Comparison of K-Means and K-Medoids Clustering for Grouping The Sub-Districts In Bojonegoro Regency Based On Educational Supporting Factors

Alif Yuanita Kartini[(1)], Syarif Husen[(2)]

[1,2]Universitas Nahdlatul Ulama Sunan Giri

e-mail: alifyuanita@unugiri.ac.id[(1)], syarifhusenbafaqih@gmail.com[(2)]

ABSTRAK

Pendidikan di Bojonegoro saat ini masih belum merata. Hal ini dikarenakan upaya pemerataan pendidikan yang telah dilakukan mempunyai banyak kendala. Kendala yang sering terjadi adalah masyarakat yang berada di daerah terpencil dan jauh dari perkotaan kesulitan dalam mengakses layanan pendidikan. Oleh karena itu pengelompokan wilayah perlu dilakukan agar pemerintah kabupaten Bojonegoro dapat memperhatikan cluster wilayah yang memperlukan perbaikan pendidikan. Pada penelitian ini menggunakan metode K-Means dan K-Medoids untuk pengelompokan kecamatan di kabupaten Bojonegoro berdasarkan faktor pendukung pendidikan. K-Means merupakan salah satu metode unsupervised learning yang digunakan untuk menganalisis data dengan melakukan pengelompokan. Sementara itu K-Medoids merupakan metode pengelompokan partisi yang mengelompokkan sekumpulan n objek menjadi sejumlah k cluster. Data yang digunakan dalam penelitian ini merupakan data sekunder yang didapatkan dari Dinas Pendidikan kabupaten Bojonegoro berupa data faktor pendukung pendidikan yang meliputi jumlah sekolahan, jumlah tenaga pendidik, dan jumlah rombongan belajar (ROMBEL) tahun 2022 pada masing-masing kecamatan di kabupaten Bojonegoro. Dari hasil penelitian didapatkan metode K-Means lebih baik dibandingkan dengan metode K-Medoids. Hasil pengelompokan menggunakan K-Means didapatkan sebanyak 5 cluster, dimana cluster 1 beranggotakan 1 kecamatan, cluster 2 beranggotakan 7 kecamatan, cluster 3 beranggotakan 1 kecamatan, cluster 4 beranggotakan 12 kecamatan dan cluster 5 beranggotakan 7 kecamatan. Berdasarkan karakteristik dari masing-masing cluster yang didapatkan, diharapkan dapat digunakan sebagai masukan bagi dinas Pendidikan sebagai Upaya pemerataan Pendidikan di kabupaten Bojonegoro.
Kata kunci: K-Means; K-Medoids; Pemerataan Pendidikan; Pengelompokan

ABSTRACT

Education in Bojonegoro is currently still uneven. This is because efforts to equalize education that have been carried out have many obstacles. The obstacle that often occurs is that people who are in remote areas and far from urban areas have difficulty accessing education services. Therefore, regional grouping needs to be done so that the Bojonegoro district government can pay attention to regional clusters that need education improvement. This study used the K-Means and K-Medoids methods to group sub-districts in Bojonegoro district based on educational supporting factors. K-Means is one of the unsupervised learning methods used to analyze data by grouping. Meanwhile, K-Medoids is a partition grouping method that groups a set of n objects into a number of k clusters. The data used in this study is secondary data obtained from the Bojonegoro district Education Office in the form of data on education supporting factors which include the number of schools, the number of educators, and the number of learning groups (ROMBEL) in 2022 in each sub-district in Bojonegoro district. From the research results, it was found that the K-Means method was better than the K-Medoids method. The results of grouping using K-Means obtained as many as 5 clusters, cluster 1 consists of 1 sub-district, cluster 2 consists of 7 sub-districts, cluster 3 consists of 1 sub-district, cluster 4 consists of 12 sub-districts and cluster 5 consists of 7 sub-districts. Based on the characteristics of each cluster obtained, it is expected to be used as input for the Education office for equal distribution of education in Bojonegoro district.
Keywords: *K-Means; K-Medoids; Education Equity; Grouping*

INTRODUCTION

Education is one of the main factors in a nation's progress [1]. A developed nation has a good education. The benchmark of good education is equal distribution of opportunities to get proper education or commonly referred to as equal distribution of education [2]. Education equity has long been an issue that has received a lot of attention, especially in developing countries. Equality of education includes two important aspects, namely equal opportunities to obtain education and equal justice in society [3].

