Volume 8, Nomor 1, Juni 2023, 1-18
ISSN 2407-4292 (Print), ISSN 2721-6780 (Online)
Doi: https://doi.org/10.33701/j-3p.v8i1.3254
Available Online: http://ejournal.ipdn.ac.id/jpdpp

GROUPING PROVINCES IN INDONESIA BASED ON YOUTH DEVELOPMENT INDICATORS IN 2021

Muhamad Zidan Nuralifian*[1], Rani Nooraeni[2]
[1]Politeknik Statistika STIS, Jl. Otto Iskandardinata No.64C, RT.1/RW.4, Bidara Cina, Kecamatan Jatinegara 13330, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia; e-mail: zidan.zd14@gmail.com
[2]Politeknik Statistika STIS, Jl. Otto Iskandardinata No.64C, RT.1/RW.4, Bidara Cina, Kecamatan Jatinegara 13330, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia; e-mail: raninoor@stis.ac.id

*Correspondence
Received: 08-04-2023; Accepted: 27-06-2023; Published: 30-06-2023

Abstract: Youth have a vital role in development. Indonesia has an uneven distribution of youth across provinces, with over half concentrated in Java. Many young people are also only sometimes in line with their qualities. Youth development was observed through youth development indicators in 2021. The method used is a multivariate method using cluster analysis. The cluster method applied in this research is hierarchy and partition. Based on internal and stability validity, the hierarchical method for five clusters and the number of clusters is the best. The hierarchical method that has the most significant agglomeration coefficient is complete linkage. There is one province with indicators that are very different from other provinces: Papua as cluster 1. Papua requires massive development in all aspects. Cluster 2 comprises Riau Island, Jakarta, the Special Region of Yogyakarta, Bali, and East Kalimantan. Cluster 3 consists of West Nusa Tenggara, Bengkulu, and Lampung. Cluster 4 consists of West Java, Banten, Central Java, Gorontalo, South Sumatra, and East Java. Meanwhile, cluster 5 consists of the remaining members, with the remaining 19 provinces having the most members.

Keywords: *Youth Development Indicators, Cluster Analysis, Hierarchy, Partition*

Abstrak: Pemuda memiliki peran yang sangat penting bagi pembangunan. Jumlah pemuda di Indonesia tidak merata antarprovinsi dengan lebih dari separuhnya terkonsentrasi di Pulau Jawa. Jumlah pemuda yang banyak tersebut juga tidak selalu selaras dengan kualitasnya. Pembangunan pemuda diamati melalui indikator pembangunan pemuda tahun 2021. Metode yang digunakan adalah secara multivariat menggunakan analisis klaster. Metode klaster yang diterapkan dalam penelitian ini adalah hierarki dan partisi. Berdasarkan validitas internal dan stabilitas, metode dan jumlah klaster terbaik adalah dengan menggunakan metode hierarki untuk dibentuk sebanyak lima klaster. Metode hierarki yang memiliki koefisien aglomerasi terbesar adalah dengan complete linkage. Terdapat satu provinsi yang memiliki capaian indikator yang sangat berbeda dengan provinsi-provinsi lainnya, yaitu Provinsi Papua sebagai klaster 1. Provinsi Papua memerlukan pembangunan yang masif secara keseluruhan. Klaster 2 terdiri atas Kep. Riau, DKI Jakarta, DI Yogyakarta, Bali, dan Kalimantan Timur. Klaster 3 terdiri atas Nusa Tenggara Barat, Bengkulu, dan Lampung. Klaster 4 terdiri atas Jawa Barat, Banten, Jawa Tengah, Gorontalo, Sumatera Selatan, dan Jawa Timur. Sementara itu, klaster 5 beranggotakan sisanya dengan merupakan anggota yang terbanyak hingga 19 provinsi sisanya.

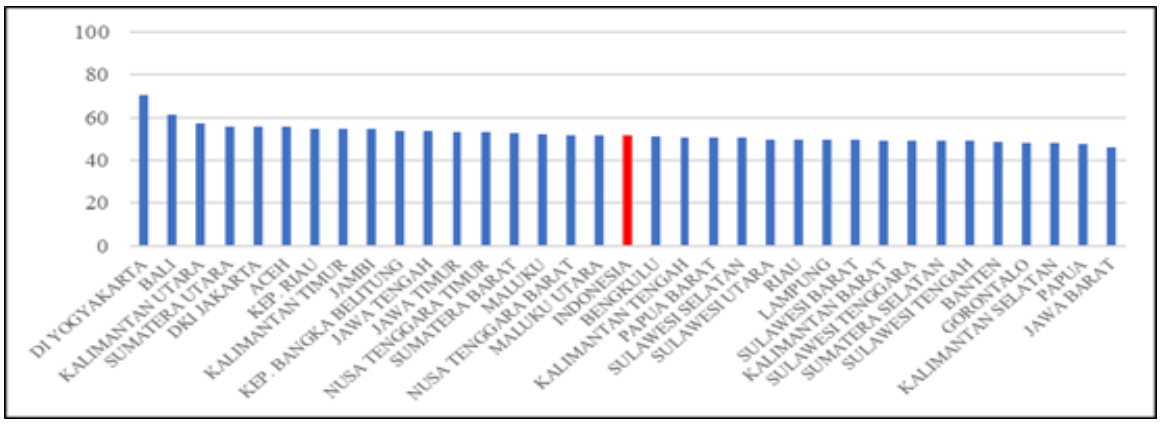
Kata Kunci: Indikator Pembangunan Pemuda, Analisis Klaster, Hierarki, Partisi

I. Introduction

Creating a resilient, independent, and competitive next generation is one of the strategic development priorities. Youth play a critical role in development. According to Law No. 40/2009 on Youth, youth ages 16 to 30. According to the United Nations (2010), youth in all countries are critical human resources for development and vital agents of social change, economic development, and technical innovation.

