

Multidimensional Poverty Clustering using K-Means Algorithm with Dimensional Reduction by Principal Component Analysis

Admi Salma¹ and Zilrahmi¹

¹ Statistics Department, Universitas Negeri Padang

Abstract- The level of Multidimensional poverty in each province in Indonesia varies, similar policies is ineffective to reduce the poverty. Several poverty indicators also influence other factors. General policies established to overcome poverty have proven ineffective, making it urgent to identify the needs of each province in overcoming this condition. Grouping provinces based on similar multidimensional poverty which use cluster analysis, will help address this situation. The aim of this study is to group provinces based on multidimensional poverty indicators using the k-means clustering method. Principal Component Analysis (PCA) was also used to reduce variables and multicollinearity. The clustering results showed seven clusters. The highest multidimensional poverty was found in cluster 2, which consisted of one province, namely Papua Pegunungan. This province shows deficiencies in education, health, and living standards compared to other clusters. Meanwhile, the lowest multidimensional poverty was found in cluster 7. There are three provinces in this cluster, namely Bali, Jakarta, and DIY Jogjakarta. These provinces experience minimal multidimensional poverty which is able to provide a better quality of life. The policies and development strategies in these provinces could serve as role models to develop other provinces based on their specific deficiencies and needs. Each cluster is well separated, as Davies Bouldin Index (DB) is lower, at 0.4.

Copyright ©2020 The Authors. Published by Rangkiang Mathematics Journal which permits unrestricted use, provided the original author and source are credited

1. Introduction

Poverty remains a critical issue and pressing challenge in many countries, including Indonesia. Poverty is a multidimensional issue. It is not only from an economic perspective but also considered in terms of social, cultural, and political aspects (Suhendar et al., 2024). Structural poverty includes a lack of education, public

facilities, health care, and job opportunities (Darma et al., 2024; Hasan, 2021).

As a central focus of the Sustainable Development Goals (SDGs), Indonesia's part in eradicating poverty represents a global commitment to ending it in all its forms, everywhere. Indonesia is an archipelagic country characterized by diverse geographical and economic features. However, this diversity leads to vastly varying socio-economic conditions. Economic inequality is a reality in Indonesia (Edison & Andriansyah, 2023).

Statistics Indonesia (BPS) reported that Indonesian's poverty as of September 2024 was 8,57% or around 24, 06 million people. Three provinces with high Poverty rate are Papua Pegunungan, with poverty rate of 29,66%, Papua Tengah at 27,6% and Papua Barat at 21,09%. In contrast, the lowest province poverty rate was Bali, at 3,8%. The data shows that Indonesia has a huge disparity in poverty. Some provinces, such as Bali, have a very low poverty rate, while others continue to face poverty issues. This disparity can be seen in various aspects of life such as education, income, healthcare, and public services. Economic inequality can hinder economic growth itself, potentially leading to higher crime rates, social discontent, and political instability (Sastiono & Nuryakin, 2019). Several factors influence poverty in Indonesia are average length of education, government spending on infrastructure, GDP, and life expectancy. Policies to promote economic growth, education, and health and government spending on infrastructure positively affect the declining number of poor people (Asrol & Ahmad, 2018; Faharuddin & Endrawati, 2022).

Even though some policies to overcome poverty have been stipulated, poverty still remains problematic in certain areas, especially rural regions. However, poverty elements such as education, employment, and health need to be enhanced (Idrus & Rosida, 2020). Therefore, to effectively handle poverty, the government needs policies that are fit for specific provincial conditions. A uniform policy will ineffectively address the poverty issue in Indonesia.

Previous research by Edna Safitri et al. (2022) utilized regression analysis to identify factors affecting poverty within a single province. This study has limitations, which is geographically limited to a single province. The result only focuses on one province. Meanwhile, other research by (Purwanti, 2024) applied descriptive analysis to profile poverty, concluding that the impoverished population is predominantly characterized by low educational attainment. While these studies provide valuable insights, they exhibit significant limitations. The research using key monetary metrics (poverty rate, poverty gap index, poverty severity index), remains descriptive in nature. It effectively quantifies the number of poor people and the intensity of monetary poverty but does not advance to classify or group provinces based on their distinct poverty profiles.

