DATA MINING STUDY FOR GROUPING ELEMENTARY SCHOOLS IN BOJONEGORO REGENCY BASED ON CAPACITY AND EDUCATIONAL FACILITIES

Denny Nurdiansyah[1*], Saniyatus Saidah[2], Nita Cahyani[3]

[1,2,3]*Statistics, Faculty of Science and Technology, Universitas Nahdlatul Ulama Sunan Giri
Jendral Ahmad Yani Street No.10, Bojonegoro, 62115, Indoonesia*

*Corresponding author's e-mail: * denny.nur@unugiri.ac.id*

ABSTRACT

The implementation of national education must ensure equitable distribution of educational facilities. However, based on data from the Regional Education Balance Sheet (NPD) in 2021, elementary schools in Bojonegoro District still need to meet the criteria for overall equality. It is mainly related to educational capacity and facilities. It is necessary to group elementary schools based on capacity and educational facilities to solve this problem by applying the clustering method. The research aims to conduct a comparative study of three clustering methods to get the best way to be used for clustering elementary schools in Bojonegoro Regency. This study applies three clustering methods, namely K-Means, K-Medoids, and Random Clustering, which are compared to get the best clustering method. The data used is secondary data representing educational capacity and facilities, namely the number of students, teachers, classrooms, and study groups (Rombel) from the Bojonegoro District Education Office. Obtained the resulting comparison of clustering methods with the best way falls on the K-Means method, which forms 5 clusters. It explained that elementary schools with educational capacity and facilities get highly complete 14 schools (cluster_3), complete 236 schools (cluster_2), fairly complete 176 schools (cluster_4), less complete 310 schools (cluster_1), and incomplete 177 schools (cluster_0). The conclusion that comparing Clustering methods obtained grouping of Elementary School data with the best way falls on the K-Means method by getting 5 clusters.

1. INTRODUCTION

Talking about the quality of education must be connected to the learning process in the classroom. Learning in the school includes two critical aspects, namely, teachers and students. Teachers have the task of teaching, while students have the purpose of learning. Good environmental conditions can have a positive impact on the teaching-learning process. An attractive environment will create a learning atmosphere in the school. The classroom's physical environment also influences elementary school students' development process. Environmental elements such as objects around students, including class size, room temperature, classroom environment, and student comfort, influence students' cognitive development [1]. A good education will produce a quality generation and become the key to improving the country's development. According to UNESCO, education in Indonesia ranks 10th out of 14 developing countries, where 21% of schools in urban areas, 37% in rural areas, and 66% in remote areas still lack teachers [2]. The most significant contributor to the low quality of education is undoubtedly from rural and remote areas. This is due to the need for more reliable educators willing to live in remote areas, the uneven distribution of teachers, and the uneven quality of teachers.

According to Law No. 2 of 1989 on the national education system, Elementary education is organized to develop attitudes and abilities. Elementary education is a nine-year education, consisting of a six-year program in Elementary school and a three-year education program in junior high school. Elementary education aims to provide students with essential skills to develop their lives as individuals, members of society, citizens, and members of humanity and to prepare students to follow education [3].

Improving a country's education quality must also begin with improving the quality of education in remote areas. Good quality education must be spread evenly in each region. The lack of equitable distribution of education occurs a lot in remote provinces, so there is much distribution of education subsidies that have not been evenly and thoroughly distributed [4].

Teachers are essential components in the teaching and learning process and participate in efforts to form potential human resources in the development field. According to experts, professional teachers are all people who have the authority and responsibility for the education of students, both individually and in groups, at school or outside school [1].

Students are members of society who seek to develop their potential through the learning process available at specific paths, levels, and types of education [5]. Students are one of the components of education that cannot be abandoned because, without students, it would not be possible for the learning process to run. Students are human components that occupy a central position in the teaching and learning process where students as parties who want to achieve goals, have goals, and then want to achieve them optimally.

The development of the world of education needs the latest innovations related to facilities tailored to educational capacity to improve the quality of education in Bojonegoro Regency. The development of capacity and facilities to advance civilization indicates that the world of education needs an upgrade to support education so that learning is carried out more effectively and efficiently. Capacity and facilities in the world of education have the advantage of making it easier to provide good comfort for teaching staff and students. A wide distribution of facilities is expected to improve today's quality of education. It is necessary to group elementary schools to solve the problem in the Bojonegoro district based on the completeness of capacity and educational facilities. The process of clustering elementary schools is statistically done with the clustering method. With the clustering of Elementary schools, policymakers can determine more efficient policies for improving Elementary education in the Bojonegoro district.

Based on data from the Regional Education Balance Sheet (NPD) in 2021, data on Elementary schools in the Bojonegoro district is presented. It is obtained with a total of 723 schools with 72,103 students and 19 dropouts, 5,041 teachers with 2,581 civil servants and 2,460 non-civil servants, 4,585 classrooms, and 4,853 classrooms with 1,456 in good condition and 3,397 in a lightly damaged condition [6].