Education in Bojonegoro is currently still uneven. There are still many people who have not received the education they should have received since of 6 years old [4]. This is because efforts to equalize education have many obstacles. The obstacle that often occurs is that people who are in remote areas and far from urban areas have difficulty accessing education services [5]. In the implementation of education, there are several supporting factors in the implementation of effective and efficient teaching and learning activities, including the provision of schools, the provision of learning groups (ROMBEL) and the provision of educators [6]. Supporting factors in providing education must be felt by all regions, especially those in Bojonegoro district. Therefore, regional grouping needs to be done so that the Bojonegoro district government can pay attention to which regional clusters need improved education.

One method that can be used for clustering is cluster analysis [7]. Cluster analysis is the merging of objects or data based on similar characteristics [8]. There are many methods in cluster analysis, including K-Means and K-Medoids. In a study entitled "Mapping Excellent Class Students Using the K-Means Clustering Algorithm" stated that K-Means is an algorithm that is easy to implement [9]. In another study entitled "K-Means and K-Medoids Algorithm for Grouping Subdistricts of Social Assistance Recipients in Bojonegoro Regency" stated that the K-Means method has a relatively small complexity of space and time [10]. Meanwhile, in a study entitled "Hybrid K Means-Multivariate Adaptive Regression Splines For Distribution Of Dengue Fever Risk Mapping In Bojonegoro District" stated that the K-Means method is able to group big data very quickly [11]. Another study showed that K-Means is an effective method for grouping datasets [12], [13],[14]. Meanwhile, the K-Medoids method also has many advantages in several studies. In a study entitled "Implementation of K-Medoids Clustering Method for Grouping Data on Potential Forest/Land Fires Based on the Distribution of Hot Spots (Hotspot)" stated that K-Medoids is a method that is sensitive to outliers [15]. In another study entitled "Analysis of K-Medoids in Grouping the Ratio of Students to Teachers, Students with Rombel, and the Ratio of Rombel to Elementary and Junior High School Education Grades by Province" stated that K-Medoids are more reliable when there is data noise [16]. Based on research entitled "Comparison of Distance Measure on K-Medoids Clustering for ARI Disease Grouping" states that the K-Medoids algorithm can get the closest grouping results to an object [17]. Another study states that K-Medoids are not affected by other extreme data [18], [19].

Based on this background, research will be conducted for the grouping of sub-districts in Bojonegoro sub-district based on educational supporting factors using the K-Means and K-Medoids algorithms. From this research, it is expected to provide advice and input for the government to pay attention to areas that lack schools, educators, and learning groups (ROMBEL) in accordance with the cluster results obtained. This is an effort to equalize education in Bojonegoro district.

METHOD

The data used in this study is secondary data obtained from the Bojonegoro District Education Office which is published on the Bojonegoro one data web. The data is in the form of data on education supporting factors which include the number of schools, the number of educators, and the number of learning groups (ROMBEL) in 2022 in each sub-district in Bojonegoro district. The variables used in this study include number of elementary schools (x_1), number of junior high schools (x_2), number of high schools (x_3), number of elementary school educators (x_4), number of junior high school educators (x_5), number of high school educators (x_6), number of elementary school learning groups (x_7), number of junior high school learning groups (x_8), and number of high school learning groups (x_9).

The methods used in this study are K-Means and K-Medoids. K-Means is a distance-based clustering method that divides data into clusters. K-Means algorithm only works on numeric attributes. K-Means algorithm includes partitioning clustering that separates data into k separate regions. K-Means algorithm is well known for its ease and ability to segment big data and data outliers very quickly. In K-Means algorithm, each data must belong to a specific cluster. In K-Means algorithm it is possible for any data belonging to a particular cluster at one stage of the process, at the next stage moving to another cluster.

Meanwhile, K-Medoids is a classic partitioning technique of clustering that clusters datasets of n objects into k clusters known as a priori. K-Medoids algorithm operates on the principle of minimizing the number of similarities between each object and its corresponding reference points. The K-Medoids algorithm uses data objects as representatives (medoids) of the cluster center. K-Medoids algorithm is used to overcome the weaknesses of the k-means algorithm which is very sensitive to outliers. This is because these objects are very far from the majority of other data, so that if entered into a cluster, this kind of data can distort the average value (mean) of the cluster.
In this study using the help of Rapidminer software. The analysis steps used are as follows.