The research gap is a lack of studies that provide a grouping of Indonesian provinces based on the multifaceted characteristics of poverty. Existing research either lacks geographical scope, and analytical depth. The novelty of this study lies in its approach to address this gap. This research will apply the K-Means clustering algorithm to a comprehensive set of multidimensional poverty indicators (e.g., health, education, standard of living). K-Means Cluster Algorithm is one of clustering methods. K-means is also a well-known clustering method with the aim of minimizing the Euclidean distance between each point. In addition, Advantages of k-means are its simplicity and speed (Shutaywi & Kachouie, 2021)

This methodology will generate a novel provincial grouping that reveals the unique combination of deprivations defining each cluster. Consequently, this study will not only show the differences between groups but also provide a nuanced map of poverty characteristics across Indonesia, offering a robust evidence base for designing targeted and effective policy interventions.

This study aims to address this need by grouping Indonesia's provinces based on multidimensional poverty characteristics. Cluster analysis is a powerful analytical method for grouping objects based on similarities. A cluster is understood as a conceptually meaningful group of objects that have common characteristics (Anitha & Patil, 2022). This analysis can be used to group provinces based on poverty characteristics.

2. Methods

The method used in this study is non-hierarchy cluster analysis, employing the k-means algorithm. Reducing dimension by Principal Component Analysis (PCA) also applied in this study.

(a) Data and Variables

This study used data collected by BPS on multidimensional poverty across 38 provinces. There are 8 variables used as listed in Table 1.

Table 1. Variables and Definition

Variables	Definition	Unit
X1	Mean Years School of Population Aged 15 Years and Over	Years
X2	Percentage of Household with Access to Improved Drinking Water Source	Percentage
X3	Percentage of Household With Access to Adequate Housing	Percentage
X4	Percentage of Households Using Safely Managed Sanitation Service	Percentage
X5	Prevalence of Undernourishment	Percentage
X6	Population Growth Rate	Percentage
X7	Life Expectancy Rate	Year
X8	Percentage of malnourished children under-fives receiving supplementary feeding	Percentage

(b) The Step of Analysis

The data in this study was analyzed using Rstudio with following steps:

1. Data Description
In this step, the data will be examined through descriptive analysis.
2. Principal Component Analysis Process
At this stage, PCA process is applied to reduce dimension of dataset and also reduce multicollinearity (Abdulhafedh, 2021).
3. Determining best number of cluster using silhouette
At this stage, determine the number of clusters needed and select k arbitrary centers.
4. Calculating the distance
Clustering analysis groups objects based on their similarity, measured by the distance between the cluster's centroid and each object (Alzami dkk., 2023) The euclidian distance formula is used for calculation.

$$d_{(x,c)} = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \quad (2.1)$$

Where x_i is data point- i and c_i is centriot i

5. Creating new centroid
The new centroid for cluster is calculated by taking the average of the objects that are members of each cluster.
6. Repeat the process
Repeat the process from the second step. In addition, the K-Means algorithm will stop running when there is no change in cluster assignments.
7. Validating of K-means Clustering
The evaluation of cluster output is important to data modeling. One of the ways to measure the validation of clusters is Davies Boulding Index (DB). The lower dunn index value, the better cluster

separation (Efmi Maiyana & Wizra Aulia, 2025; Mukti dkk., 2025). The Davies Bouldin Index is calculated using following formula:

$$DB = \frac{1}{n} \sum_{i=1}^n \frac{S_i + S_j}{d(c_i c_j)} \quad (2.2)$$

Where n is number of cluster, S_i is average distance of all points in cluster i from cluster centroid ci measure inter-cluster distance, d measure intra-cluster distance, then i, j and k are indices cluster.