In 2021, the youth population in Indonesia was around 64.92 million, or 23.9% of the total population (Central Bureau of Statistics, 2021). Such a large number emphasizes the demographic advantage when the productive age population dominates the population structure. As part of the productive-age population, youngsters must be strategic in accepting positions and opportunities. Indonesia's ability to capitalize on the demographic dividend may be threatened if youth cannot raise their value in development.

According to Salim (2018), the various actions and roles of youth in various historical periods strengthen the existence of youth in creating change. The changes that occur in this country place youth as the main actors. Investment in enhancing youth quality is required. As the nation's hopeful individuals, youth should prepare to reach their full potential and contribute to its development.



Source: Ministry of National Development Planning/Bappenas

Figure 1. The Youth Development Index (YDI) of provinces in Indonesia in 2018

The Youth Development Index (YDI) by the Ministry of National Development Planning/Bappenas (2020) has increased over the past three years. In 2018, the YDI rate in Indonesia was 51.5 percent, an increase from 49.33 percent in 2017. Based on Figure 1, the range of YDI values at the provincial level is between 46.17 percent and 70.33 percent. West Java has the lowest YDI, whereas DI Yogyakarta has the highest.

According to Efendi (2020), there is a positive or unidirectional association between the YDI and the Human Development Index (HDI). The higher the YDI, the higher the HDI. The YDI captures components of human resource development that are more extensive and particular to the youth age group than the HDI. However, according to the YDI, youth development achievements have not been equally spread across Indonesia.

The distribution of population and youth in Indonesia is uneven between provinces. Uneven distribution creates a gap between the center and the regions. Poverty, unemployment, environmental degradation, pollution, increased crime, and the rise of slums in urban areas are some of the consequences of unequal population distribution (Falikhah, 2017).

Each region's distribution of youth does not necessarily correspond to their quality. Youth must excel in quantity, quality, and competitiveness—these are demands in each region. Each province must be ready for youth development, yet the diversity of Indonesia's regions generates variances in youth development. Therefore, it is necessary to group provinces in Indonesia based on youth development indicators in 2021.

II. Literature Review

In Nurkholis (2018), human development should focus not only on efforts to improve human abilities or capabilities but also on efforts to maximize the utilization of these human abilities. In Indonesia, economic development inequality remains high, and interprovincial inequality is moderate to high (Azim et al., 2022). One of the causes is the disparity in the amount and quality of youth in Indonesian provinces. As the primary development pillar, youth development can only function optimally in some locations.

Bassani (2007) stated the dimensions of social capital theory in the study of youth. The dimensions of social capital include various forms of capital that affect welfare. Social capital is vital and has a good association with welfare. Social resources are transformed into social capital, creating a complex process. Then, there is an interaction between the social capital possessed, resulting in the influence of youth-owned social capital. Education, health, and work can all provide social capital.

Youth are one of the population groups with particular characteristics that demand a different approach to assure their quality of life, according to the 2020–2024 National Medium-Term Development Plan (National RPJMN) (Bappenas, 2019). Increasing the equal distribution of quality education services is one of the development policy methods

implemented. Education is a critical component in the process of sustainable development, particularly when it comes to people (Simanjuntak, 2017).

Youth who can build the country must also be healthy and prosperous. Every human being's health is an investment, a right, and an obligation. Youth have the finest physical health compared to other age groups (Isfandari & Lolong, 2014).

Youth are part of the working-age population expected to engage in productive activities. The huge working-age population can be absorbed as human capital in the labor market, resulting in economic growth (Dewi et al., 2018). In other words, youth production aids economic growth.

Gender equality and reproductive health are other vital issues among youth that cannot be ignored. Challenges such as early marriage and unwanted pregnancies can have current and future implications for youth health, education, career opportunities, and empowerment. Youth, particularly teenagers, have the right to accurate and complete reproductive health information (Permatasari & Suprayitno, 2021).

Clustering provinces in Indonesia are one technique to observe the achievement of youth development indicators. Multivariate cluster analysis is one way that can be used. Hierarchical and partitioning cluster approaches were used in this study.

III. Research Methodologies

In 2018, indicators based on multiple YDI components and other data were used to assess the role of youth as both subject and object in development. Individual growth, livelihood and welfare development, and engagement in numerous sectors of life are all mentioned in the planning process of the 2018 YDI. The 2021 youth development indicators complement it more comprehensively to capture broader cross-cutting aspects of youth development. Youth development indicators include education, health and welfare, employment, and gender and reproductive health. This study categorizes Indonesian provinces based on youth development metrics. The approach taken is cluster analysis of hierarchical and partition methods. The data is secondary data collected from the Central Bureau of Statistics (BPS) publication 2021 Youth Statistics. This study's unit of analysis is Indonesia's 34 provinces. Table 1 shows the variables that were used.

Table 1. Variables List

Dimension	Indicator	
Education	X11	Average Year of Schooling
	X12	Percentage of Youth with the Last Education Completed at Higher Education
	X13	Percentage of Youth who Used the Internet in the Last Month
Health and Welfare	X21	Youth Morbidity Rate
	X22	Percentage of Youth Who Have Smoked in the Last Month
	X23	Percentage of Youth Who Have Been Victims of Crime in the Last Year
	X24	Percentage of Youth Living in an Adequate House
Employment	X31	Youth Open Unemployment Rate
	X32	Percentage of Youth with White Collar Primary Job Type
Gender and Reproductive Health	X41	Percentage of Female Youth Working in the Formal Sector
	X42	Percentage of Females Who Have/Are Using Contraception/Traditional Methods
	X43	Percentage of Youth with Marriage Age < 19 Years

Analysis Methods

Factor analysis is a technique used to simplify data with multiple variables. It is only necessary when the correlation between variables is higher than 0.8. In such cases, even a small sample size can be used for factor analysis, per Williams et al. (2010). However, factor analysis is not required if there is no correlation value exceeding 0.8.