1. Group sub-districts in Bojonegoro based on education supporting factors using the k-means algorithm with the following steps [20],[21],[22],[23] :
   a. Determine the number of clusters to use (k)
   b. Split data into k clusters
   c. Calculate the centroid value or average of data in each cluster formed
   d. Calculates Euclidean distances to determine the distance to all mean values in each cluster
   e. Allocate data for clusters that have the average closest to the data
   f. Loop starting in step (b) if there is still data moving to the cluster
   g. Get clustering results along with members from each cluster
   h. Validate cluster quality by calculating Silhoutte values Coefficient
2. Grouping sub-districts in Bojonegoro based on education supporting factors using the K-Medoids algorithm with the following steps [15],[16],[17],[24]:
   a. Determine the number of clusters to be used (k)
   b. Determine medoids by randomly taking objects from objects to be grouped
   c. Calculate Euclidean distance to obtain the distance of non-medoids objects
   d. Determine the members of each cluster based on the closest distance to the grouping object
   e. Calculates the total distance of non-medoids objects closest to the grouping object

f. Randomly select one non-medoids object for each cluster that is used as a candidate for new medoids. If the selected object has been a medoid, it is not allowed to be re-selected.
g. Calculates the Euclidean distance at the distance of non-medoids objects to new medoids candidates for each cluster.
h. Determine the members of each cluster based on the closest distance to prospective new medoids
i. Calculates the total distance of nonmedoids objects closest to potential new medoids
j. Calculates the difference between the total distance and the new prospective medoids with the total distance and the medoids
k. If the value of the total distance difference is less than zero, then the candidate medoids become new medoids in the next iteration
l. Repeating from step (f) until there is no change in medoids
m. If the value of the total distance difference is more than zero, the iteration will stop and get the cluster and the members of each cluster

3. Comparing the results of sub-district grouping in Bojonegoro district based on supporting factors of education using K-Means and K-Medoids algorithms by looking at cluster distance performance values [25],[26].
4. Create a clustered map with the best method
5. Getting the characteristics of each cluster from the best method

RESULT AND DISCUSSION

Clusterization Results Using K-Means Algorithm

The first step to grouping using the K-Means algorithm is to preprocess the data. Data preprocessing is preparing data and standardizing it. The next step after preprocessing data is to process data. The number of clusters (k) used in this study was 2, 3, 4 and 5. Furthermore, to evaluate the optimal grouping results, namely by looking at the smallest average value within centroid distance. From the results of the analysis, the smallest average within centroid distance value in the number of clusters (k) was 5, which was 1.833. So that the number of clusters used to group sub-districts in Bojonegoro district based on supporting factors for education with the K-Means algorithm is 5. After obtaining the number of clusters to be used, the next step is to determine the initial centroid value randomly, calculate the Euclidian distance, grouping objects with the shortest distance between objects, until a fixed centroid value is obtained and cluster members do not move to another cluster. Based on these steps, clustering results using K-Means are obtained as shown in Table 1 below.

Table 1. Number and Members of Clusters Formed Using K-Means

Cluster	Number of cluster members	Cluster Members
Cluster 1	1	Kedungadem
Cluster 2	7	Baureno, Dander, Kalitidu, Kepohbaru, Ngasem, Padangan, Sumberejo
Cluster 3	1	Bojonegoro

Cluster	Number of cluster members	Cluster Members
Cluster 4	12	Balen, Bubulan, Gayam, Gondang, Kapas, Kasiman, Kedewan, Margomulyo, Ngambon, Ngraho, Sekar, Trucuk
Cluster 5	7	Kanor, Malo, Purwosari, Sugihwaras, Sukosewu, Tambakrejo, Temayang

Clusterization Results Using K-Medoids Algorithm

Just like grouping using the K-Means algorithm, grouping using the K-Medoids algorithm also begins with determining the number of clusters. The number of clusters (k) used in this study is the same as that used in K-Means, which are 2, 3, 4 and 5. To evaluate the best grouping results, namely by using the smallest average value within centroid distance. From the results of the analysis, the smallest average within centroid distance value in the number of clusters (k) was 5, which was 2.786. So that the number of clusters used to group sub-districts in Bojonegoro based on supporting factors for education with the K-Means algorithm is 5.

The next step after determining the number of clusters is to determine the initial centroid value as medoids. Furthermore, after going through several stages until the clustering process is stopped, the clusterization results are obtained using the K-Medoids algorithm as shown in Table 2 below.

