COMPARISON PERFORMANCE OF K-MEDOIDS AND K-MEANS ALGORITHMS IN CLUSTERING COMMUNITY EDUCATION LEVELS

Diana Dwi Aulia[1], Nurahman[2]

[1,2]Sistem Informasi, Universitas Darwan Ali Sampit, Indonesia

email: dianadwiaulia@gmail.com[1], nurrahman.ikhtiar@gmail.com[2]

Abstract

Education is a mandatory right of all citizens and the key to the nation's superiority in global competition that must get top priority to be examined critically and comprehensively. It is known that compulsory education is at least 12 years, but not all people can do it because of minimal economic conditions. In past years, COVID-19 has also had an impact on the economy, school dropout rates, and falling academic achievement, for example in Central Kalimantan. The size of Central Kalimantan, however, makes it difficult for the government to identify the areas with the worst levels of education. To determine which regions fall into the low and high education categories, it is required to group the province's educational levels. This study also compares two algorithms by measuring their accuracy. By looking at which algorithm has the lowest Davies Bouldin Index (DBI) value, the best degree of performance can be ascertained. To process the data from as many as 1,565 sources, data mining techniques, including the clustering method, were used. K-Means and K-Medoids algorithms were employed in this work as clustering techniques. Based on the outcomes of the cluster created, both algorithms are also put to the test for performance. The results of this study obtained 6 clusters in K-Means with the lowest DBI value of -0.439, while the results in K-Medoids were in 3 clusters with the lowest DBI of -0.866. Based on accuracy testing using DBI, it is known that K-Means results are more optimal with the lowest DBI value in the grouping of education levels compared to K-Medoids. It is also known from the formation of 6 clusters of the K-Means algorithm that the low education level is in cluster_0 which is 1484 villages and the higher education level is as many as 3 villages in cluster_3.

Keywords : Education, Data mining, Clustering, K-Means, K-Medoids.

INTRODUCTION

Education is a fundamental need for every human being and a crucial component in achieving success and leading a good life in the future[1]. Science and education can be acquired through learning in schools, as schools facilitate the realization of our ideals and future success [2]. But in Central Kalimantan, a lot of parents still struggle with financial barriers that prevent them from pursuing a higher degree. As a result, not all children who are of compulsory school age participate equally in academic learning. Since many businesses and organizations require at least a high school certificate for job openings, the 12-year compulsory education program seeks to lower the community's unemployment rate[3].

Central Kalimantan is the province with the second-largest land area in Indonesia. However, it still faces numerous challenges that contribute to the decline in education. The decrease in public education in Central Kalimantan is mainly attributed to limited access to educational services, particularly in remote areas or impoverished neighborhoods. Factors such as distance, costs, and inadequate infrastructure, including a lack of educational media and teaching materials, hinder access to education[4]. In some cases, community education programs can be hampered by a lack of government support and limited resources. The decline in public education is also caused by insufficient government budget allocation and management institutions, such as scholarships for underprivileged and outstanding children funded by the provincial and state budgets. These scholarships have limited quotas and are unable to reach all students from different regions. As a result, the educational programs conducted suffer from suboptimal facilities and the quality of teaching staff. Furthermore, the decline in education can be attributed to a lack of understanding and interest among the community, as they may not have sufficient

information about the importance of education or the benefits derived from educational programs. Without proper enforcement, the quality and services of education will inevitably decrease. Therefore, various parties involved in education provision and management must address this issue. To assist the government in identifying areas that require special attention, this research aims to group villages based on their education levels[5].

It is required to carry out village attribute clustering in Central Kalimantan after determining the challenges of schooling in this region. Clustering is used to identify which communities need to enhance their educational systems, including how students will receive financial aid and where schools will be located[6]. With so many villages, it will be challenging for the government to pinpoint regions with poor levels of education. As a result, the population must be grouped according to educational level. Two algorithms—K-Means and K-Medoid—will be used for this clustering process [7]. These algorithms are being used by researchers to compare their performance in accuracy tests[8].

The popular clustering algorithm K-Means is frequently used to group data. In comparison to other clustering approaches, it is preferred for its simplicity and ease of implementation[9]. K-Means has the benefit of quicker computation times. The algorithm attempts to divide data sets into many clusters, grouping related data into one cluster while assigning distinct data to other clusters [10]. K-Medoids, which are comparable to K-Means in an application, are another common clustering approach. The selection process for the cluster centroids or centers is where the biggest distinction lies. The average of all the data points in the cluster is used to calculate centroids in the K-Means technique. The representative data points chosen by K-Medoids, in contrast, are those that are most comparable to the other data points in the cluster[11][12]. To determine the shortest mean distance across data groupings, both algorithms use the Euclidean Distance function. The K-Means algorithm determines the average distance of all the data points inside each cluster, whereas K-Medoids chooses representative data points to decrease distances between data points[13].