The analysis in this study used cluster analysis. According to Han et al. (2022), cluster analysis divides data objects (or observations) into sets of portions known as clusters. Cluster analysis is a natural grouping technique based on indicators of similarity or difference between people (Williams et al., 2010).

In this study, cluster validation is used to analyze the data. It helps determine the ideal method and number of clusters to use. This process involves evaluating the homogeneity and separation of individuals within each cluster (Liu et al., 2010).

Cluster Analysis

Cluster analysis analyzes a data set to identify patterns and group data into clusters. Members within a cluster are homogeneous, while members between clusters are heterogeneous. There are two clustering methods: hierarchical or partitioning, as explained

by Madhulatha (2012). The flow of clustering using cluster analysis is illustrated in Figure 2.

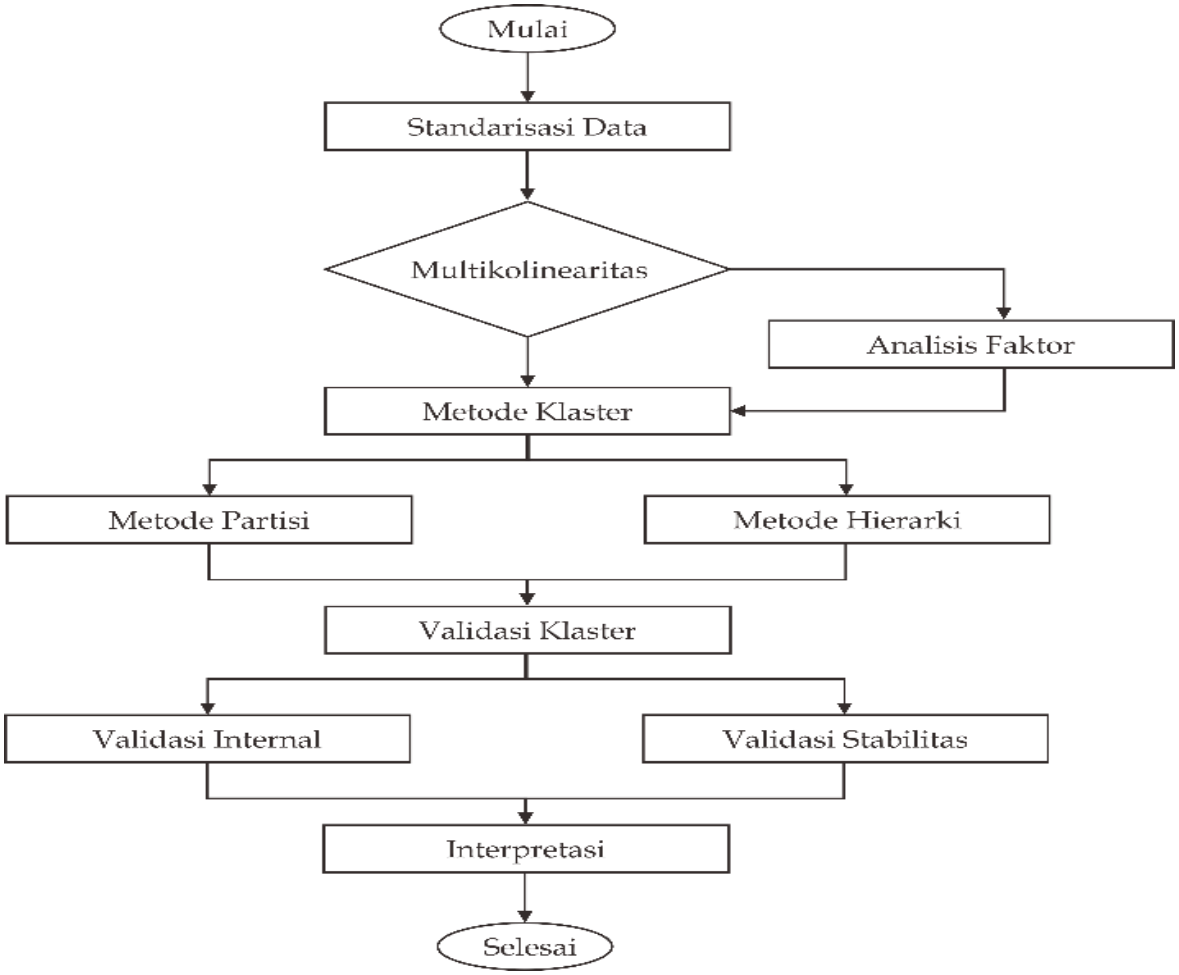


Figure 2. Cluster Analysis Flowchart in the Research

Cluster analysis requires a similarity measure in the process. The Euclidean distance formula is the most widely recognized distance measure in the case of numerical data. Here is the formula for Euclidean distance.

$$d = \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2}$$

Hierarchy Method

Like a tree structure, the hierarchical technique creates clusters in stages and levels (Ramadhani et al., 2018). This approach creates clusters by repeating partitions both top-down and bottom-up. The agglomerative approach is one form of hierarchical method. At an early stage, agglomerative approaches analyze each individual as a cluster and then combine based on a measure of similarity (Rokach & Maimon, 2005). In hierarchical

approaches, there are various merging methods, including single linkage, complete linkage, average linkage, and Ward's method.

Single linkage uses individuals in clusters closest to other individuals to consider merging. The following is an example formulation to calculate the distance of cluster *ij* to individual *k*.

$$d_{(ij)k} = \min\{d_{ik}, d_{jk}\}$$

Meanwhile, complete linkage uses individuals in the cluster with the furthest distance.

$$d_{(ij)k} = \max\{d_{ik}, d_{jk}\}$$

Another method is average linkage. This technique considers the average distance between cluster members and other individuals.

$$d_{(ij)k} = \frac{d_{ik} + d_{jk}}{2}$$

In addition, Ward's method groups objects by minimizing the sum of square errors in the cluster.

$$SSE = \sum_{i=1}^{n} (\boldsymbol{x_i} - \overline{\boldsymbol{x}})'(\boldsymbol{x_i} - \overline{\boldsymbol{x}})$$

The results and clustering process of the hierarchical method can be visualized in a dendrogram.