In the Bojonegoro sub-district, there are Elementary schools with less than five new students. It is such as Banjarejo 3 Elementary School, which has only two new students, Klangon 1 Elementary School has only three students, and Ledok Kulon 3 Elementary school has only five new students. The small number of new students is due to the need for more school-age children in the neighborhood and the proximity between elementary schools. The elementary schools concerned have requested the Education Office policy for equal distribution of students in elementary schools so that all schools can get students according to the room's capacity and teachers [7].

As of 2021, four elementary school buildings have collapsed, including Semanding Elementary School, Pandan 2 Elementary School, Besah 1 Elementary School, and Sumberagung 2 Elementary School, Ngraho. The collapsed buildings that have yet to be repaired have resulted in students and teachers having to carry out teaching and learning activities (KBM) alternately in other classrooms [8].

Based on research conducted by Winarti et al. (2021) related to teacher professionalism and the quality of education at Mojodelik 1 State Elementary School with a qualitative approach shows that, in general, teachers are professional in mastering the material, concept structure, and scientific mindset [9]. So far, educational research in Bojonegoro Regency still needs to be improved, so it is a critical research area.

In this research, the aim is to compare clustering methods to get the best method to be used for clustering elementary schools in Bojonegoro Regency based on educational capacity and facilities. The clustering methods used are K-Means, K-Medoids, and Random clustering.

The clustering method is appropriate in dividing or grouping elementary schools in Bojonegoro Regency based on educational capacity and facilities. The clustering method is the process of grouping some data based on the characteristics of the data obtained. It can explain the relationship between one member and another by grouping the members' similarities in one group and making a difference with other groups regarding their respective data characteristics [10].

K-Means research conducted by Priambodo and Prasetyo (2018) obtained the results of mapping the Banten Province area based on the level of education that has a shortage, sufficiency, and excess of teachers according to the district/city [2]. The benefit of the results of this study is that it can be a suggestion for the Banten Provincial Education Office in terms of teacher equity. Furthermore, K-Means research by Pradanyana and Permana (2018) concluded that the number of clusters and the amount of data used affect the quality of clusters formed by the K-Means and K-Nearest Neighbors methods used [11]. From the K-Means research conducted by Rangan et al. (2018), the resulting accuracy measure of the Support Vector Machine method with an accuracy rate of 93.33%, while the K-means clustering method with Support Vector Machine (SVM) support obtained an accuracy rate of 99.33% [12]. In the research of Nugraha and Hairani (2018), the K-Means Clustering method is used by displaying the results in the form of a map so that the education office or institutions that handle education in Indonesia can compare the quality of education in each province in Indonesia [13]. Anggraeni and Putra's research (2021) stated that the K-Means clustering method is a method with optimal cluster accuracy results. However, the use of this method is longer with the determination of the centroid, which must be accurate so that the output results are as expected [14].

In the research of Astria et al. (2020), the K-Medoids method is used to group data with the help of RapidMiner software so that 2 clusters are obtained [15]. The research of Rhamadani et al. (2020) obtained clustering results using K-Means and K-Medoids algorithms well. There are differences in the number of clusters in the performance of each algorithm. Each iteration has a different calculation pattern depending on the dataset used and the centroid points used as calculations in the algorithm [16]. From the research of Silitonga et al. (2019), the K-Medoid method was chosen because it can overcome the shortcomings and weaknesses in other methods, especially the K-Means method, where this method can overcome the weaknesses in the data, including outlier data and noise [17].

In the research of Gates and Ahn (2019), the Random Clustering method is used for clustering Cancer types where the clustering process is more like Hierarchical Clustering than random Clustering [18]. This method is considered easier to implement and works for real cluster weights q > 0 with results for several q that can be found at once in one simulation. However, this method must improve accuracy for large values when q > 1. Random Clustering performs random flat Clustering on observational data, which groups object that are similar to each other and different from objects that belong to other clusters [19].

In this study, the best clustering model criteria used a principle similar to the Elbow method in Putu et al.'s research (2021), which measures cluster density performance and produces an average within cluster distance measure with the help of Rapidminer software for each clustering method experiment on clusters 2 to 10. The best cluster k is selected at the average within-cluster distance value, close to zero, on the first stable sloping curve [20].

In the field of education in Bojonegoro Regency, it is still considered less targeted because it needs to know which groups of elementary schools need attention related to increasing educational capacity and facilities. Grouping elementary schools with clustering methods can determine groups of elementary schools based on educational capacity and facilities in the Bojonegoro district. This research benefits the Bojonegoro

education office in supporting equalizing educational capacity and facilities for all elementary schools. Thus, it is necessary to propose research titled "Data Mining Study for Grouping Elementary Schools in Bojonegoro Regency Based on Educational Capacity and Facilities".

