Quantitative Economics and Management Studies (QEMS) *ISSN: 2722-6247 (online)*
Vol. 6 No. 2 (2025)
https://doi.org/10.35877/454RI.qems3940

Implementation K-Medoids Algorithm for Clustering Indonesian Provinces by Poverty and Economic Indicators

Hardianti Hafid*, Sitti Masyitah Meliyana, Isma Muthahharah, & Zakiyah Mar'ah

Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia

Abstract

Regional development disparities in Indonesia remain one of the main challenges in formulating national development policies. This study aims to classify the 38 provinces in Indonesia based on four key indicators: the percentage of the population living in poverty, Gross Regional Domestic Product (GRDP) per capita, the open unemployment rate, and the Human Development Index (HDI), using the K-Medoids algorithm. This method was chosen due to its robustness to outliers and its ability to produce representative clusters. The data used are secondary data obtained from the Central Bureau of Statistics (BPS). The analysis process began with data standardization, determination of the optimal number of clusters using the Elbow and Silhouette methods, followed by clustering implementation and result interpretation. The analysis results identified four main clusters with distinct socioeconomic characteristics. Cluster 1 reflects provinces with moderate conditions, Cluster 2 represents more developed provinces, Cluster 3 highlights regions facing significant development challenges, and Cluster 4 consists of provinces with the most underdeveloped socioeconomic conditions. These findings indicate that the K-Medoids algorithm is effective in identifying inter-provincial disparity patterns and can serve as a foundation for formulating more targeted and inclusive development policies.

Keywords: K-Medoids, Poverty, Gross Regional Domestic Product (GRDP), Human Development Index (HDI), Regional Development

Received: 14 January 2025 Revised: 3 April 2025 Accepted: 25 April 2025

1. Introduction

Sustainable and equitable national development is one of the main goals in the Indonesian government's strategic planning (Putra et al., 2024). As an archipelagic country with 38 provinces that have highly diverse geographical, demographic, and economic characteristics, Indonesia faces major challenges in managing regional development disparities. These inequalities are reflected in the disparities of key socioeconomic indicators such as poverty rate, Gross Regional Domestic Product (GRDP), open unemployment rate, and Human Development Index (HDI). The differences in performance across provinces on these indicators indicate an imbalance in the distribution of resources, infrastructure, and economic opportunities, which can affect the overall quality of life. Efforts to address regional inequality require comprehensive mapping and analysis of the socioeconomic conditions in each region. One relevant approach is to group (classify) the provinces in Indonesia into several categories based on the similarity of certain economic and social indicators. By understanding the characteristics of each group of provinces, policymakers can develop more targeted and region-specific development strategies. For example, provinces with high poverty and unemployment rates may require different interventions compared to provinces with a high HDI but stagnant economic growth.

In the context of data analysis, regional grouping based on specific characteristics can be performed using clustering methods. Cluster analysis is a crucial subject and technique in data mining, machine learning, and statistics for supporting decision-making processes (Rizqia & Ratnasari, 2023). This analysis originates from multiple disciplines and is applied across a wide range of fields and practical uses (Zhao & Zhou, 2021). One algorithm that can be used for this purpose is the K-Medoids algorithm, a partition-based clustering method known for its robustness to outliers. K-medoids also offers better scalability for larger datasets because it is more efficient compared to k-means (Soni & Patel, 2017). Unlike the K-Means algorithm, which uses the mean of cluster points as the center, K-Medoids selects

* Corresponding author.
E-mail address: hardiantihf@unm.ac.id

actual data points as cluster centers, making it more stable when applied to economic data, which often contains extreme values.