Table 2. Number and Members of Clusters Formed Using K-Medoids

Cluster	Number of cluster members	Cluster Members
Cluster 1	12	Kalitidu , Kanor, Kasiman, Malo, Ngasem, Ngraho, Padangan, Purwosari, Sugihwaras, Sukosewu, Tambakrejo, Temayang
Cluster 2	5	Baureno, Dander, Kedungadem, Kepohbaru, Sumberejo
Cluster 3	1	Bojonegoro
Cluster 4	8	Balen, Bubulan, Gayam, Gondang, Kedewan, Margomulyo, Ngambon, Sekar
Cluster 5	2	Kapas, Trucuk

Comparison of Grouping Results Using K-Means and K-Medoids Algorithms

Based on the results of clusterization using the K-Means and K-Medoids algorithms as shown in Table 1 and Table 2, a comparison of clusterization results is obtained as shown in Table 3 below.

Table 3. Clusterization Results Using K-Means and K-Medoids Algorithms

Cluster	K-Means	K-Medoids
1	Kedungadem	Kalitidu , Kanor, Kasiman, Malo, Ngasem, Ngraho, Padangan, Purwosari, Sugihwaras, Sukosewu, Tambakrejo, Temayang
2	Baureno, Dander, Kalitidu, Kepohbaru, Ngasem, Padangan, Sumberejo	Baureno, Dander, Kedungadem, Kepohbaru, Sumberejo
3	Bojonegoro	Bojonegoro
4	Balen, Bubulan, Gayam, Gondang, Kapas, Kasiman, Kedewan, Margomulyo, Ngambon, Ngraho, Sekar, Trucuk	Balen, Bubulan, Gayam, Gondang, Kedewan, Margomulyo, Ngambon, Sekar
5	Kanor, Malo, Purwosari, Sugihwaras, Sukosewu, Tambakrejo, Temayang	Kapas, Trucuk

Furthermore, to get the best clustering method, namely by using the average value within centroid distance. The best clustering method is the one with the smallest average value within centroid distance. The results of the comparison of K-Means and K-Medoids algorithms by looking at the average value within centroid distance as shown in Table 4 below.

Table 4. Comparison Results of K-Means and K-Medoids Algorithms

Grouping Method	Average Within Centroid Distance
K-Means	**1,833**
K-Medoids	2,786

Based on Table 4, it can be seen that the average value within centroid distance for the K-Means method is 1.833. Meanwhile, using the K-Medoids method, the average value within centroid distance was 2.786. This means that the value of average within centroid distance using the K-Means method is smaller than the value of average within centroid distance using the K-Medoids method. So it can be concluded that the clusterization method using the K-Means method is better for grouping sub-districts in Bojonegoro district based on educational supporting factors.

Mapping Clusterization Results with the Best Method

Based on Table 4, the best clustering method for grouping sub-districts in Bojonegoro based on supporting factors for education is K-Means method. To create a map of clusterization results with the best method based on clusterization results using the K-Means method as shown in Table 1. The results of clusterization are in the form of maps as shown in Figure 1 below.

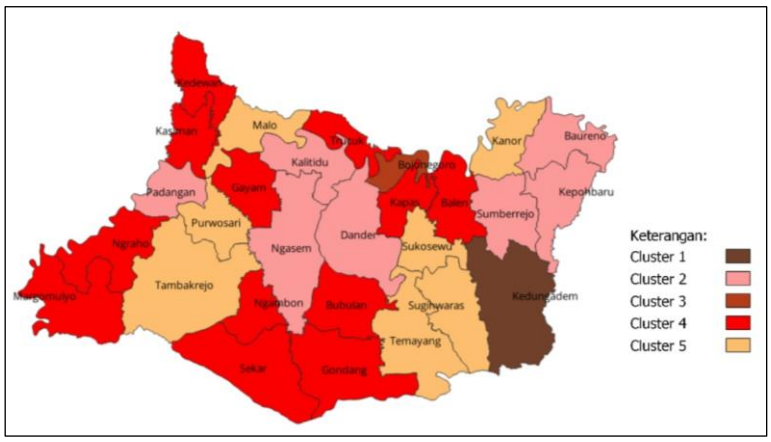

Figure 1. Map of Clusterization Results using the Best Method

Characteristics of Each Cluster Formed Based on the Best Method

To distinguish the results of clusterization formed, profiling is carried out by finding the average value of each variable. The clusterization results used are based on the results of clusterization using the K-Means method as shown in Table 1. The results of the profiling formed as shown in Table 5 below.