2. RESEARCH METHODS

2.1 Research Design

The research design applied is a quantitative research approach that uses three clustering methods, namely K-Means, K-Medoids, and Random Clustering, to be compared. The clustering method is applied using Rapidminer Software with the best method criteria measured based on cluster density performance, producing an average within cluster distance measure.

2.2 Population and Sample

The sample in this study represents the observed population. It uses secondary data representing educational capacity and facilities, namely data on the number of teachers, number of students, number of classrooms, and number of rombel of an elementary school in Bojonegoro district taken from the Ministry of Education, Culture, Research and Technology in 2021.

2.3 Sampling Technique

The sampling technique used in this research uses a purposive sample that takes data from the database according to the research objectives. As the case study application, the education data sample was taken from the official website of the Ministry of Education, Culture, Research, and Technology.

2.4 Research Subjects

The data sources used in this study are secondary data in the form of data on students, teachers, rombel, and classrooms of elementary schools in Bojonegoro Regency obtained from the website of the Ministry of Education, Culture, Research and Technology in 2021. The research variables used in this study are presented in **Table 1**:

Table 1. Definition Of Research Variables for Clustering Method

Variables	Measurement Scale	Description
Teacher	Ratio	The number of teachers teaching in the school.
Student	Ratio	The number of students attending lessons.
Classroom	Ratio	The number of classrooms in the school.
Rombel	Ratio	The number of Rombel given by the school.

The research variables used are Teacher, Student, Classroom, and Rombel variables, the description of which has been presented in **Table 1**. All variables have a measurement scale, that is, the ratio.

2.5 Data Analysis Technique

The analysis step in this research is to compare three clustering methods, including K-Means, K-Medoids, and Random Clustering, which can be done with the help of RapidMiner software. The application of clustering methods with Rapidminer software is made by making a process design arranged in **Figure 1** with several tools such as: Retrieve data, Normalize, Multiply, Method (K-Means), Data to Similarity, multiply second, and Performance.
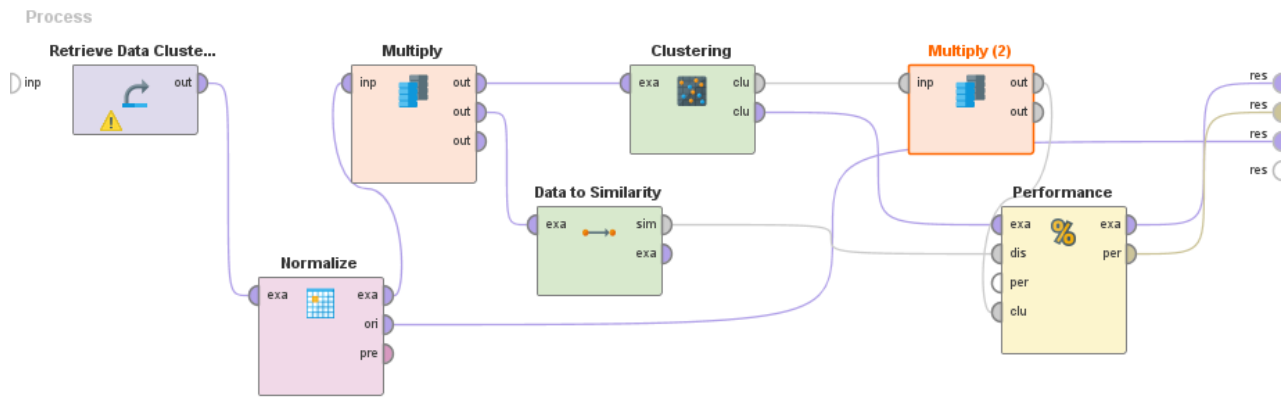
Process

Figure 1. K-Means clustering process design

In **Figure 1**, the tools or operators are given in the research. Retrieve data is used to input the dataset. Normalize to standardize is used for the observation data in a standard Normal distribution. Multiply is given to divide the process into more than one process. Method (K-Means) is used to apply the K-Mean clustering method. Data to Similarity accounts for the Similarity between examples of a dataset. Performance is used to measure the Performance of an applied method.

The Retrieve data tool helps open datasets, Normalize to standardize data, and Multiply to divide the process of 1 path into more than one path. The method can be selected for Clustering methods such as K-Means, K-Medoid, and Random Clustering methods, Data to Similarity serves to anticipate data repetition. Performance is used in testing the Performance of the Method with the provisions of Cluster density performance.   The steps of data analysis in this study are given in stages:

1. Describe the Elementary database of Elementary School education in the Bojonegoro district.
2. Evaluate the Elementary School education database in the Bojonegoro district to obtain data on the capacity and facilities of elementary schools in the Bojonegoro district.
3. Perform clustering analysis by applying the K-Means, K-Medoid, and Random Clustering methods.
4. The clustering method comparison process is carried out by comparing the value of the performance vector with the best cluster k based on the value of the Average Within Cluster Distance of the centroid-based K-Means, K-Medoid, and Random Clustering methods.
5. Determine the best clustering model with similar criteria to the Elbow method for clustering elementary schools in the Bojonegoro district.
6. Clustering Elementary Schools based on capacity and facilities is done by applying the best clustering method.