Previous studies have demonstrated the effectiveness of the K-Medoids algorithm in clustering data based on socioeconomic characteristics. For example, Lapiza et al. (2023) grouped districts/cities in the Sumatera region based on economic development indicators using the K-Medoids algorithm and produced a silhouette coefficient value of 0.13. Furthermore, Dwi et al. (2023) applied the K-Medoids method for regional mapping based on the number of poor people in the Central Java Province. The results of her study showed that six districts/cities (17%) were included in the high cluster, while 83% were included in the low cluster., Hafid & Annisa (2025) utilized K-Medoids and K-Prototypes to detect hypertension risk, showing that K-Medoids produced better clustering results in distinguishing high- and low-risk groups based on health indicators. Meanwhile, Haumahu & Matdoan (2022) clustered poverty levels in the districts and cities of the Maluku and Papua Islands, identifying four clusters with distinct socioeconomic characteristics, reinforcing the relevance of clustering approaches for mapping regional disparities.

Based on this background, this study aims to classify Indonesian provinces using four key indicators: the percentage of the population living in poverty, GRDP per capita, open unemployment rate, and Human Development Index (HDI), by applying the K-Medoids algorithm. This research is expected to contribute to understanding the distribution patterns of socioeconomic characteristics in Indonesia. The resulting classification not only serves as descriptive information but also acts as a tool to support the formulation of more equitable and inclusive regional development policies. Additionally, the use of data mining methods such as K-Medoids in regional analysis highlights how quantitative and technological approaches can be integrated into data-driven policy making—an increasingly vital aspect in today's era of digital transformation.

2. Research Method

2.1. Type and Source of Data

This study uses secondary data obtained from the official publications of Statistics Indonesia (Badan Pusat Statistik, BPS). The data includes socioeconomic indicators from 38 provinces in Indonesia. The data was collected and analyzed using the K-Medoids clustering method.

2.2. Research Variables

The variables used in the analysis are social and economic indicators that are considered to reflect the level of welfare and development of a region. The explanation of each variable is as follows:

a. Poverty Rate (X1): The proportion of the population living in poverty relative to the total population in a province.

b. GRDP per Capita (X2): The value of goods and services produced in a province divided by the number of residents.

c. Unemployment Rate (X3): The percentage of the labor force that is not employed but is actively seeking work.

d. HDI (Human Development Index) (X4): A composite measure reflecting the average achievement of a province in three basic dimensions of human development: long and healthy life, knowledge, and a decent standard of living.

2.3. Method of Analysis

The method used in this research is the K-Medoids algorithm, a non-hierarchical, partition-based clustering method. K-Medoids is similar to K-Means, but more robust to outliers, as the cluster center (medoid) is an actual data point within the cluster rather than a calculated mean.

The steps of the analysis are as follows:

a. Data Standardization: All variables are converted to a comparable scale using z-score standardization.

b. Determination of Optimal Number of Clusters: The optimal number of clusters is determined using two approaches: the Elbow Method (Sagala et al., 2022) and Silhouette method. The Silhouette method was first developed by (Kaufman & Rousseeuw, 1990), this method produces Silhouette values ranging from -1 to 1. A

value close to 1 indicates that an object fits very well within its assigned cluster. Therefore, if a model generates clusters with high Silhouette values, the model can be considered valid and appropriate for use in analysis .

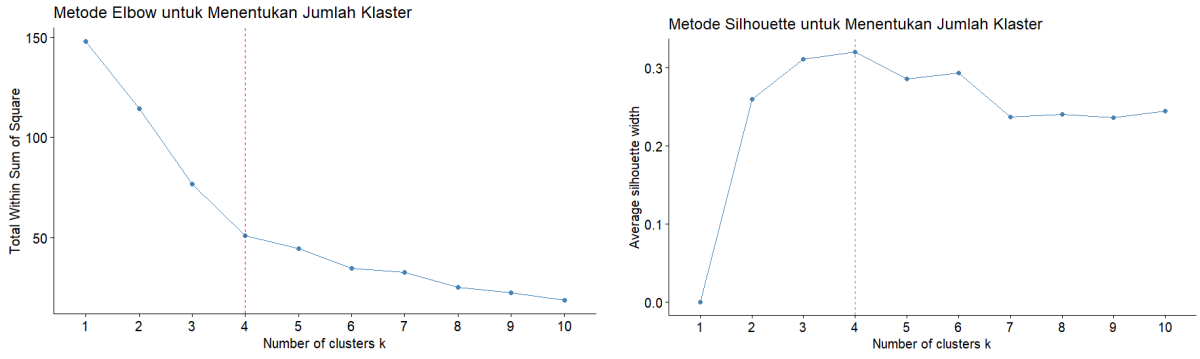
c. Clustering Process: The data is analyzed using the K-Medoids algorithm to group provinces based on similarities in socioeconomic characteristics. The steps followed in the K-Medoids clustering analysis (Azmi et al., 2023) :