Table 5. Characteristics of Each Cluster Using the Best Method

Cluster	District	Characteristics
1	Kedungadem	This cluster has the highest number of elementary schools among other clusters
2	Baureno, Dander, Kalitidu, Kepohbaru, Ngasem, Padangan, Sumberejo	This cluster has the highest number of elementary school educators and the least number of junior high school educators. In addition, this cluster has the least number of junior high school learning groups
3	Bojonegoro	This cluster has the largest number of junior high schools, the largest number of high schools, the largest number of junior high school educators, the largest number of high school educators, the largest number of junior high school learning groups and the largest number of high school learning groups among other clusters
4	Balen, Bubulan, Gayam, Gondang, Kapas, Kasiman, Kedewan, Margomulyo, Ngambon, Ngraho, Sekar, Trucuk	This cluster has the least number of elementary schools, the least number of junior high schools, the least number of high schools and the least number of elementary educators compared to other clusters
5	Kanor, Malo, Purwosari, Sugihwaras, Sukosewu, Tambakrejo, Temayang	This cluster has the highest number of elementary school study groups among other clusters. In addition, this cluster also has the least number of high school educators

and the least number of high school learning groups compared to other clusters.

Based on the results, the characteristics of each cluster as shown in Table 5 can be used as input for the Bojonegoro district government for equal distribution of education in Bojonegoro. For example, cluster 2 has the highest number of elementary educators among other clusters, while cluster 4 has the least number of elementary educators among others. From these conditions, the number of elementary school educators in cluster 2 can be delegated and distributed to cluster 4. In addition, as another example, cluster 1 has the highest number of elementary schools among other clusters, while cluster 4 has the least number of elementary schools among other clusters. From these conditions, the government can reduce the number of elementary schools in cluster 1 and increase the number of elementary schools in cluster 4.

CONCLUSION

Based on the results of the study, it was found that the K-Means method is the best method to group sub-districts in Bojonegoro district based on educational supporting factors. The results of grouping using the K-Means method obtained 5 clusters. Cluster 1 consists of 1 sub-district, namely Kedungadem. Meanwhile, cluster 2 consists of 7 sub-districts, namely Baureno, Dander, Kalitidu, Kepohbaru, Ngasem, Padangan, and Sumberejo. For cluster 3, there is 1 sub-district, namely Bojonegoro. And for cluster 4 consisting of 12 sub-districts, namely Balen, Bubulan, Gayam, Gondang, Kapas, Kasiman, Kedewan, Margomulyo, Ngambon, Ngraho, Sekar and Trucuk. For cluster 5 there are 7 sub-districts, namely Kanor, Malo, Purwosari, Sugihwaras, Sukosewu, Tambakrejo, and Temayang. From the results of clusterization, profilization and characteristics of each cluster are obtained. It is hoped that these results can be used as input for the Education office as an effort to equalize education in Bojonegoro district.

REFERENCE

[1] I. Purwaningsih, O. Oktariani, L. Hernawati, R. Wardarita, and P. I. Utami, "Pendidikan sebagai Suatu Sistem," *J. Vision. Penelit. dan Pengemb. dibidang Adm. Pendidik.*, vol. 10, no. 1, pp. 21–26, 2022.

[2] P. S. Rosmana, S. Iskandar, N. Fadilah, N. Azhar, D. Oktavini, and A. C. Munte, "Upaya Pemerataan Pendidikan Berkelanjutan di Daerah 3T," *Attadib J. Elem. Educ.*, vol. 6, no. 2, pp. 405–418, 2023.

[3] K. Anwar, "Implementasi dan Relevansi Kebijakan Dalam Pemerataan Pendidikan: Studi Literatur Pelayanan Publik," *Coopetition J. Ilm. Manaj.*, vol. 13, no. 3, pp. 419–428, 2022.

[4] M. V. Roesminingsih, T. R. Hariastuti, and F. Agustina, "Perencanaan Peningkatan Mutu Sekolah di SMKN Purwosari Bojonegoro," *J. Pendidik. Tambusai*, vol. 6, no. 1, pp. 1892–1906, 2022.

[5] S. Satiti, "Gerakan Ayo Sekolah Di Kabupaten Bojonegoro: Peningkatan Sumber Daya Manusia Melalui Pendidikan Untuk Menyongsong Bonus Demografi," *J. Kependud. Indones.*, vol. 14, no. 1, pp. 77–92, 2019.