Partitioning Method

The partitioning method allocates individuals by moving them from one cluster to another. Partition-based approaches partition *n* data objects into *k* groups. Among the most common partition-based algorithms are the K-Means and K-Medoids approaches, which use Euclidean space as their similarity measure (Mousavi et al., 2020).

The most basic and extensively used partitioning algorithm is K-Means. K-Means has a straightforward algorithm. The data is divided into *k* predetermined clusters (Kodinariya & Makwana, 2013). K-Medoids, which is related to K-means, are another method. K-Medoids uses the centermost observation as the cluster center. This method outperforms K-Means in terms of robustness but not efficiency (Rokach & Maimon, 2005).

Internal Validity

Internal validation is done with the connectivity, silhouette, and dunn indexes. The smaller the connectivity index value, the better the cluster formation (Machfudoh & Wahyuningsih, 2013).

$$Conn(C) = \sum_{i=1}^{N} \sum_{j=1}^{L} X_{i,nni(j)}$$

$nn_{i(j)}$: nearest neighbor observation from *i* to *j*

L: A parameter that determines the number of contributing neighbors.

In addition, validation can use the silhouette index. The best cluster is the one with a silhouette index closest to one (Machfudoh & Wahyuningsih, 2013).

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)}$$

$a(i)$: The average distance between *i* and all other observations in the same cluster.

$b(i)$: The average distance between *i* and the nearest cluster observation.

Another validation is the Dunn index. The most extensive Dunn index indicates that clusters are separated and better (Machfudoh & Wahyuningsih, 2013).

$$C = \frac{d_{min}}{d_{max}}$$

d_{min}: The smallest distance between observations from different clusters.

d_{max}: The largest distance between observations in the same cluster.

Stability Validity

A cluster stability metric analyzes a cluster's consistency when the variables in the cluster are removed one by one. The average proportion of non-overlap (APN), the average distance (AD), the average distance between means (ADM), and the Figure of Merit (FOM) are cluster stability indicators.

APN measures the average proportion of observations not in the same cluster when included in a cluster analysis with a complete dataset and a dataset with one of its variables removed. The APN value ranges from 0 to 1. The cluster created is more consistent when the APN is near zero (Syah, 2019).

AD calculates the average distance between observations in the same cluster when included in a cluster analysis with the entire dataset and one of the variables removed. The range of AD values is 0 to infinity. The closer AD is to 0, the more consistent the cluster formed (Syah, 2019).

ADM calculates the average distance between cluster centers when performing cluster analysis with complete and incomplete data sets. The ADM value ranges from 0 to 1. The closer ADM reaches zero, the more consistent the cluster becomes (Syah, 2019).

FOM measures the average intracluster variance of variables not used in cluster analysis. The FOM value is in the range of 0 to 1. The closer the FOM is to 0, the more consistent the cluster formed (Syah, 2019).

IV. Result And Discussion

Descriptive Statistic

The frequency distribution pattern for each variable used is shown below.

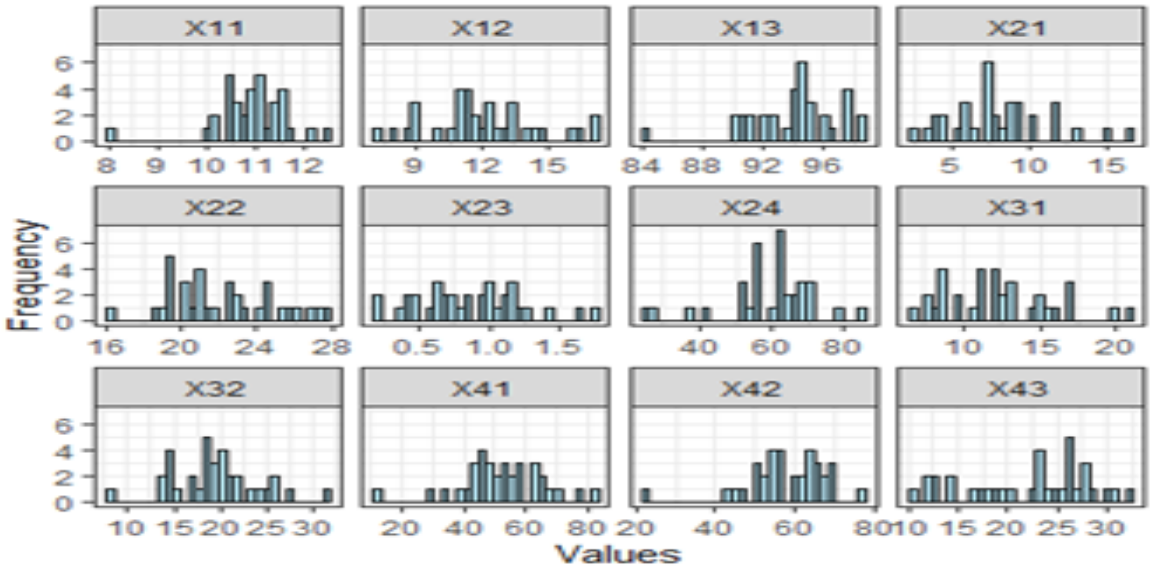


Figure 3. Frequency distribution of each indicator.

According to Figure 3, some variables have significant variety and outlier values. Variables X13 and X21 contain regions that differ from others in appearance. Due to unequal development, the distribution reflects the variability of youth development successes throughout Indonesian provinces. In the meantime, there are outliers, such as X11 and X42. Outliers demonstrate that some regions have considerably different youth development achievements than others. Each variable's distribution pattern is employed.