1) Determine the number of clusters to be formed, and randomly select K objects from n data points.
2) Perform random selection to be used as the initial medoids.
3) Calculate the distance between each medoid and all data points.
4) Assign each data point to the cluster of the nearest medoid based on the smallest distance.
5) Calculate the total cost (sum of distances between data points and their respective medoids).
6) Determine new medoids by selecting K objects from the data.
7) Recalculate the distance between each data point and the new medoids within each cluster.
8) Calculate the change in cost to determine the total deviation (S). If S < 0, repeat the iteration.
9) Continue the iteration until there is no change in the position of the medoids.

d. Result Interpretation: The clustering results are analyzed to identify the characteristics of each cluster, and then visualized using two-dimensional graphs and regional distribution maps based on the geographic coordinates from the data.

3. Results and Discussion

The determination of the optimal number of clusters in this analysis was carried out using two evaluative approaches: the Elbow Method and the Silhouette Method. Figure 1(a) presents the Elbow plot, which illustrates the relationship between the number of clusters (k) and the total distortion cost (total within-cluster dissimilarity). In the plot, a clear "elbow" point appears at k = 4, indicating that adding more clusters beyond this point does not result in a significant reduction in dissimilarity. This suggests that using four clusters is an efficient choice in terms of cluster compactness. Meanwhile, Figure 1(b) shows the evaluation results using the Silhouette Method, which measures how well each object (in this case, each province) fits within its assigned cluster. The mostly positive and relatively high Silhouette coefficient values support the selection of four clusters, as they indicate good cluster separation and strong internal cohesion.



(a) (b)

Figure 1. Optimal Cluster Determination Plots: (a) Elbow Method and (b) Silhouette Method

The visualization of the clustering results using the K-Medoids algorithm is presented in Figure 2 as a two-dimensional scatter plot. This plot illustrates the distribution of 38 provinces in Indonesia based on economic and poverty indicators, which have been reduced to two dimensions using dimensionality reduction techniques. Each point on the plot represents one province, while different colors or symbols indicate cluster membership. The clustering results form four relatively distinct clusters, reflecting the diversity of economic characteristics across regions. Cluster 1 appears as a relatively dense group, consisting of provinces with moderate levels of poverty and unemployment, as well as medium GRDP and HDI values. Cluster 2 comprises economically more advanced provinces, characterized by high HDI and GRDP values and low unemployment rates, and is clearly grouped in a specific area on the plot. Meanwhile, Cluster 3, the largest cluster, is more widely dispersed, indicating high heterogeneity among provinces facing greater economic challenges, such as low to moderate GRDP and HDI, along with relatively high poverty and

unemployment rates. Finally, Cluster 4 emerges as a very small and isolated group, consisting of provinces in Eastern Indonesia with extreme socioeconomic characteristics. Its distant position from the other groups suggests that the provinces in this cluster can be considered outliers in the cluster analysis.