[6] H. Nurhuda, "Masalah-Masalah Pendidikan Nasional; Faktor-Faktor Dan Solusi Yang Ditawarkan," *Dirasah J. Pemikir. Dan Pendidik. Dasar Islam*, vol. 5, no. 2, pp. 127–137, 2022.

[7] A. A. Maulana, A. W. Rafii, Y. A. Anjelina, and E. Widodo, "Pengelompokkan Kecamatan di

Kabupaten Bima Berdasarkan Jumlah Produksi dan Luas Panen Bawang Merah Tahun 2021 Menggunakan K-Means Clustering," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 16, no. 1, pp. 442–451, 2023.

[8] F. Alfiah, A. Almadayani, D. Al Farizi, and E. Widodo, "Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020," *J. Ilm. Sains*, vol. 22, no. 1, pp. 1–7, 2022.

[9] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022.

[10] S. Rahayu and A. Y. Kartini, "ALGORITMA K-MEANS DAN K-MEDOIDS UNTUK PENGELOMPOKAN KECAMATAN PENERIMA BANTUAN SOSIAL DI KABUPATEN BOJONEGORO," *MEDIA BINA Ilm.*, vol. 16, no. 5, pp. 6815–6822, 2021.

[11] A. Y. Kartini and N. Cahyani, "HYBRID K MEANS-MULTIVARIATE ADAPTIVE REGRESSION SPLINES FOR DISTRIBUTION OF DENGUE FEVER RISK MAPPING IN BOJONEGORO DISTRICT," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 1, pp. 313–322, 2023.

[12] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny).*, 2022.

[13] Y. Mayona, R. Buaton, and M. Simanjutak, "Data Mining Clustering Tingkat Kejahatan Dengan Metode Algoritma K-Means (Studi Kasus: Kejaksaan Negeri Binjai)," *J. Inform. Kaputama*, vol. 6, no. 3, pp. 2548–9739, 2022.

[14] L. Qadrini, "Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 13, no. 2, pp. 5–11, 2020.

[15] D. F. Pramesti, M. T. Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, p. 964X, 2017.

[16] D. M. Sinaga, A. P. Windarto, and D. Hartama, "Analisis K-Medoids Dalam Pengelompokkan Rasio Murid dengan Guru , Murid dengan Rombel , dan Rasio Rombel dengan Kelas Jenjang Pendidikan SD dan SMP Menurut Provinsi," *J. Ris. Tek. Inform. dan Data Sci.*, vol. 1, no. 1, pp. 1–6, 2022.

[17] M. N. P. Pamulang, M. N. Aini, and U. Enri3, "Komparasi Distance Measure Pada K-Medoids Clustering untuk Pengelompokkan Penyakit ISPA," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 99–107, 2021, doi: 10.29408/edumatic.v5i1.3359.

[18] R. N. H. Hutasuhut, H. Okprana, and B. E. Damanik, "Penerapan Data Mining Untuk Menentukan Penerima Program Bidikmisi Menggunakan Algoritma K-Medoids," *TIN Terap. Inform. Nusant.*, vol. 2, no. 11, pp. 667–672, 2022.

[19] R. Anjariansyah and A. Triayudi, "Clustering Kebutuhan Makanan untuk Meminimasi Standar Deviasi Angka Kebutuhan Gizi Menggunakan Algoritma K-Means dan K-Medoids," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, pp. 597–607, 2022.

[20] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," *Qual. Quant.*, vol. 56, no. 3, pp. 1283–1291, 2022.

[21] M. R. Nugroho, I. E. Hendrawan, and P. P. Purwantoro, "Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI," *NUANSA Inform.*, vol. 16, no. 1, pp. 125–133, 2022.

[22] Sanusi and J. Husna, "Utilization of Rapidminer using the K-Means Clustering Algorithm for

Classification of Dengue Hemorrhagic Fever (DHF) Spread in Banda Aceh City," *J. Inotera*, vol. 5, no. 2 SE-Articles, pp. 146–151, Oct. 2020, doi: 10.31572/inotera.Vol5.Iss2.2020.ID119.

[23] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma k-means untuk pengelompokkan penyakit pasien pada puskesmas cigugur tengah," *J. Inf. Technol.*, vol. 2, no. 2, pp. 39–44, 2020.

[24] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.

[25] M. Herviany, S. P. Delima, T. Nurhidayah, and K. Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat: Comparison of K-Means and K-Medoids Algorithms for Grouping Landslide Prone Areas in West Java Province," *Malcom Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021.

[26] S. Ramadhani, D. Azzahra, and Z. Tomi, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 24–33, 2022.