J-KOMA **Journal of Computer Science and Applications**

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

Open Access

E-ISSN: *2620 - 4872*
Vol.*08*, ***No***.*02*
Doi: *10.21009/JKOMA.082.04*

Recivied: *13 Dec 2025*
Accepted: *22 Dec 2025*
Published: *26 Dec 2025*

Keywords:
Clustering;
K-Medoids;
Partitioning Around Medoids (PAM);
Povert.

Correspondence Email:
violaokta11@gmail.com

Clustering Analysis Of Districts/Cities In Sumatra Region Based On Poverty Percentage Using The K-Medoids Method

Viola Oktamelisa[1*], Muhammad Arib Alwansyah[2], Sigit Nugroho[3], Etis Sunandi[4]

[1,3,4]Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Bengkulu, Indonesia
[2]Department of Mathematics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta, Indonesia

Abstract
Clustering is a data grouping method applied to identify groups formed by combining elements that have the same characteristics. One of the clustering methods that can be used is the K-Medoids method known as Partitioning Around Medoids (PAM). This study aims to obtain groupings and determine the characteristics of the results of grouping regencies/cities in the Sumatra Region based on the percentage of poverty using the K-medoids cluster method. The data used are poverty data per district/city totaling 154 in the Sumatra Region with the variables used being the expected length of schooling, average length of schooling, open unemployment rate, and percentage of poor population. The results obtained in this study are that districts/cities in the Sumatra Region have 2 optimum clusters as seen from the silhouette index value and Davies-Bouldin index value. Cluster 1 reflects an area with a fairly high poverty rate, with an average expected length of schooling of 13,78 years, an average length of schooling of 8,65 years, an open unemployment rate of 3,67%, and a percentage of poor population of 14,76%. Cluster 2 reflects an area with a moderate poverty rate with an average expected length of schooling of 13,36 years, an average length of schooling of 8,58 years, an open unemployment rate of 4,75%, and a percentage of poor people of 7,34%..

INTRODUCTION

Development is the goal of a country, where the country is more advanced when there is an increase in its development [1]. One of the targets of national development is to reduce the number of poor people. Efforts to achieve this goal involve various policies and programs, including improving infrastructure, education and training, health, agriculture and fisheries, industry and technology, social policies [2]. Poverty is a condition where there is a lack of resources such as: food, clothing, shelter and drinking water, natural resources, human resources and things closely related to the quality of life. The impact of this poverty is that Indonesia is still classified as a developing country. The term developing country is used to refer to poor countries [3]. The problem of poverty characteristics can be done by grouping regions that have similarities and differences in characteristics [4]. One of the grouping methods in the field of statistics is cluster analysis [5]. Clustering is also often referred to as data segmentation. The advantage of using clustering is that it is a very effective data segmentation method for predicting and analyzing various problems including regional zone mapping [6]. K-medoids is a partitioning method in clustering that groups n objects into k clusters. In this method, a medoid from each cluster is selected as the object that best

Universitas Negeri Jakarta
Phone: (021) 4890856
East Jakarta City
Jakarta, 13220

represents that cluster. A medoid is an object in a cluster that has the lowest total distance to other objects in the same cluster [7].

Research on K-medoids clustering analysis has been conducted by [7], obtained the results of grouping the characteristics of each cluster formed based on the poverty indicator values in East Java in 2020 as many as 2 clusters. Cluster 1 consists of 30 districts/cities. While cluster 2 consists of 8 districts/cities. Based on previous research and explanations of K-medoids clusters, researchers are interested in grouping districts/cities in the Sumatra Region that have the same characteristics based on the percentage of poverty using the K-medoids clustering method. The purpose of this study is to obtain a grouping of districts/cities in the Sumatra Region based on the percentage of poverty using the K-medoids cluster method and to obtain the characteristics of the results of grouping districts/cities in the Sumatra Region based on the percentage of poverty using the K-medoids cluster method.