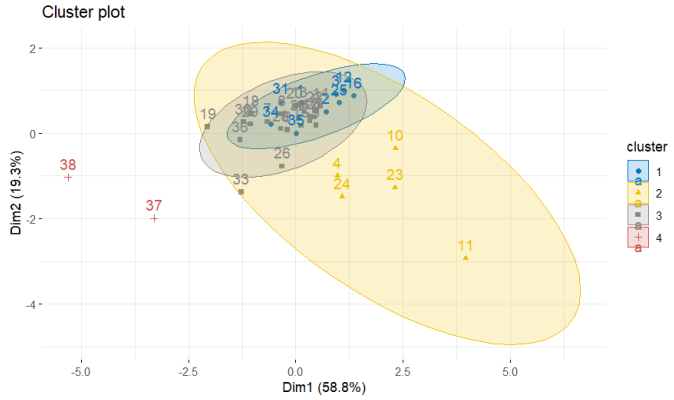


Figure 2. Cluster Result Plot

Table 1. Regional Characteristics of 38 Provinces Based on Clustering Results

Cluster	Number of Provinces	List of Provinces	Characteristics
1	9	Aceh, North Sumatra, West Sumatra, West Java, Central Java, Banten, South Kalimantan, Central Sulawesi, West Sulawesi	Has a moderate poverty rate, relatively average GRDP, medium HDI, and a fairly balanced unemployment rate.
2	5	Riau, Riau Islands, DKI Jakarta, East Kalimantan, Bali	High HDI, very high GRDP, and relatively low unemployment rate. This represents the more economically advanced provinces.
3	22	Jambi, South Sumatra, Bengkulu, Lampung, Central Kalimantan, West Kalimantan, North Kalimantan, North Sulawesi, Southeast Sulawesi, Gorontalo, Maluku, North Maluku, Papua, West Papua, West Nusa Tenggara, East Nusa Tenggara, South Sulawesi, East Java, Special Region of Yogyakarta (DIY), Southwest Papua, South Papua, Bangka Belitung	The majority of provinces fall into this cluster. GRDP is low to moderate, HDI varies but tends to be in the medium range, and both unemployment and poverty rates are relatively high.
4	2	Central Papua, Highland Papua	This is the smallest cluster and is identified as an outlier or noise. Unique characteristics include very high poverty, low HDI, and extreme socioeconomic conditions compared to other regions.

Based on the clustering results presented in Table 1, which analyzed 38 provinces in Indonesia using the K-Medoids algorithm, the provinces are grouped into four clusters based on key indicators: poverty percentage, GRDP per capita, unemployment rate, and Human Development Index (HDI). Each cluster reflects specific economic and social characteristics that represent the current state of regional development in Indonesia.

Cluster 1, consisting of 9 provinces such as Aceh, North Sumatra, West Sumatra, and Central Java, displays a profile of moderate poverty, average GRDP, medium HDI, and a relatively balanced unemployment rate. This cluster represents regions with relatively stable development but not yet considered advanced. Provinces in this cluster typically have strong traditional and agrarian economic foundations. For instance, in Aceh is the study conducted by Jamal (2017), which analyzed the geographical economic concentration (GEC) in Aceh Province using the AGC Decomposition Index for the period 2001–2014. The study showed that the GEC in Aceh tended to decline by

approximately -7,09 percent per year. Since 2011, growth increased by about 1, 27 percent per year. From 2008 to 2014, population density began to emerge as an influential factor. In Central Java, The role of several government regulations in poverty alleviation, such as the minimum wage policy, has effectively reduced the poverty rate. However, unemployment has significantly increased the number of poor people.(Sriyana, 2018). Other studies also show that the poverty gap index and the poverty severity index in rural areas of Sumatra Island are relatively higher compared to urban areas (Mustika & Nurjanah, 2021). These examples reflect provinces with gradual development progress but without significant poverty reduction.

Cluster 2, made up of 5 provinces—Riau, Riau Islands, Jakarta, East Kalimantan, and Bali—represents the most developed regions. These provinces have high HDI, very high GRDP per capita, and relatively low unemployment rates. These characteristics indicate areas that serve as national economic growth centers, supported by industry, services, and international trade. Jakarta, as the (former) capital and center of government and finance (Satria et al., 2023), along with East Kalimantan which serves as the national energy hub, it has a coal mining industry that forms the backbone of the economy. This province, located on the island of Borneo, contributes a significant portion of the economic foundation needed for infrastructure development and energy supply in Indonesia (Izza & Afkarina, 2019). In addition, Bali Province has a unique economic structure compared to other provinces in Indonesia, where the majority of people's livelihoods come from the tertiary sector (tourism), which also contributes significantly to Indonesia's economy (Sukriani et al., 2023). Reinforces this cluster’s role as the symbol of advanced regions in Indonesia.