Papua Province, once determined, has a lower level of youth development than other provinces. In terms of education, the average number of years of schooling in Papua is merely 8.07 years. Papua province also has poor welfare outcomes, with only 26.26% of the population living in adequate housing, when all other provinces have reached more than 50%.

Correlation Testing

Before conducting cluster analysis, it is necessary to determine whether data reduction from the variables employed is required. The correlation coefficient between variables determines whether or not to reduce the variables. Variable reduction is related to

multicollinearity testing. The correlation values of the variables in the study are shown below.

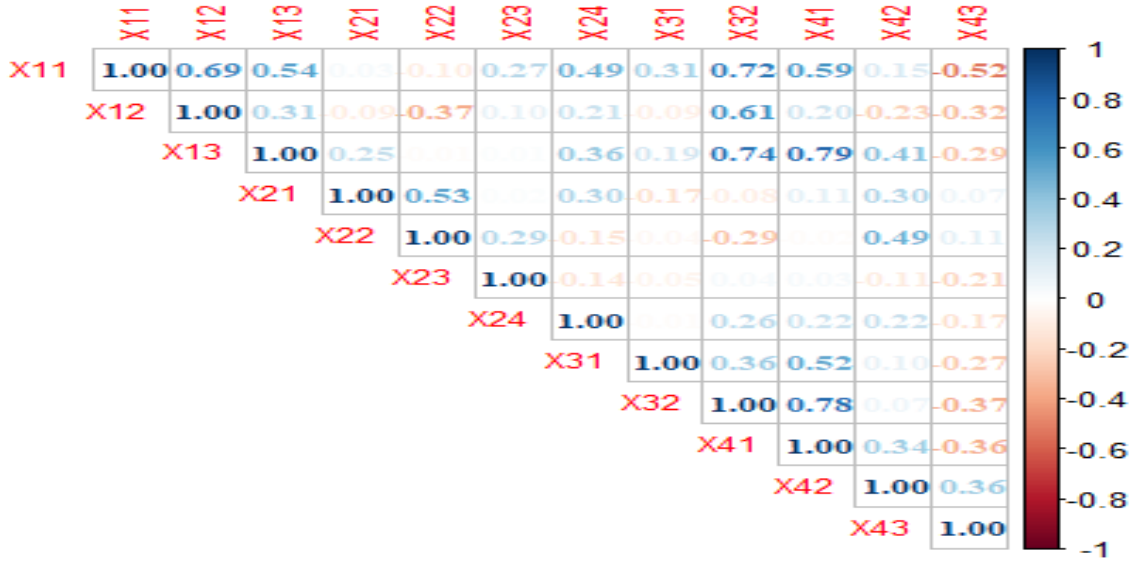


Figure 4. Inter-variable Correlation Plot

A correlation value between variables greater than 0.8 indicates the presence of multicollinearity (Shrestha, 2020). Figure 4 reveals that no variables in the study have a strong association greater than 0.8. As a result of the lack of multicollinearity, factor analysis is optional to decrease the data.

Cluster Analysis

The hierarchical approach and the K-Means and K-Medoids partitioning methods were used to do cluster analysis. Internal validity and stability evaluation criteria were used to determine the optimal method and number of clusters. The Connectivity, Silhouette, and Dunn indices were utilized as internal validity criteria. Meanwhile, the APN, AD, ADM, and FOM are the stability validity requirements.

Cluster Validation

Because there are outliers in the data, the number of two clusters does not yield relevant findings. The first cluster will be Papua province, identified as an outlier, while the second cluster will include other provinces. Observing the diversity of development achievements in provinces other than Papua is impossible. As a result, the number of clusters of two is not considered when finding the ideal number of clusters. The following shows the internal validity criteria for each method according to the number of clusters.

Table 2. Internal Validity Index Values

Cluster	*Connectivity*	*Dunn*	*Silhouette*
K-Means Method			
3	14,2544	0,3399	0,2418
4	15,9544	0,3399	0,2114
5	32,1909	0,2865	0,1584
6	41,1103	0,326	0,147
7	43,8179	0,326	0,168
K-Medoids Method			
3	19,7778	0,2845	0,1678
4	23,6179	0,2845	0,1638
5	35,3683	0,2845	0,1149
6	34,5627	0,2845	0,1423
7	37,8988	0,3419	0,1697
Hierarchy Method			
3	11,0198	0,4315	0,2791
4	11,3448	0,4315	0,2713
5	14,2738	0,5557	0,2072
6	18,8929	0,4996	0,1726
7	27,0627	0,3671	0,1607

Based on Table 2, the hierarchical technique is the best strategy for clustering youth development indicators in 2021. According to the Connectivity and Dunn indices, the ideal number of clusters produced is three, while the Silhouette index recommends five. Meanwhile, the outcomes utilizing the stability validity criterion are as follows.

Table 3. Stability Validity Index Value

Cluster	APN	AD	ADM	FOM
K-Means Method				
3	0,1191	3,7082	0,4853	0,8705
4	0,3130	3,6385	1,1141	0,8475
5	0,1384	3,1933	0,6120	0,8446
6	0,1836	3,0587	0,8797	0,8178
7	0,2350	2,9376	1,1388	0,7891
K-Medoids Method				
3	0,1075	3,6754	0,4653	0,8648
4	0,2758	3,5655	0,9879	0,8603
5	0,2655	3,3770	1,0900	0,8508
6	0,2720	3,0973	1,0036	0,8043
7	0,2703	2,8682	0,9295	0,7666

Hierarchy Method				
3	0,0531	3,8212	0,5000	0,8942
4	0,0559	3,6076	0,4230	0,8602
5	0,0448	3,3024	0,2312	0,8101
6	0,0613	3,1258	0,3537	0,8040
7	0,1013	2,9987	0,8117	0,7902

According to the stability validity criterion results in Table 3, the optimum technique for clustering youth development indicators in 2021 combines the hierarchical method and K-Medoids. According to the AD and FOM values, the optimal approach and number of clusters are as many as seven with K-Medoids. Furthermore, the best approach and number of clusters based on the APN and ADM values are hierarchical, with five clusters.