METHODS

2.1 Clustering K-Medoids

K-medoids clustering uses the partition clustering method to cluster a set of n objects into k clusters. The K-medoids algorithm uses objects in the set to represent a cluster. Objects representing a cluster are called medoids. Clusters are constructed by calculating the proximity between medoids and non-medoid objects [8]. Steps for K-medoids analysis [9]:

1. Initialize k cluster centers (the number of clusters).
2. Allocate each data object to the closest cluster using the Euclidean distance measure.
3. Randomly select an object in each cluster as a new candidate medoid.
4. Calculate the distance between each object in each cluster and the new candidate medoid.
5. Calculate the total deviation (S) by subtracting the total new distance from the total old distance. If $S < 0$, swap the object with the cluster data to form a new set of objects as medoids.
6. Repeat steps 3 through 5 until there are no changes in the medoids, resulting in clusters and their respective members.
7. The clustering results use the best clustering results obtained in the study using the silhouette index and dunn index and Davies-Bouldin index (DBI).

2.2 Euclidean distance

Euclidean distance is the calculation of the distance between two points. Euclidean is related to the Pythagorean Theorem, namely by calculating the square root. For example, A and B are in p-dimensional space, given $A(x_1, x_2, \dots, x_p)$ and $B(y_1, y_2, \dots, y_p)$, hen the Euclidean distance from A to B is [10]:

$$d_{AB} = \sqrt{\sum_{i=1}^{p} (x_i - y_i)^2} \quad (1)$$

2.3 Davies Bouldin Index

The Davies Bouldin Index (DBI) maximizes inter-cluster distances while simultaneously minimizing the distance between points within a cluster [11].

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq 1}(R_{ij}) \quad (2)$$

2.4 Silhouette Index

The Silhouette Index evaluates the placement of each object within each cluster by comparing the average distance between objects within a cluster and the distance between objects in different clusters [12].

$$S(i) = \frac{b(i) - a(i)}{max\,(a(i), b(i))} \quad (3)$$

Given:

$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)} & , jika\ a(i) < b(i) \\ 0 & , jika\ a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1 & , jika\ a(i) > b(i) \end{cases} \quad (4)$$

The range of values of the silhouette coefficient is -1 to 1. If the value of the silhouette coefficient is close to 1, then the object is in the right cluster, if it is around 0 then the object can be between 2 clusters, and if the result is negative, then the object may be in the wrong cluster. The best number of clusters or the optimal number of clusters is the number of clusters with the highest average silhouette score where the average is taken from the silhouette value of each cluster.

2.5 *Dunn* Index

This measure of cluster validity is based on the fact that separate clusters typically have large inter-cluster distances and small intra-cluster diameters [13].

$$D = \frac{\min_{i \neq j}\{d(C_i C_j)\}}{\max_k\{\delta(C_k)\}} \quad (5)$$

Where $\delta(C_i C_j)$ is the distance between clusters C_i dan C_j and $\Delta(C_k)$ is the diameter of the cluster C_k (the maximum distance within a cluster). The larger the Dunn Index, the better the clustering results.

RESULTS AND DISCUSSION

3.1 Descriptive Statistics

This study was conducted to determine the percentage of poverty groups in the Sumatra region using the K-Medoids method.

TABLE 1. Results of the Optimal Cluster Test

Total Data	K Value	Dunn Index Value	Silhouette Index Value	DBI Value
	2	0,0508	0,4240	0,9474
	3	0,0754	0,2948	1,2143
	4	0,0600	0,2756	1,3170
154	5	0,0903	0,2824	1,0975
	6	0,0859	0,2724	1,1818
	7	0,0859	0,2932	1,1479
	8	0,0889	0,3068	1,0818
	9	0,0851	0,3012	1,0660
	10	0,1070	0,3075	1,0578

Based on Table 1 above, the testing results show that the three validation measures demonstrate the optimal number of clusters selected for each validation measure. There are differences in the Dunn Index, Silhouette Index, and DBI values for optimal clustering results. As seen in the table, the Dunn Index validation test showed the optimal number of clusters with a value of 0.1070. In the Silhouette Index test, a cluster count of 2 indicated the optimal number of clusters with a value of 0.4240. Based on the Davies-Bouldin Index test, the optimum clustering result was two clusters with a value of 0.9474.