Cluster 3 is the largest cluster, comprising 22 provinces including Jambi, Lampung, West Kalimantan, East Nusa Tenggara, and Papua. This cluster represents areas facing significant development challenges, with low to moderate GRDP per capita, relatively high poverty and unemployment rates, and medium to low HDI. This highlights ongoing regional development disparities, especially outside Java and in Eastern Indonesia. Despite their natural resource wealth, many of these provinces still struggle with gaps in infrastructure, education access, and healthcare services. For example, poverty in Lampung Province is still relatively high, resulting in low and uneven welfare among the community in the region (Pertiwi & Purnomo, 2022). A similar situation also occurs with economic growth, which tends to increase every year in the districts/cities of Jambi Province but has not been followed by a significant decrease in poverty levels (Safri, 2021).

Cluster 4, consisting only of Central Papua and Highland Papua, is the smallest cluster and appears to be an outlier in the data distribution. These provinces show very high poverty levels, low HDI, and economies that are significantly lagging behind the rest of the country. This reflects the real conditions in the highland regions of Papua, which have historically faced structural disadvantages due to geographic barriers, limited access to basic infrastructure, and insufficient public service distribution. Papua faces complex challenges, including high poverty, social inequality, and low HDI, exacerbated by remote geography and limited infrastructure (Dalimunthe et al., 2022). The recent formation of new provinces in Papua is part of the government’s effort to accelerate development in this region.ni.

In conclusion, the clustering results clearly illustrate the persistent development disparities among Indonesian provinces. Only a few provinces fall into the advanced cluster (Cluster 2), while the majority remain in the lower to middle categories. These findings can serve as a foundation for developing more targeted and region-specific development policies, taking into account the unique characteristics of each cluster and prioritizing interventions based on regional needs.

4. Conclusion

This study successfully classified 38 provinces in Indonesia into four clusters based on indicators such as poverty percentage, Gross Regional Domestic Product (GRDP) per capita, open unemployment rate, and Human Development Index (HDI) using the K-Medoids algorithm. The clustering results reveal significant variations in socioeconomic characteristics across provinces, reflecting regional development disparities in Indonesia. The formed clusters represent groups of provinces with similar welfare characteristics, ranging from economically stable regions and advanced provinces to those facing major development challenges. One of the clusters even consists of provinces categorized as outliers due to their extreme socioeconomic conditions. These findings highlight the importance of data-driven approaches in formulating more targeted and region-specific development policies. By understanding the profile of each cluster, the government can design more effective intervention strategies to reduce interregional disparities and promote more inclusive and sustainable development across Indonesia.

References

Azmi, M., Putra, A. A., Vionanda, D., & Salma, A. (2023). *Comparison the Performance of K-Means and K-Medoids Algorithms in Grouping Regencies / Cities in Sumatera Based on Poverty Indicators*. *1*, 59–66. https://doi.org/10.24036/ujsds/vol1-iss2/25

Dalimunthe, A. A., Fitrianto, A., Sartono, B., Oktarina, S. D., & Articles, I. (2022). Regency Clusterization Based on Village Characteristics to Increase the Human Development Index (IPM) in Papua Province. *Jurnal Ekonomi Pembangunan*, *20*(02), 153–168.

Dwi, F., Sari, R., & Ediwijojo, S. P. (2023). Clustering Analysis Using K-Medoids on Poverty Level Problems in Central Java by District / City. *International Conference on Advance & Scientific Innovation*, *2023*, 78–87. https://doi.org/10.18502/kss.v8i9.13321

Hafid, H., & Annisa, S. (2025). *Implementation of K-Medoids and K-Prototypes Clustering For Early Detection of Hypertension*. *19*(1), 465–476.

Haumahu, G., & Matdoan, M. Y. (2022). K-Medoids Clustering Algorithm for Classification of Poverty Levels in Districts and Cities in The Maluku Islands and Papua. *Variance: Journal of Statsitics and Its Applications*, *4*(2), 81–87.