2.6 Flow Chart

In this research, a flow chart can provide a detailed description of the data analysis steps in this study. The flow chart is used following the visualization in **Figure 2**:
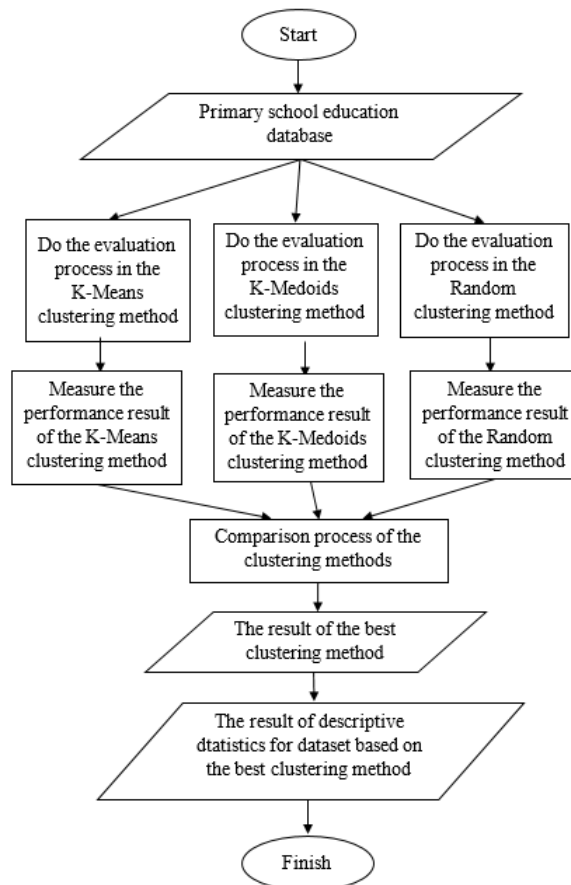
Figure 2. **Flowchart of clustering comparison**

Figure 2 gives a visualization of the stages of comparison between three clustering methods: the methods of K-Means, K-Medoids, and Random Clustering. Starting with the input dataset as a primary school education database and then evaluating the three clustering methods. The process continues by measuring the performance results of the three clustering methods and then comparing the results of the three clustering methods. The output given is the result of the best clustering method, then the output for the descriptive statistics of the dataset based on the grouping of the best clustering method.

3.   RESULTS AND DISCUSSION

3.1.   Results

In this study, the amount of data used is 723 in the form of elementary schools in Bojonegoro Regency regarding the number of teachers, students, classrooms, and rombel in 2021. Elementary School data is taken in the vulnerable months of January - December 2021. The data criteria used consist of the number of Students, Teachers, Classrooms, and Rombel of Elementary Schools in Bojonegoro Regency. After the data selection process, 723 datasets were obtained, which are summarized in the descriptive statistics table in **Table 2**.

Table 2. **Descriptive Statistics For Elementary School Data In Bojonegoro District.**

Statistic	Student	Teacher	Classroom	Rombel
Number of data	723	723	723	723
Minimum	2	1	0	1
Maximum	632	22	24	24
Total	72,103	5,041	4,853	4,585
Mean	99.057	6.957	6.469	6.342
Standart Deviation	69.014	2.293	2.129	1.601

Table 2 obtained descriptive statistical results for elementary school data in Bojonegoro Regency, a total of 723 elementary schools, with the lowest value of students being two located at Soko III Temayang State Elementary School and the highest value of 632 located at Muhammadiyah 2 Bojonegoro Elementary School with the total number of students in Bojonegoro as many as 72,103 with an average of 99.06 and a standard deviation of 69.014. The second variable is teachers, with the lowest value of 1, located in Bumiayu Baureno Public Elementary School, and the highest value is 22, located in Luqman Al-Hakim Bojonegoro Integral Elementary School. The total number of Bojonegoro Elementary School teachers recorded in Dapodik is 5,041, with a mean value of 6.96 and a standard deviation of 2.293. For the third variable, the classroom has the lowest value of 0. It is located in Nurul Ulum Ngemplak Baureno Elementary School, Ngablak II Dander State Elementary School, and Mori II Trucuk State Elementary School. The highest value classroom is 24, located in Luqman Al-Hakim Bojonegoro Integral Elementary School and Muhammadiyah 2 Bojonegoro Elementary School. Classrooms totaled 4,853 with an average value of 6.47, while the standard deviation value was 2,129. The fourth variable is Rombel, with the lowest value of 1 in Soko III Temayang State Elementary School and the highest value of 24 located in Muhammadiyah 2 Bojonegoro Elementary School. Rombel in Bojonegoro Elementary Schools amounted to 4,585, with an average value of 6.34 and a standard deviation of 1,601. The descriptive statistics for Elementary School data in Bojonegoro district.