Previous research has already been conducted on clustering methods, specifically utilizing the K-Means algorithm, to cluster the levels of community education in the Kapuas district. This study incorporates village attributes and encompasses a wide range of education levels, including non-school attendance, elementary/SD, junior high/SMP, vocational/SMA, and tertiary education. The data analysis resulted in the formation of eight distinct education-level clusters, comprising a total of 229 records. Notably, the villages with the lowest education levels were identified within Cluster 1, comprising 33 villages [14]. Moreover, another study employed the K-Means algorithm to cluster education levels in DKI Jakarta. This particular study utilized attributes such as sub-districts, elementary schools/SD, junior high schools/SMP, high schools/SMA, and tertiary institutions. The findings yielded educational data categorized into the highest and lowest groups. Cluster 2 represented the lowest education level, while Cluster 0 represented the highest education level[15].

Further research was conducted on grouping and comparison algorithms for COVID-19 cases in Indonesia. The study focused on two attributes: death cases and confirmed cases. The analysis involved comparing the effectiveness of the K-Means and K-Medoids algorithms for clustering the spread of COVID-19 in Indonesia. The K-Means algorithm produced a range of values from K2 to K9, with the smallest DBI value observed at K5, resulting in an accuracy value of 0.064. On the other hand, the K-Medoids algorithm yielded a value of 0.411 at K-2. Based on the results of both clustering methods, it can be concluded that the K-Means algorithm is more effective in grouping the reported cases [16].

Some of the previous studies above are different from the research that has been done now. This study introduces a novelty approach by clustering based on the level of education, utilizing various attributes such as villages and education levels that encompass Not school, elementary/SD, junior high/SMP, high school/SMA, D1, D2, D3, S1, S2, and S3. Another notable distinction in this study is the evaluation of accuracy using the Davies Bouldin Index method. This evaluation helps determine the optimal level of performance among the formed clusters, striving for a more refined and efficient clustering process. Additionally, while previous studies focused on grouping education levels solely with the K-Means algorithm, this research employs two algorithms, namely the K-Means algorithm and K-Medoids, to facilitate a comparison of their accuracy in clustering, assessed through the performance value of the Davies Bouldin Index.

This study aims to apply two clustering methods, namely the K-Means algorithm and K-Medoids, to cluster village data based on the

level of education of the community in Central Kalimantan Province. In this context, the use of K-Means and K-Medoids algorithms aims to compare the performance of both and determine which algorithm provides the best results. Performance is known after processing data from each algorithm and testing accuracy using the performance of the Davies Bouldin Index. After knowing which algorithm is better, the results of the more optimal DBI algorithm will be used as a consideration for approximately which villages have high and low education levels. The results of the grouping are expected to make it easier for the government or the Ministry of Education and Culture to distribute educational assistance to students and school facilities.

METHOD

Methods in research serve as a means to streamline the course of targeted and structured research activities. The research methodology involves designing a research flow that encompasses the entire process from the initial stage to completion [17]. The initial stages of this study include Problem Analysis, Data Collection, Data Preprocessing, Clustering Process using K-Means and K-Medoids algorithms, and Accuracy Testing using the DBI performance metric. The stages of this study are illustrated in Figure 1 below.
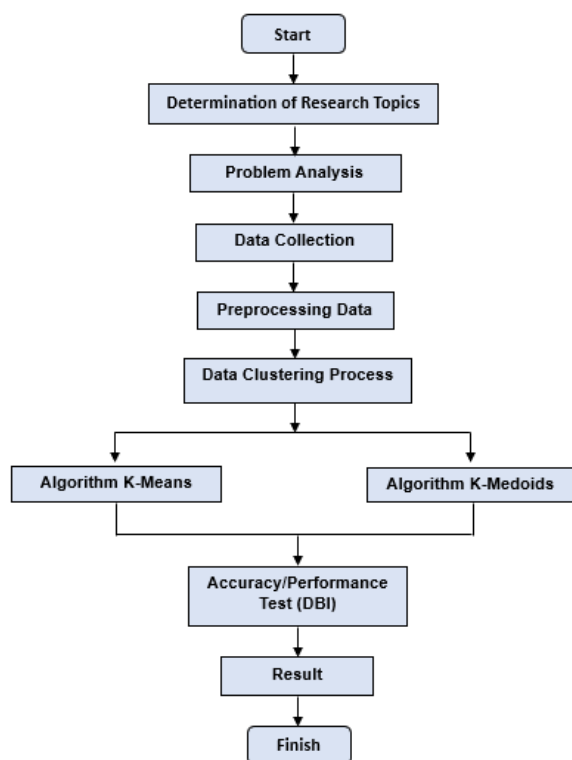

Figure 1. Research Methodology Framework

1. Determination of Research Topics

The first stage in this study is to determine the research topic to be carried out. The research topic was raised because it was known that the previous COVID-19 pandemic had a lot of impacts on society, such as an increase in out-of-school children and a decrease in learning achievement. Therefore, the research topic taken is Community education.

