKI Jakarta.

From the research results, the clusters that need more attention from the government are cluster 4. The provinces of West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Gorontalo, Sulawesi Barat, West Papua, and Papua which are members of cluster 4

coincidentally are provinces in the eastern region. Indonesia. This shows the high disparity between the western and eastern regions. Human development still needs to be boosted, especially cluster 4, but that does not mean ignoring other clusters such as cluster 1 and cluster 2 whose HDI achievement is relatively moderate. Cluster 3 and cluster 5 are required to maintain HDI achievements, even increase them.

V. REFERENCES

- [1]. [BPS] Badan Pusat Statistik. 2020. Jakarta [ID]. Dynamic table – Indeks Pembangunan Manusia. [downloaded 23 July 2020]. From: <https://www.bps.go.id/subject/26/indeks-pembangunan-manusia.html#subjekViewTab3>
- [2]. Cahyani, Adella Sari. 2020. Comparison of Multiple Distances On Panel Data in Non-Hierarchical Clustering Method. *International Journal of Scientific and Research Publications*, Volume 10, Issue 7, July 2020
- [3]. Fidella, Rita. 2020. Pengaruh Belanja Pemerintah Daerah Terhadap Indeks Pembangunan Manusia [Thesis]. Bogor (ID) : IPB
- [4]. Genolini C, Alacoque X, Sentenac M, Arnaud C. 2015. KML and KML3D:R Packages to Cluster Longitudinal Data. *Journal of Statistical Software*. Volume 65, Issue 4.
- [5]. Genolini C, Ecohard R, Benghezal M, Arnaud C, et al. 2016. kmlShape:An Efficient Method to Cluster Longitudinal Data (Time-Series) According to Their Shapes.
- [6]. Mattjik, A. A., Sumertajaya, I., Wibawa, G. N. A., & Hadi, A. F. 2011. Sidik peubah ganda dengan menggunakan SAS. Department of Statistics, IPB.
- [7]. Mirza, Denni Sulistio. 2011. Pengaruh Kemiskinan, Pertumbuhan Ekonomi dan Belanja Modal terhadap IPM Jawa Tengah. *JEJAK, Jurnal Ekonomi dan Kebijakan Unnes*. Vol. 4 No. 2
- [8]. Musgrave, Richard, et al. 1989. *Public Finance : In Theory and Practice*. New York: McGraw-Hill Book Co.

Cite this article as :

Azlam Nas, Sri Mulatsih, Muhammad Findi , "Regional Clustering Based On HDI Components In Indonesia", *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 8 Issue 1, pp. 21-25, January-February 2021. Available at doi : <https://doi.org/10.32628/IJSRSET21813>
Journal URL : <http://ijsrset.com/IJSRSET21813>