8. Interpretation of the clusters. After clusters are formed, the characteristics of clusters will be interpreted based on provincial data within those groups.
9. Conclusion. This section contains conclusions and recommendations for the next studies.

3. Results and Discussion

(a) Descriptive Statistics

Table 2 shows descriptive statistics of multidimensional poverty in Indonesia. Some variables have a wide range, for example, the variable of households with access to adequate housing (X3). This variable has a wide range among provinces. There are provinces with less than 5% of livable houses, while other provinces already have 86.68% livable houses. A similar pattern also occurs in other variables.

Table 2. Descriptive Statistics of Data

Variables	Min	Mean	Median	Max
X1	4.21	8.89	8.885	11.49
X2	30.64	89.02	89.015	99.96
X3	4.44	65.32	65.32	86.68
X4	0.64	6.96	6.96	22.93
X5	2.55	9.71	9.705	37.69
X6	0.31	1.37	1.365	1.93
X7	67.44	73.92	739.225	76.06
X8	18.7	76.87	76.865	100

The Correlation between variable also appear in this study. Figure 1 shows correlation between variables. The darker the color of the circle on the plot, the higher the correlation between the variables.

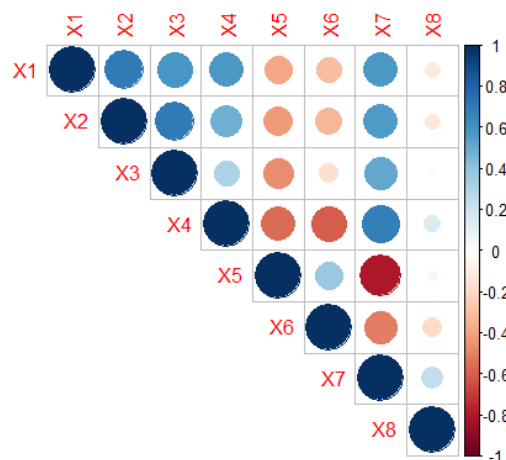


Figure 1. Correlation Plot of Variables

Several variables have correlation as shown in Figure 1. To handle the number of variables and correlation between variables, Principal Component Analysis (PCA) is employed in the process to obtain better clusters.

(b) PCA Process

PCA is one of analysis utilizing to reduce variables in a dataset, can see in Table 3. It also can handle multicollinearity.

Table 3. Eigen value of PCA

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Eigen Values	4.08	1.37	0.804	0.68	0.45	0.24	0.231	0.11

Eigen values that show in Table 3 are used to determine the number of Principal Components (PC). The requirement is eigenvalues greater than 1. In this study, there are 2 PCs that are used in future analysis. Initially, there are 8 variables, then after reducing by PCA, there are 2 PCs.

(c) K-Means Cluster and Validation

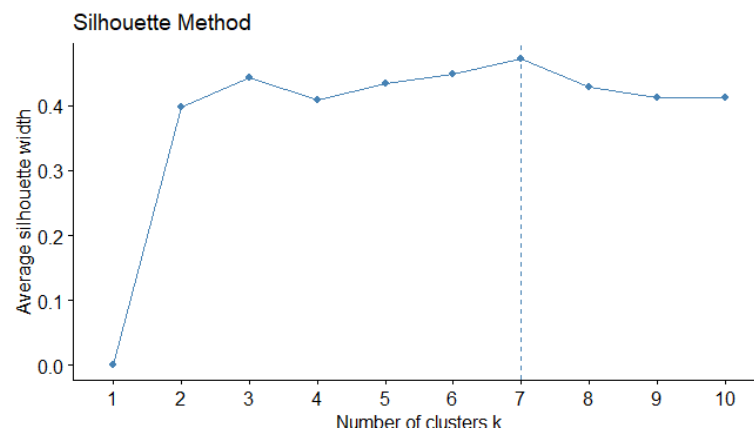


Figure 2. Number of cluster with silhouette Method

After earning new PCs for k-means analysis, the data frame is employed to select the optimal number of clusters using silhouette method. The scree plot used to select the optimal number of clusters is shown in Figure 2. Based on the plot, the optimal number of clusters in this study is 7 clusters.