After conducting the Silhouette Index, Dunn Index, and Davies-Bouldin Index tests, the optimum clustering result was two clusters. The following is a visualization of the resulting clusters:

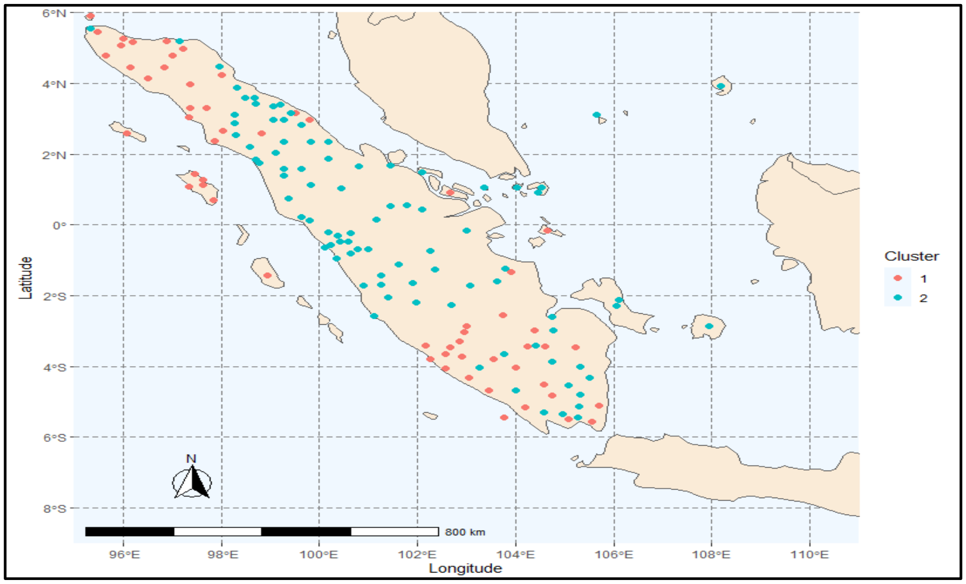

FIGURE 1. Visualization of Optimal Cluster Results

TABLE 2. Description of Optimal Cluster Results

Cluster	**Mean X_1**	**Mean X_2**	**Mean X_3**	**Mean X_4**	**Total**
1	13,78	8,65	3,67	14,76	57
2	13,36	8,58	4,75	7,34	97

Based on Table 2, it can be seen that cluster 1 has 57 members of districts/cities with an average value of Expected Years of Schooling (X_1) of 13.78 years in the fairly high category, Average Years of Schooling (X_2) of 8.65 years in the moderate category, Open Unemployment Rate (X_3) of 3.67% which is quite low, and the Percentage of the Poor Population (X_4) of 14.76% which tends to be high. From these characteristics, it can be concluded that cluster 1 reflects an area with a high level of poverty. Cluster 2 has 97 members of districts/cities with an average value of Expected Years of Schooling (X_1) of 13.36 years which is categorized as quite high, Average Years of Schooling (X_2) of 8.58 years which is categorized as moderate, Open Unemployment Rate (X_3) of 4.75% which is categorized as moderate, and the Percentage of the Poor Population (X_4) of 7.34% which is categorized as moderate. From these characteristics, it can be concluded that cluster 2 reflects an area with a moderate level of poverty.

CONCLUSION

Based on the results of a cluster analysis of poverty rates in the Sumatra region using 154 observations, it can be concluded that using $k = 2,3,4, \dots ,10$. The optimum clusters obtained based on the Silhouette Index and Davies-Bouldin Index values were two clusters, with values of 0.4240 and 0.9474, respectively. Cluster 1 has 57 districts/cities, and Cluster 2 has 97 districts/cities.