Table 4. Stability Validity Indec Value

Criteria	Value	Method	Number of Cluster
Internal Validity			
Connectivity	11,0198	Hierarchy	3
Dunn	0,5557	Hierarchy	5
Silhouette	0,2791	Hierarchy	3
Validity Stability			
APN	0,0448	Hierarchy	5
AD	2,8662	*K-Medoids*	7
ADM	0,2312	Hierarchy	5
FOM	0,7666	*K-Medoids*	7

Table 4 displays the best approach and the number of clusters for each validation criterion. Based on these findings, the hierarchical technique produced the best grouping, with up to five clusters. Clustering using the hierarchical technique has numerous methods, including single linkage, complete linkage, average linkage, and Ward's method.

Cluster Result

The agglomeration coefficient determines the type of distance or linkage used. The agglomeration coefficient from the cluster results utilizing single linkage, complete linkage, average linkage, and Ward's technique is shown below.

Tabel 5. Agglomeration Coefficient

Method	Agglomeration Coefficient
Single Linkage	0,5977
Average Linkage	0,6787
Complete Linkage	0,7557
Ward's Method	0,7462

According to Table 5, the entire connection technique has the highest agglomeration coefficient. As a result, in 2021, the grouping of youth development indicators is done hierarchically utilizing the entire linkage approach, with up to five clusters.

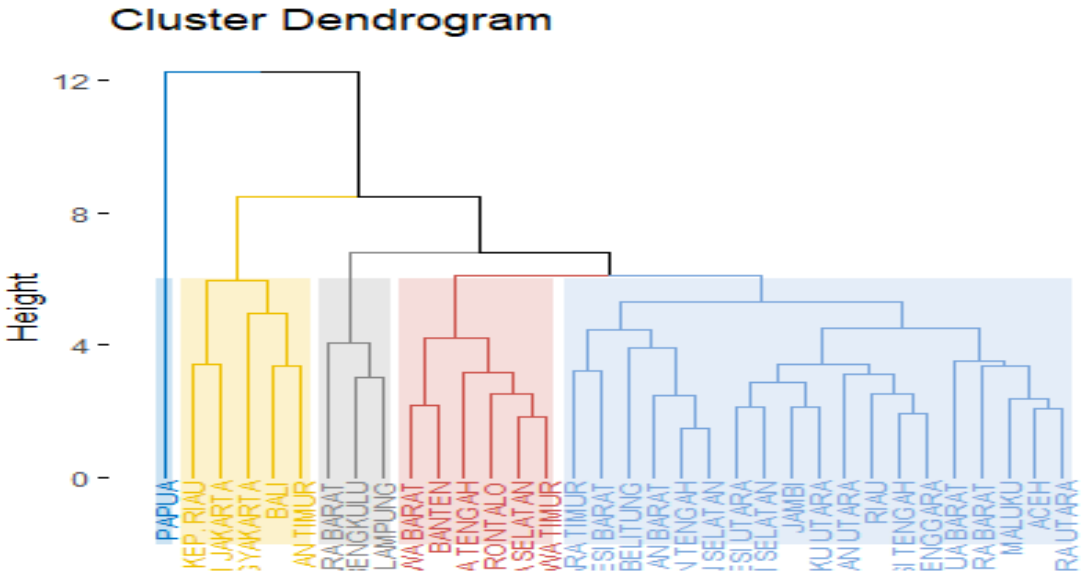


Figure 5. Dendrogram

Figure 5 depicts the members of each cluster. The first cluster has only one province, Papua, whereas the fifth cluster has the most members, with 19 members. The members of each created cluster are listed below.

Table 6. Members of Each Cluster

Cluster	Members of Cluster
1	Papua
2	Riau Island, Jakarta, Special Region of Yogyakarta, Bali, and East Kalimantan
3	West Nusa Tenggara, Bengkulu, and Lampung
4	West Java, Banten, Central Java, Gorontalo, South Sumatera, and East Java
5	East Nusa Tenggara, West Sulawesi, Bangka Belitung, West Kalimantan, Central Kalimantan, South Kalimantan, North Sulawesi, South Sulawesi, Jambi, North Maluku, North Kalimantan, Riau, Central Sulawesi, South East Sulawesi, West Papua, West Sumatera, Maluku, Aceh, and North Sumatera

Table 6 shows the distribution of attainment of youth development indicators in 2021. Papua is a province in cluster 1 that is an outlier. Aside from Papua Province, the other provinces' results in the surrounding areas could be more diverse. Except for East Kalimantan and Gorontalo, the provinces of Kalimantan and Sulawesi are commonly grouped as cluster 5. The achievement of youth development indicators on Java Island differs from provinces outside of Java Island and is equally distributed within it. According to Hidayat (2021), this is one of the indirect causes of stagnation in other regions due to Indonesia's structural problems, which tend to be Java-centric. According to Wilonoyudho (2009), the fundamental issue that must be addressed is a long-term vision based on social justice and a political-economic system that upholds specific regions.