Izza, K., & Afkarina, I. (2019). Coal Mining Sector Contribution to Environmental Conditions and Human Development Index in East. *Journal of Environmental Science and Sustainable Development*, *2*(2), 192–207.

Jamal, A. (2017). Geographical Economic Concentration , Growth and Decentralization : Empirical Evidence in Indonesia. *Jurnal Ekonomi Pembangunan*, *18*(2), 142–158. https://doi.org/10.23917/jep.v18i2.2786

Kaufman, L., & Rousseeuw, P. . (1990). *Finding groups data: An introduction to cluster analysis.* John Wiley & Sons.

Lapiza, R., Amalita, N., & Fitria, D. (2023). Grouping The Districts in Sumatera Region Based on Economic Development Indicators Using K-Medoids and CLARA Methods. *UNP Journal of Statistics and Data Science*, *1*(1), 16–22. https://doi.org/10.24036/ujsds/vol1-iss1/13

Mustika, C., & Nurjanah, R. (2021). Rural and urban poverty models on Sumatra Island. *Jurnal Perspektif Pembiayaan Dan Pembangunan Daerah*, *9*(1), 107–114. https://doi.org/10.22437/ppd.v9i1.10684

Pertiwi, E., & Purnomo, D. (2022). Analysis of the Effect of Gross Regional Domestic Product (GRDP), Human Development Index (IPM), and Open Unemployment Rate (TPT) on Poverty Rate in Lampung Province. *2 Nd International Conference on Islamic Economics, Islamic Finance, & Islamic Law (ICIEIFIL)*, 47–61. https://proceedings.ums.ac.id/index.php/icieifil/article/download/315/314

Putra, A. A., Hasibuan, H. S., Tambunan, R. P., & Lautetu, L. M. (2024). Integration of the Sustainable Development Goals into a Regional Development Plan in Indonesia. *Suistainability*, *16*(10235). https://doi.org/https://doi.org/10.3390/su162310235

Rizqia, N., & Ratnasari, P. (2023). Comparative Study of k-Mean , k-Medoid , and Hierarchical Clustering. *Indonesian Journal Oo Life Sciences*, *5*(2), 9–20.

Safri, M. (2021). The Analysis Related to The Factors Which Affect The Poverty Levels of Districts/Cities in Jambi Province During 2014-2018. *Paradigma*, *12*(1), 1–12.

Sagala, N. T. M., Agung, A., & Gunawan, S. (2022). Discovering the Optimal Number of Crime Cluster Using Elbow , Silhouette , Gap Statistics , and NbClust Methods. *ComTech: Computer, Mathematics and Engineering Applications*, *13*(1), 1–10. https://doi.org/10.21512/comtech.v13i1.7270

Satria, A., Syaban, N., Appiah-opoku, S., Satria, A., Syaban, N., & Appiah-opoku, S. (2023). Building Indonesia ' s new capital city : an in-depth analysis of prospects and challenges from current capital city of Jakarta to Kalimantan prospects and challenges from current capital city of Jakarta to Kalimantan. *Urban, Planning and Transport Research*, *11*(1), 1–29. https://doi.org/10.1080/21650020.2023.2276415

Soni, K. G., & Patel, A. (2017). Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data. *Journal of Computational Intelegence Research*, *13*(5), 899–906.

Sriyana, J. (2018). Reducing Regional Poverty Rate in Central Java. *JEJAK : Jurnal Ekonomi Dan Kebijakan*, *11*(1), 1–11.

Sukriani, N. G. A. A., Suarbawa, I. W., Murthi, N. W., & Djayastra, I. K. (2023). Analysis Of Factors Affecting The Human Development Index In Districts/Cities In Bali Province. *Jurnal Ganec Swara*, *17*(4), 1568–1579.

Zhao, Y., & Zhou, X. (2021). K-means Clustering Algorithm and Its Improvement Research K-means Clustering Algorithm and Its Improvement. *Journal of Physics:Conference Series*, *1873*. https://doi.org/10.1088/1742-6596/1873/1/012074