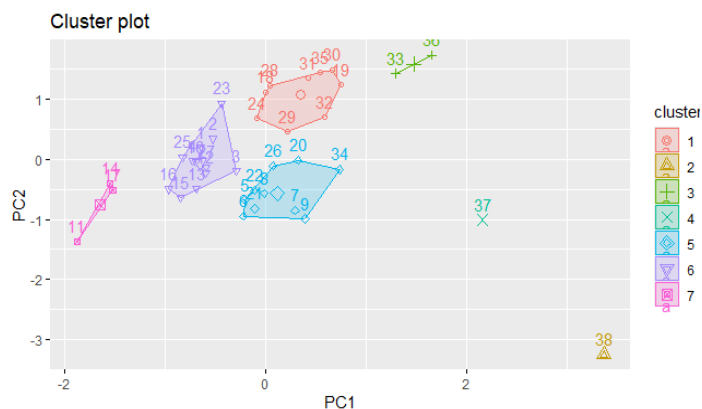


Figure 3. Cluster of Province based on multidimensional poverty in Indonesia

The K-means method was used to cluster provinces in Indonesia. This algorithm produces 7 clusters, which are shown in Figure 3. Two clusters consist of one province, one cluster consists of two provinces, one cluster consists of three provinces, one cluster consists of 12 provinces, one cluster consists of 10 provinces, and one cluster consists of 9 provinces. Table 4 provides information about the members of

each cluster.

Table 4. Clustering of provinces

Cluster	Provinsi
1	Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Utara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua
2	Papua Pegunungan
3	Papua Barat, Papua Selatan
4	Papua Tengah
5	Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Papua Barat Daya
6	Aceh, Sumatera Utara, Sumatera Barat, Riau, Kep. Riau, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Kalimantan Timur, Sulawesi Utara, Sulawesi Selatan
7	Dki Jakarta, Di Yogyakarta, Bali

The characteristics of these clusters are shown in [Table 5](#). Papua Pegunungan, Papua Selatan, Papua Tengah, and Papua Barat Daya are young provinces in Indonesia. However, they are in different clusters, which indicates that their poverty characteristics also differ.

Table 5. Profiling Cluster

Clusters	X1	X2	X3	X4	X5	X6	X7	X8
1	8.95	90.42	65.47	3.99	16.53	1.48	71.55	60.61
2	4.21	30.64	4.44	0.64	27.26	1.43	67.44	100.00
3	8.12	77.14	48.28	2.81	25.59	1.58	68.51	34.50
4	6.12	81.53	26.79	1.71	37.69	1.54	68.23	98.66
5	8.58	81.67	58.97	6.70	10.47	1.31	73.34	84.20
6	9.34	92.82	68.73	11.16	7.68	1.25	74.60	76.09
7	10.32	98.40	70.56	18.27	5.26	0.56	75.64	75.90

The validation cluster is determined by Davies Bouden Index (DB). In this study, the value of DB is 0.47. It shows that provinces separate well. Another validation is the Calinski-Harabasz Index, which has a value of 61.92. It means that each cluster has the lowest similarity.

4. Discussion

In this research, we obtain 7 clusters from 8 variable multidimensional poverty. Three areas in multidimensional poverty are education, health, and living standards. The lowest multidimensional poverty was found in provinces in cluster 7. The levels of education, health, and living standards in these provinces are better than in other provinces in Indonesia, even though they have the lowest population growth rates. The education indicator in this study is determined by the mean years of schooling. In cluster 7, the average year of schooling is more than 10 years, which is the highest number of schooling years. This shows that higher education increases the probability of becoming a non-poor family (Arsani et al., 2020; Artha & Dartanto, 2018; Kusumaningsih dkk., 2022). The second cluster with lower multidimensional poverty is cluster 6, consisting of 12 Provinces. Other clusters having good quality in education, health and standard living are provinces in cluster 5 and cluster 1. Cluster 1 has better education and standard of living indicators compared cluster 5. On the other hand cluster 5 is better in the health area compared to cluster 1.