Outliers will affect the mean value in the observation data so that the median will replace the centering value. The presence of outliers in the observation data is indicated by the Z-score value outside the Normal curve ($-3 <$ Z-score < 3).

Table 3. Descriptive Statistics For Z-Score Calculation Results.

Statistic	Student	Teacher	Classroom	Rombel
Number of data	723	723	723	723
Minimum	-1.4063	-3.0387	-3.3364	-2.5981
Maximum	7.7222	8.2351	11.0296	6.5608
Total	0	0	0	0
Mean	1	1	1	1
Standart Deviation	723	723	723	723

Table 3 provides descriptive statistics related to the results of the calculation of the Z-score value. It is obtained from 723 observations that the Student, Teacher, Classroom, and Rombel variables have many outliers in the distribution of observation data. It happened because the Z-score value of each variable has a value above three, so outliers are identified. Observational data can also be detected for outliers with the following Box-plot diagram visualization:
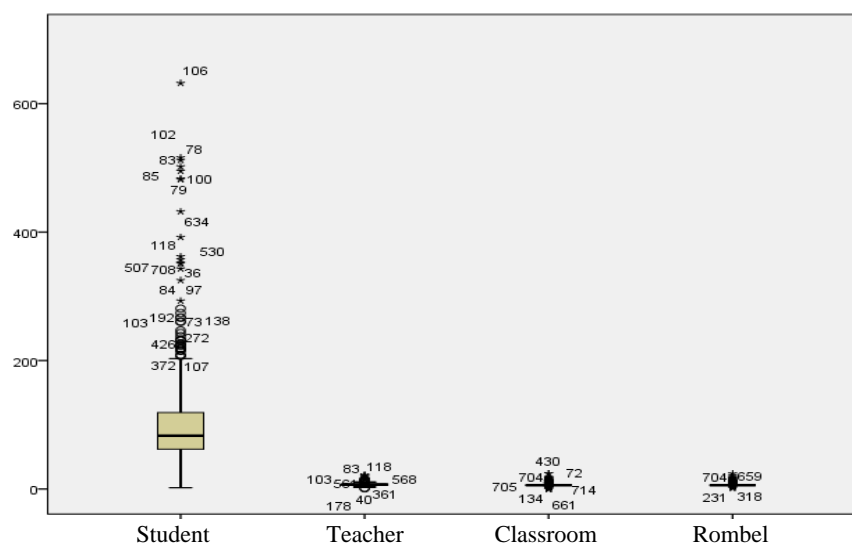

Figure 3. Boxplot results of research variables.

Figure 3 shows the visualization of the outliers detection using Boxplot. Outliers are identified when the observation data points are above or below the boxplot diagram given to each research variable.

The Boxplot identification results lead to the conclusion that there are very many outliers in all research variables.

The performance vector value with the best results from each clustering method is used in the clustering method comparison process. Based on the clustering method evaluation process of the centroid-based K-Means, K-Medoids, and Random Clustering methods, the comparison results of the three clustering methods are described in **Table 4**.

Table 4. Comparison Of Clustering Methods

Cluster (k)	K-Means Clustering	K-Medoids Clustering	Random Clustering
2	-26324.474	-14278.866	-2272.325
3	-11480.223	-10599.635	-1524.329
4	-6823.103	-4590.788	-11333.53
5	-3569.258	-4321.351	-9121.728
6	-2217.916	-2754.42	-7639.181
7	-1511.311	-1666.381	-6465.694
8	-1189.554	-1324.929	-5653.377
9	-867.843	-1082.863	-5055.305
10	-840.101	-988.359	-4541.986