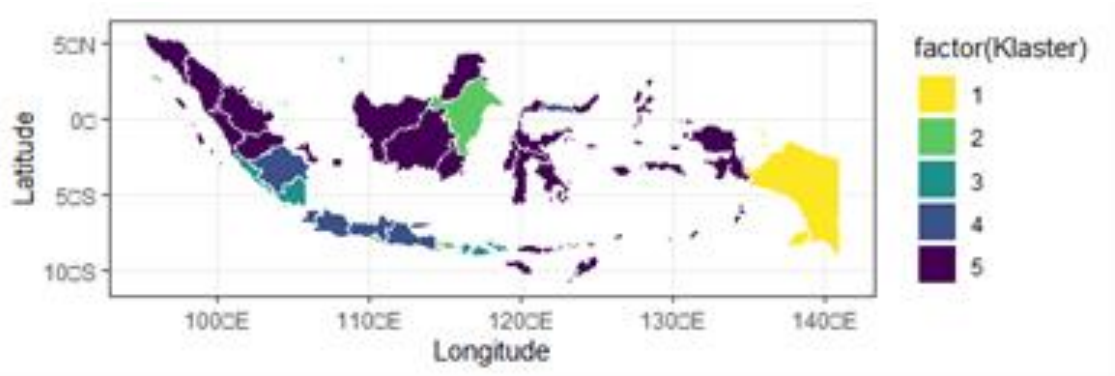


Figure 6 Map by Grouping of Youth Development Indicators in 2021

Papua is a province that has shown signs of youth development issues. Figure 6 depicts how the province's development achievements differ from other provinces. The existing disparities show that the youth situation in Papua is exceedingly unfair and appears to be ignored. Regional disparities in a variety of variables mainly cause inequality. As a result, the ability of a region to drive its regional development process is also different (Hartati, 2022).

Raafi'i et al. (2018) discovered that no Papua region is typologically dominant as developed or underdeveloped. Based on statements from previous researchers, the issue in Papua appears pervasive. Budget allocations for education and health in Papua affect poverty, intimately tied to human resource development (Sofilda, 2019). Papua Province needs more infrastructure for youth development, such as education and health care. The average of each indicator and its comparison for the five clusters produced are shown below to show which areas of development could have been better or more adequate.

Table 7. Average Indicator by Cluster

Indicator	Cluster				
	1	2	3	4	5
X11	8,07	11,81	10,93	10,66	10,9
X12	7,15	15,08	10,73	9,72	12,42
X13	84,36	97,99	93,63	95,05	93,27
X21	4,13	7,49	10,53	12,19	6,75
X22	19,5	19,55	27,36	24,87	21,27
X23	0,66	0,82	1,33	0,74	0,84
X24	26,26	64,58	58,88	61,91	58,64
X31	7,51	13,04	9,24	14,6	12,26
X32	7,79	27,01	16,25	18,92	18,83
X41	12,32	69,04	42,26	61	49,02
X42	22,15	52,82	66,55	64,14	57,9
X43	25,4	14,13	23,38	22,95	23,66

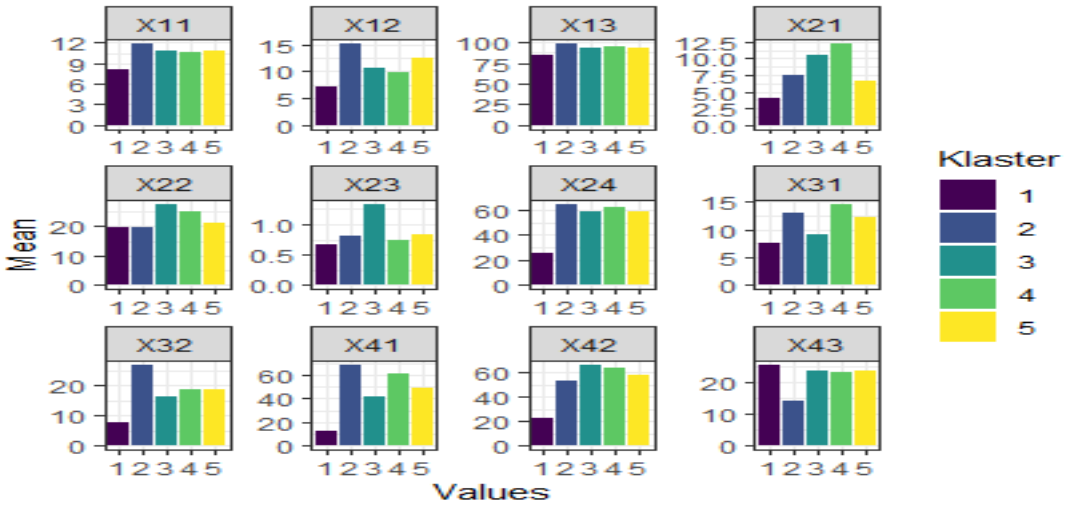


Figure 7. Average Indicator by Cluster

Table 7 and Figure 7 show the average differences between clusters. Cluster 1, or Papua Province, consistently has the lowest outcomes compared to the other clusters. This means that positive indices like average years of schooling, the percentage of young people living in adequate housing, and the percentage of youth working in white-collar jobs remain low. However, some negative indicators, such as youth morbidity and open unemployment rates, are beneficial because of their low values. Youth development in Papua necessitates extensive overall growth in all areas.

Cluster 2 has made significant progress in education, health, and welfare. The results for employment, gender, and reproductive health are likewise quite favorable. Unemployment and family planning engagement are two youth development factors that

might be addressed further for Cluster 2. Meanwhile, Cluster 3 does well on school indicators, but other indicators, such as early smoking, youth employment in white-collar jobs, and early marriage, require improvement.

Cluster 4 showed the most outstanding results in the metrics of youth morbidity rate and youth open unemployment rate. These should be the primary considerations in cluster 4 for promoting youth development. Cluster 5, on the other hand, has the lowest rate of youth morbidity. Cluster 5 is the outcome of the cluster with the most members, which yields average indicator values in most cases. Gender and employment are two indicators that can be addressed for improvement in Cluster 5.

V. Conclusion

Factor analysis is unnecessary for the 2021 youth development indicators because there is no tangible link between variables. Direct application of cluster analysis is possible. The cluster approach used is complete hierarchical linkage, with five clusters formed.

As cluster 1, Papua Province has significantly different indicator achievements than the other provinces. Riau Islands, DKI Jakarta, DI Yogyakarta, Bali, and East Kalimantan make up Cluster 2. West Nusa Tenggara, Bengkulu, and Lampung make up Cluster 3. West Java, Banten, Central Java, Gorontalo, South Sumatra, and East Java comprise Cluster 4. Cluster 5 comprises the remaining 19 provinces with the most members.