Cluster 1 has an average Expected Years of Schooling (X_1) of 13.78 years, an Average Years of Schooling (X_2) of 8.65 years, an Open Unemployment Rate (X_3) of 3.67%, and a Percentage of Poor Population (X_4) of 14.76%. Based on these characteristics, it can be concluded that Cluster 1 reflects an area with a high poverty rate. Cluster 2 has an average Expected Years of Schooling (X_1) of 13.36 years, Average Years of Schooling (X_2) of 8.58 years, Open Unemployment Rate (X_3) of 4.75%, and the Percentage of Poor Population (X_4) of 7.34%. From these characteristics, it can be concluded that cluster 2 reflects an area with a moderate level of poverty.

REFERENCES

[1] Wardhana, G. W., Yenni, K. 2024. Dampak Indeks Pembangunan Manusia Terhadap Laju Kemiskinan di Indonesia. *Jurnal Manajemen, Bisnis, dan Pendidkan Indonesia*. Vol. 11, No. 02.

[2] Nurhaeni. 2022. Pengaruh IPM, Pertumbuhan Ekonomi Dan Pengangguran Terhadap Kemiskinan Kab&Kota di Prov Jawa Tengah. *Jurnal Ekonomi*, 11(1), 26.

[3] Kasim, F. S. 2021. Determinan Kemiskinan Kabupaten Dan Kota Di Provinsi Sulawesi Tengah (Periode 2011-2020). *Tolis Ilmiah; Jurnal Penelitian*, 3(1), 124–129.

[4] Widodo, E., Ermayani, P., Laila, L. N., & Madani, A. T. 2021. Pengelompokkan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering. *Seminar Nasional Official Statistics*, Vol. 2021, No. 1.

[5] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. 2019. Multivariate Data Analysis Eighth Edition.

[6] Astri, D. F., & Martanto. 2024. Clustering Penduduk Miskin Menggunakan Algoritma K-means Pada Wilayah Jawa Barat. *Jurnal Mahasiswa Teknik Informatika*. Vol 8, No.2.

[7] Alfiah, F., Almadayani, A., Al Farizi, D., & Widodo, E. 2021. Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020. *Jurnal Ilmiah Sains*, 22(1), 1.

[8] M. A. Alwansyah, D. I. Fadli, D. S. Rini, M. A. Nabhan, and R. Rachmawati, "Grouping of Earthquakes in Sumatra Region Using Partitioning Around Medoids (PAM) and Clustering Large Applications (CLARA) Methods," *AIP Conf. Proc.*, vol. 3285, no. 1, 2025

[9] Wira, B., Budianto, A.E. & Wiguna, A.S. 2019. Implementasi Metode K-Medoids Clustering Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang. *Jurnal Terapan Sains & Teknologi*, 1(3): 54-69.

[10] M. A. Alwansyah, N. Hidayat, D. I. Fadli, D. S. Rini, and P. Jana, "Spatial zone analysis of earthquake in Sumatra," *Proceeding 7Th Int. Conf. Sci. Technol. Interdiscip. Res. (Ic-Star 2021)*, vol. 2601, p. 020009, 2023.

[11] Fauzi, I. B., Mochzen, G. R., dan Teguh, I. H. 2023. Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit. *Jurnal Masyarakat Informatika Unjani*. 7(2),1-15.

[12] Nahdliyah, M. A., Widiharih, T., & Prahutama, A. 2019. Metode K-Meoids Clustering Dengan Validasi Silhouette Index dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018). *Jurnal Gaussian*, 8(2), 161–170.

[13] Fajrina, E. A. N. 2021. *Analisis Cluster Robust MenggunakanMetode K-Medoids pada Data Mengandung Pencilan*. Universitas Lampung.