Table 4 provides the results of the comparison of the three clustering methods based on the Average Within Cluster Distance value. The best group k value is obtained from observing the movement of the average within-cluster distance value which slopes first to the minimum value where the best value falls on k equal to 5. The best cluster is selected from the average within-cluster distance value that rises the first stable slope, namely in cluster 5 with a K-Means performance value of -3569.258, K-Medoid of -4321.351, and Random Clustering of -9121.728. This result will be seen in the visualization of the multi-line diagram in **Figure 4**.
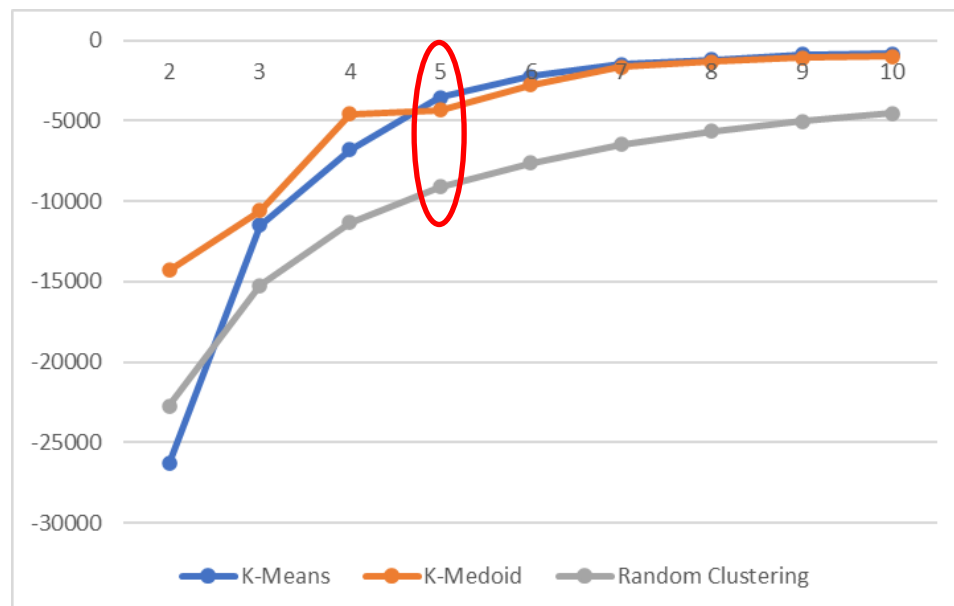

Figure 4. The multi-line diagram from the average within-cluster distance value of clustering methods

Figure 4 shows the visualization of multiple lines that give the movement of the average within-cluster distance value. The best cluster is selected from the average within-cluster distance value with the first stable slope increase, which is cluster 5 with K-Means. The best clustering result for clustering elementary schools in Bojonegoro Regency is the K-Means clustering method. For group categorization, a summary of descriptive statistics is given in **Table 5**.

Table 5. **Descriptive Statistics for Each Cluster From K-Means Clustering.**

Cluster_0	Number of data	Minimum	Maximum	Mean	Standard Deviation
Student	177	2	61	42.03	13.41
Teacher	177	1	10	5.76	1.549
Classroom	177	0	9	5.55	1.201
Rombel	177	1	6	5.89	0.542
Cluster_1	**Number of data**	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**
Student	310	62	105	80.91	12.37
Teacher	310	3	10	6.65	1.467
Classroom	310	0	12	6.07	0.764
Rombel	310	6	6	6	0
Cluster_2	**Number of data**	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**
Student	236	106	632	165.657	83.4256
Teacher	236	3	22	8.26271	2.93309
Classroom	236	0	24	7.6822	3.12244
Rombel	236	6	24	7.13136	2.59207
Cluster_3	**Number of data**	**Minimum**	**Maximum**	**Mean**	**Standart Deviation**
Student	14	343	632	443.643	87.0611
Teacher	14	11	22	16.5	3.45855
Classroom	14	12	24	16.4286	3.77674
Rombel	14	12	24	15.5	3.65324
Cluster_4	**Number of data**	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**
Student	176	106	172	129.875	16.7795
Teacher	176	3	12	7.23295	1.59544
Classroom	176	4	14	6.52273	1.39777
Rombel	176	6	8	6.09091	0.32554

Based on **Table 5**, the clustering results using the K-Means method based on the capacity and completeness of elementary school facilities are divided into 5 clusters. For the first group, cluster_0 consists of 177 elementary schools with a number of students with a minimum of 2, a maximum of 61, and an average of 42.03; The number of teachers has a minimum of 1, a maximum of 10, and an average of 5.76; a minimum number of classrooms of 0, a maximum of 9, and an average of 5.55; and a minimum rombel of 1, a maximum of 6, and an average of 5.89. For the second group, cluster_1 consists of 310 elementary schools with a number of students a minimum of 62, a maximum of 105, and an average of 80.91; The number of teachers has a minimum of 3, a maximum of 10, and an average of 6.65; The number of classrooms with a minimum value of 0, a maximum value of 12, and an average value of 6.07; and The number of classrooms with a minimum of 6, a maximum of 6, and an average of 6. For the third group, cluster_2 consists of 236 elementary schools with a number of students minimum of 106, a maximum of 632, and an average of 165,657; a minimum number of teachers of 3, a maximum of 22, and an average of 8.26271; a minimum number of classrooms of 0, a maximum of 24, and an average of 7.6822; and a minimum number of classrooms of 6, a maximum of 24, and an average of 7.13136. For the fourth group, cluster_3 consists of 14 elementary schools with a number of students with a minimum of 343, a maximum of 632, and an average of 443,643; a minimum number of teachers of 11, a maximum of 22, and an average of 16.5; a minimum number of classrooms of 12, a maximum of 24, and an average of 16.4286; and a minimum number of classrooms