Cluster 3 consists of 2 provinces. In this cluster, the mean year of schooling is almost 9 years, but it does not achieve the Indonesian program called “9-year Compulsory Education Program”. It is important to improve the Compulsory Education Program (Artha & Dartanto, 2018). The living standards need to be improved because less than 50% of the population lives in decent housing, and less than 5% have adequate sanitation according to standards. In addition, the health sector needs to be improved in terms of malnutrition prevalence.

Cluster 4 is a province with a high level of multidimensional poverty. That province is Papua Tengah. The

last cluster, cluster 2, has the highest level of multidimensional poverty, namely the Papua Pegunungan. Both Papua Tengah and the Papua Pegunungan have lower average years of education, 6.12 and 4.21, respectively. This shows that, in general, the population of Papua Tengah only completes elementary school. Meanwhile, the population of the Papua Pegunungan only attends elementary school and does not complete it. This result is also consistent with the previous study which found the impoverished people in Papua do not finish primary school (Wahyuni & Damayanti, 2014). The average years of education have an impact on poverty. This is in line with the results of a previous study by Rochman dkk (2025). The study shows that an increase in the average years of education will reduce the poverty rate (Rahim et al., 2024; Surbakti et al., 2023). This two clusters need special policies to improve their condition.

5. Conclusion

In this study, the K-means Cluster method was applied to group provinces in Indonesia based on multidimensional poverty. PCA was also applied to reduce the number of variables and multicollinearity. There were 7 clusters obtained with different characteristics. Based on validation, the clusters formed were well separated, as indicated by lower DBI scores. Two clusters (clusters 4 and 2) were the clusters with the most severe multidimensional poverty compared to the other 5 clusters. These two clusters require improvement in all areas, such as education, health, and living standards. As the data in this study contains correlations between several variables, the recommendation for further research is to compare clustering using multicollinearity with clustering where correlations have been addressed.

References

1. Abdulhafedh, A. (2021). Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation. *Journal of City and Development*, 3, 12–30.
2. Alzami, F., Sambasri, F. D., Nabila, M., Megantara, R. A., Akrom, A., Pramunendar, R. A., Prabowo, D. P., & Sulistiyawati, P. (2023). Implementation of RFM Method and K-Means Algorithm for Customer Segmentation in E-Commerce with Streamlit. *ILKOM Jurnal Ilmiah*, 15(1), 32–44. <https://doi.org/10.33096/ilkom.v15i1.1524.32-44>
3. Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences*, 34(5), 1785–1792. <https://doi.org/10.1016/j.jksuci.2019.12.011>
4. Arsani, A. M., Ario, B., & Ramadhan, A. F. (2020). Impact of Education on Poverty and Health: Evidence from Indonesia. *Economics Development Analysis Journal*, 9(1), 87–96. <https://doi.org/10.15294/edaj.v9i1.34921>
5. Artha, D. R. P., & Dartanto, T. (2018). The Multidimensional Approach to Poverty Measurement in Indonesia: Measurements, Determinants and its Policy Implications. *Journal of Economic Cooperation and Development*, 39(3), 1–38.
6. Asrol, A., & Ahmad, H. (2018). Analysis of factors that affect poverty in Indonesia.
7. Darma, A. V., Wicesa, N. A., Setyanti, A. M., Rochmat, T. M. (2024). Is Extreme Poverty in Indonesia is Going to End? *Journal of Indonesian Applied Economics*, 12(1), 126–136. <https://doi.org/10.21776/ub.jiae.2024.012.01.9>
8. Edison, E., & Andriansyah, M. (2023). Pertumbuhan Ekonomi dan Ketimpangan Sosial: Tinjauan Terhadap Kebijakan Pembangunan di Indonesia. *Journal Development*, 11(2), 134–146. <https://doi.org/10.53978/jd.v11i2.323>
9. Edna Safitri, S., Triwahyuningtyas, N., & Sugianto, S. (2022). Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Kemiskinan Di Provinsi Banten. *Sibatik Journal: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, 1(4), 259–274. <https://doi.org/10.54443/sibatik.v1i4.30>
10. Efmi Maiyana & Wizra Aulia. (2025). Optimization Of Stunting Infant Data Clustering With K-Means++ Algorithm Using Dbi Evaluation. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 11(1), 162–170. <https://doi.org/10.33480/jitk.v11i1.7007>
11. Faharuddin, F., & Endrawati, D. (2022). Determinants of working poverty in Indonesia. *Journal of Economics and Development*, 24(3), 230–246. <https://doi.org/10.1108/JED-09-2021-0151>
12. Hasan, Z. (2021). The Effect Of Economic Growth And Human Development Index On Poverty In Indonesia. *Journal Of Economics And Sustainability*, 3(No.1), 42–53.