Suggestions include the government and related parties focusing on development more equally across Indonesian provinces, particularly in Papua Province. Because
the role of youth is critical to development, equitable development should be the primary focus. The government can boost youth development by first prioritizing aspects of significant concerns based on the formation of groups. Furthermore, additional researchers are encouraged to examine the diversity in Papua Province and its environs.

VI. Bibliography

Azim, A. N., Sutjipto, H., & Ginanjar, R. A. F. (2022). Determinan Ketimpangan Pembangunan Ekonomi Antarprovinsi di Indonesia. *Jurnal Riset Ilmu Ekonomi*, *2*(1), 1-16.

Badan Perencanaan Pembangunan Nasional. (2020). Rencana Pembangunan Jangka Menengah Nasional 2020-2024. Jakarta: Badan Perencanaan Pembangunan Nasional.

Badan Pusat Statistik. (2021). Statistik Pemuda 2021. Jakarta: Badan Pusat Statistik.

Bassani, C. (2007). Five Dimensions of Social Capital Theory as They Pertain to Youth Studies. *Journal of youth studies*, *10*(1), 17-34.

Dewi, S., Listyowati, D., & Napitupulu, B. E. (2018). Bonus Demografi di Indonesia: Suatu Anugerah atau Petaka. Journal of Information System, Applied, Management, Accounting and Research, 2(3), 17-23.

Efendi, A. (2020). Studi Korelasi Indeks Pembangunan Pemuda (IPP) dengan Indeks Pembangunan Manusia (IPM) di Indonesia. *Jurnal Paradigma (JP)*, *9*(1), 3948.

Falikhah, N. (2017). Bonus Demografi Peluang dan Tantangan bagi Indonesia. *Alhadharah: Jurnal Ilmu Dakwah*, *16*(32).

Han, J., Pei, J., & Tong, H. (2022). Data Mining: Concepts and Techniques. Morgan kaufmann.

Hartati, Y. S. (2022). Analisis Ketimpangan Ekonomi di Provinsi Papua. Jurnal Ekonomi dan Bisnis, 14(2), 19-29.

Hidayat, A. R. (2021). Pembangunan Dan Ketimpangan: Pandangan Keadilan Sosial. Journal of Humanity and Social Justice, 3(2), 145-149.

Kementerian Perencanaan Pembangunan Nasional (PPN)/Badan Perencanaan Pembangunan Nasional. (2020). Indeks Pembangunan Pemuda 2019. Jakarta: Kementerian PPN/Bappenas.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on Determining Number of Cluster in K-Means Clustering. International Journal, 1(6), 90-95.

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010, December). Understanding of Internal Clustering Validation Measures. In 2010 IEEE international conference on data mining (pp. 911-916). IEEE.

Lolong, D. B., & Isfandari, S. (2014). Analisa Faktor Risiko dan Status Kesehatan Remaja Indonesia pada Dekade Mendatang. Indonesian Bulletin of Health Research, 42(2).

Machfudhoh, S., & Wahyuningsih, N. (2013). Analisis Cluster Kabupaten/Kota Berdasarkan Pertumbuhan Ekonomi Jawa Timur. Jurnal Sains dan Seni POMITS, 2(1), 2337-3520.

Madhulatha, T. S. (2012). An Overview on Clustering Methods. arXiv preprint arXiv:1205.1117.

Mousavi, S. H. A. H. L. A., Boroujeni, F. Z., & Aryanmehr, S. A. E. E. D. (2020). Improving customer clustering by optimal selection of cluster centroids in k-means and k-medoids algorithms. Journal of Theoretical and Applied Information Technology, 98(18).

Nurkholis, A. (2018). Teori Pembangunan Sumber Daya Manusia: Human Capital Theory, Human Investment Theory, Human Development Theory, Sustainable Development Theory, People Centered Development Theory.

Permatasari, D., & Suprayitno, E. (2021). Pendidikan Kesehatan Reproduksi pada Remaja. Jurnal Empathy Pengabdian Kepada Masyarakat, 8-12.20079.

Raafi'i, A., Hakim, D. B., & Putri, E. I. K. (2018). Ketimpangan Pembangunan Antarwilayah Pengembangan di Provinsi Papua Barat. *Journal of Regional and Rural Development Planning (Jurnal Perencanaan Pembangunan Wilayah dan Perdesaan)*, *2*(3), 244-257.

Ramadhani, L., Purnamasari, I., & Amijaya, F. D. T. (2018). Penerapan Metode Complete Linkage dan Metode Hierarchical Clustering Multiscale Bootstrap. Eksponensial, 9(1), 1-10.

RI (Republik Indonesia). (2009). Undang-Undang No. 40 Tahun 2009 tentang Kepemudaan. Lembaga Negara RI Tahun 2009.

Rokach, L., & Maimon, O. (2005). Clustering Methods. In Data Mining and Knowledge Discovery Handbook (pp. 321-352). Springer, Boston, MA.

Salim, A. (2018). Pendidikan Karakter dan Eksistensi Pemuda. *Potret Pemikiran*, *19*(2).

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. American Journal of Applied Mathematics and Statistics, 8(2), 39-42.

Simanjuntak, F. N. (2017). Pendidikan untuk Pembangunan Berkelanjutan. Jurnal Dinamika Pendidikan, 10(2), 169-195.

Sofilda, E. (2019). Kualitas Pembangunan Manusia di Provinsi Papua. Media Ekonomi, 27(2), 171-182.

Syah, R. H. (2019). Penerapan Metode K-Means dalam Pengelompokan Kabupaten/Kota di Jawa Tengah. Media Edukasi Data Ilmiah dan Analisis (MEDIAN), 2(01), 44-58.

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory Factor Analysis: A Five-Step Guide for Novices. Australasian journal of paramedicine, 8(3).

Wilonoyudho, S. (2009). Kesenjangan dalam Pembangunan Kewilayahan.