of 12, a maximum of 24, and an average of 15.5. For the fifth group, cluster_4 consists of 176 elementary schools with a number of students with a minimum of 106, a maximum of 172, and an average of 129,875; a minimum number of teachers of 3, a maximum of 12, and an average of 7.2395; minimum number of classrooms of 4, maximum of 14, and an average of 6.52273; and minimum number of classrooms of 6, maximum of 8, and an average of 6.09091.

3.2.   Discussion

In **Table 2**, the variant value of each observation variable has a reasonably diverse value, so the data distribution is not normal. The standard deviation of all variables is enormous because the average is also huge. The results of identifying outlier values in the observation data are obtained, but in applying the K-Median method, the outlier values can be adequately resolved. Figure 3 shows that the Box-plot output for the variables of Students, Teachers, Classrooms, and Rombel found outliers. The most significant difference between the minimum and maximum values is shown from the Student's variable, so this variable needs more attention from decision-makers.

In evaluating the clustering method, the resulting performance vector value of the clustering method in the K-Means, K-Medoids, and Random Clustering methods occurs in selecting cluster k of 5. It is also following the equalization of the capacity and facilities of elementary schools, namely: Very Complete, Complete, Quite Complete, Incomplete, and Incomplete. In comparing clustering methods in **Table 4**, the best clustering method is obtained with the performance vector value closest to 0 and the stability of the results, which falls on the K-Means method. The results of the clustering application produce clusters of elementary schools with almost the same amount of capacity and facilities in each cluster.

In **Table 5**, cluster_0 consists of 177 elementary schools with the lowest average number of students, the lowest average number of teachers, the lowest average number of classrooms, and the lowest average number of rombel. Thus cluster_0 includes the capacity and facilities of elementary schools classified as Incomplete. In cluster_1, there are 310 elementary schools with the second lowest average students, second lowest average teachers, second lowest average classrooms, and second lowest average number of rombel. Thus cluster_1 includes the capacity and facilities of elementary schools classified as less complete. From cluster_2, there are 236 elementary schools with the second highest average number of students, the second highest average number of teachers, the second highest average number of classrooms, and the second highest average number of rombel. Thus cluster_2 includes the capacity and facilities of elementary Schools classified as Complete. In cluster_3, there are 14 elementary schools with the highest average number of students, the highest average number of teachers, the highest average number of classrooms, and the highest average number of rombel. Thus cluster_3 includes the capacity and facilities of elementary schools classified as Highly Complete. From cluster_4, there are 176 elementary schools with the third highest average students, third highest average teachers, third highest average classrooms, and third highest average rombel. Thus cluster_4 includes elementary school capacities and facilities classified as Fairly Complete.

4.   CONCLUSIONS

Based on the results that researchers have carried out, the following conclusions can be obtained:

1. Data on the capacity and completeness of elementary school facilities is obtained with a large number of outliers with maximum and minimum values with a huge difference so that the observation data is far from the Normal distribution.
2. From the comparison results of Clustering methods, it is obtained that the grouping of Primary School data with the best method falls on the K-Means method by getting 5 clusters.
3. The group of schools with highly complete primary school capacity and facilities is 14 schools (cluster_3), complete 236 schools (cluster_2), fairly complete 176 schools (cluster_4), less complete 310 schools (cluster_1), and incomplete 177 schools (cluster_0).

ACKNOWLEDGMENT

REFERENCES

[1] C. P. Pratiwi and S. Ediyono, "Analisis Keterampilan Guru Sekolah Dasar Dalam Menerapkan Variasi Pembelajaran," *J. Sekol.*, vol. 4, no. 1, pp. 1–8, 2019.

[2] Y. A. Priambodo and S. Y. J. Prasetyo, "Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Metode Spatial Clustering K-Means (Studi kasus : Wilayah Provinsi Banten)," *Indones. J. Comput. Model.*, vol. 1, no. 1, pp. 18–27, Jan. 2018, doi: 10.24246/j.icm.2018.v1.i1.p18-27.

[3] P. R. Indonesia, *Undang-Undang Nomor 2*, no. 2. 1989, pp. 1–17. [Online]. Available: https://peraturan.bpk.go.id/Home/Details/46794/uu-no-2-tahun-1989

[4] S. Suryana, "Permasalahan Mutu Pendidikan Dalam Perspektif Pembangunan Pendidikan," *Edukasi*, vol. 14, no. 1, May 2020, doi: 10.15294/edukasi.v14i1.971.