- <https://doi.org/10.32890/jes2021.3.1.5>
13. Idrus, S., & Rosida, L. (2020). Poverty In Indonesia: Critical Review. *Archives of Business Research*, 8(6), 127–132. <https://doi.org/10.14738/abr.86.8326>
 14. Kusumaningsih, M., Setyowati, E., & Ridhwan, H. R. (2022). Study on the Impact of Economic Growth, Unemployment, and Education on South Kalimantan Province's Poverty Level from 2014 to 2020. *Business and Management Research*, 218, 170–176.
 15. Mukti, B. P., Hariguna, T., & Tahyudin, I. (2025). Model Klastering Hybrid Menggunakan Inisialisasi K-means++ dan Algoritma Optimasi Grey Wolf. 13(2).
 16. Purwanti, E. (2024). Analisis Deskriptif Profil Kemiskinan Indonesia Berdasarkan Data BPS Tahun 2023. *AKADEMIK: Jurnal Mahasiswa Humanis*, 4(1), 1–10. <https://doi.org/10.37481/jmh.v4i1.653>
 17. Rahim, A., Haryadi, W., & Muliawansyah, D. (2024). Analisis Faktor Rata-Rata Lama Sekolah Dan Pengangguran Terbuka Dalam Mempengaruhi Tingkat Kemiskinan Di Kabupaten Sumbawa. *Jurnal Ekonomi & Bisnis*, 12(1), 14–25. <https://doi.org/10.58406/jeb.v12i1.1528>
 18. Rochman, S., Indriyani, F., Sapitri, S. D., Anggreini, P., & Yuliani, S. (2025). Peran Pendidikan Berkualitas dalam Memutus Kemiskinan untuk Mendukung SDGs di Indonesia. 4.
 19. Sastiono, P., & Nuryakin, C. (2019). Inklusi Keuangan Melalui Program Layanan Keuangan Digital dan Laku Pandai. *Jurnal Ekonomi dan Pembangunan Indonesia*, 19(2), 242–262. <https://doi.org/10.21002/jepi.2019.15>
 20. Shutaywi, M., & Kachouie, N. N. (2021). Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering. *Entropy*, 23(6), 759. <https://doi.org/10.3390/e23060759>
 21. Suhendar, F. A., Sari, R. V., Pangesti, T., Putra, Z. M. G., & Santoso, A. P. A. (2024). The Impact of Poverty in Indonesia on Education. *JISIP (Jurnal Ilmu Sosial Dan Pendidikan)*, 8(2), 1119. <https://doi.org/10.58258/jisip.v8i2.6682>
 22. Surbakti, S. P. P., Muchtar, M., & Sihombing, P. R. (2023). Analisis Pengaruh Tingkat Pendidikan terhadap Kemiskinan di Indonesia Periode 2015-2021. *Ecoplan*, 6(1), 37–45. <https://doi.org/10.20527/ecoplan.v6i1.631>
 23. Wahyuni, R. N. T., & Damayanti, A. (2014). Faktor-Faktor yang Menyebabkan Kemiskinan di Provinsi Papua: Analisis.