[5] P. R. Indonesia, *Undang-Undang Nomor 20*, no. 20. 2003, pp. 1–26.

[6] Kemendikbud, "Neraca Pendidikan Daerah 2021," 2022. https://npd.kemdikbud.go.id/?appid=download (accessed Jan. 10, 2023).

[7] Y. M. Purwanto, "Empat Gedung SD di Bojonegoro Ambruk, Hanya Satu Diperbaiki Tahun Ini," *radarbojonegoro*, 2022. https://radarbojonegoro.jawapos.com/daerah/bojonegoro/31/03/2022/empat-gedung-sd-di-bojonegoro-ambruk-hanya-satu-diperbaiki-tahun-ini/ (accessed Jan. 10, 2023).

[8] Y. M. Purwanto, "Tiga SDN di Kota Dapat Siswa Kurang dari Lima Anak," *radarbojonegoro*, 2022. https://radarbojonegoro.jawapos.com/daerah/bojonegoro/20/07/2022/tiga-sdn-di-kota-dapat-siswa-kurang-dari-lima-anak/ (accessed Jan. 10, 2023).

[9] Winarti, C. S. Hartati, and N. M. Wibowo, "Kompetensi Profesional Guru Dalam Peningkatan Mutu Pendidikan Di SD Negeri Mojodelik I Kecamatan Gayam, Kabupaten Bojonegoro," *JMP Online*, vol. 5, no. 2, pp. 155–164, 2021.

[10] L. Rahmawati, S. Widya Sihwi, and E. Suryani, "Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, FMIPA, Universitas Sebelas Maret)," *J. Teknol. Inf. ITSmart*, vol. 3, no. 2, pp. 66–73, Aug. 2014, doi: 10.20961/its.v3i2.654.

[11] G. A. Pradnyana and A. A. J. Permana, "Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode K-Means Dan K-Nearest Neighbors Untuk Meningkatkan Kualitas Pembelajaran," *JUTI J. Ilm. Teknol. Inf.*, vol. 16, no. 1, pp. 59–68, Jan. 2018, doi: 10.12962/j24068535.v16i1.a696.

[12] A. Y. Rangan, M. I. Ukkas, and S. Qomariah, "Implementasi Metode Support Vector Mechine Dan K-Means Clustering Dengan Support Vector Mechine Tema Karya Ilmiah Pada STMIK Widya Cipta Dharma," in *SNITT (Seminar Nasional Inovasi Teknologi Terapan) Poltekba 2018*, 2018, vol. 3, pp. 274–279.

[13] G. S. Nugraha and H. Hairani, "Aplikasi Pemetaan Kualitas Pendidikan di Indonesia Menggunakan Metode K-Means," *J. MATRIK*, vol. 17, no. 2, pp. 13–23, May 2018, doi: 10.30812/matrik.v17i2.84.

[14] D. Anggraeni, R. Rizaldi, and G. M. Putra, "Penerapan K-Means Clustering Untuk Pengelompokan Kelas Pada Taman Kanak-Kanak," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 400–404, Dec. 2021, doi: 10.47065/bits.v3i3.1125.

[15] C. Astria, D. Hartama, A. P. Windarto, and I. Sudahri, "Pengembangan Metode Datamining K-Medoid Pada Kasus Distribusi Listrik di Indonesia," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2020, pp. 276–281.

[16] N. N. Rhamadani, A. Fauzi, E. Nurlaelasar, and A. R. Pratama, "Implementasi Algoritma K-Means Dan K-Medoids Dalam Pengelompokan Nilai Ujian Nasional Tingkat SMK," in *CIASTECH 2020 "Peranan Strategis Teknologi Dalam Kehidupan di Era New Normal,"* 2020, no. 3, pp. 717–726.

[17] S. Silitonga, E. Irawan, S. Saifullah, M. R. Lubis, and I. Parlina, "Pengelompokan Nilai Akademik untuk Menentukan Kenaikan Kelas Menggunakan Algoritma K-Medoids," in *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, Sep. 2019, vol. 1, pp. 740–749. doi: 10.30645/senaris.v1i0.80.

[18] A. Gates and Y.-Y. Ahn, "CluSim: a python package for calculating clustering similarity," *J. Open Source Softw.*, vol. 4, no. 35, pp. 1–5, 2019, doi: 10.21105/joss.01264.

[19] Rapidminer, "Random Clustering: RapidMiner Studio Core," 2022. https://docs.rapidminer.com/latest/studio/operators/modeling/segmentation/random_clustering.html (accessed Jan. 13, 2023).

[20] N. Putu, M. Novita, I. Bagus, and G. Dwidsamara, "Implementation of K-Modes Algorithm for Clustering of Stress Causes in University Students," *J. Elektron. Ilmu Komput. Udayana*, vol. 9, no. 3, pp. 419–428, 2021.