2. Problem Analysis

The next stage is to analyze the problems regarding education in Central Kalimantan Province. The activities carried out at this stage seek information from the Kalteng website regarding the main problems of education that take place. Where the education is still not optimal in the quality of learning in th e classroom, which is in the form of limited teaching materials. And there is a gap in the quality between the Education Units, especially in the Satdik in the city and the interior. After the problem is known, then in helping to solve the problem, it must be grouped areas based on their level of education.

3. Data Collection

The data processed in this study consists of the number of individuals with varying education levels from different regions in Central Kalimantan. The data was obtained from the official website of Dukcapil (https://gis.dukcapil.kemendagri.go.id/peta) in June 2022, resulting in a dataset comprising 1565 entries. This dataset includes 11 attributes, namely villages, number of schools, completion of elementary school/SD, completion of junior high school/SMP, completion of high school/SMA, D1 and D2, D3, S1, S2, and S3. It is worth noting that individuals who receive education go through different age ranges: from not yet attending school to completing elementary school at the ages of 5-12, attending junior high school from 13-16, and high school from 16-19. In contrast, the age range for individuals pursuing a college education, from D1 to S3, is not limited or is unknown. The data collection process involved manual extraction from the Dukcapil website, wherein village names in Central Kalimantan were searched one by one. The education level data, ranging from incomplete schooling to college education, was manually entered into Excel sheets based on their respective villages, resulting in a collection of data from 1565 villages.

4. Pre-Processing Data

The data preprocessing stage is crucial before proceeding to the next phase. Its objective is to minimize the failure rate in subsequent processing. During this stage, it has been identified that the collected data contains three missing values in the S3 attribute and two missing values in the S2 attribute. Once the presence of missing data in the dataset is acknowledged, the next step is to re-collect the data to fill in the empty fields in the S2 and S3 columns.

5. Processing of Clustering Data

The data clustering process in this study utilized two algorithms, namely K-Means and K-Medoids. The steps of the K-Means algorithm [18] are as follows:

a. Specify the initial value of cluster K
b. Specify the initial value of K as the center of the cluster, which can be done randomly. The center of the cluster is also assigned a value with a random number.
c. Determine the distance between each object's points and the centroid center by calculating using the formula in equation (1) which is the Euclidean distance, until finding the closest distance to the centroid value for the cluster.

$$d = \left(\sum_{l=1}^{n}(x_i - \mu_i)^2 \right)^{\frac{1}{2}} \qquad (1)$$

d. Randomly determine the new cluster center and then proceed to recalculate the Euclidean Distance [19].
e. Iterate through step b and step c until the centroid value is equal to the average of the cluster item values or until the cluster values remain unchanged.
f. The update of the centroid point value is intended to determine whether there has been a change in the centroid value. If there is no change, the centroid calculation is halted. The flow of the K-Means algorithm can be seen in Figure 2.

The stages for the clustering process of the K-Medoids algorithm are [20] :

a. Determines the K distance value of each pair of objects based on different selections.
b. Calculate the equation Vj for object j using the following formula: (2):

$$V_j = \sum_{i=1}^{n} \frac{d_{ij}}{\sum_{i=1}^{n} d_{ij}} j = 1 \dots . n \qquad (2)$$

Sort Vj from largest to smallest, and then select the object k that has the smallest K value as the initial medoid.
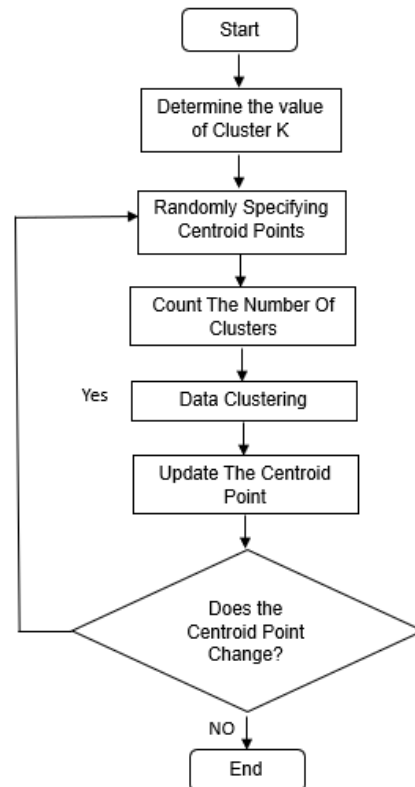

Figure 2. K-Means Algorithm Process

c. Use the preceding result to initialize the clusters by specifying the nearest object as the medoid.
d. Update the medoids of each cluster by minimizing the total distance within the cluster.
e. Set each object to its nearest medoid to obtain the cluster results in the second stage. Then calculate the sum of all the distances between the object and its medoid. The flow of the K-Medoids algorithm is known as shown in Figure 3.

6. Performance or Accuracy Test

After the clustering process using K-Means and K-Medoids, the grouping results are tested for validity using the DBI (Davies Bouldin Index) method. The DBI method criteria are derived from the values of data entanglement within each cluster (intra-cluster) and the distances between the centroid clusters (inter-cluster)[21]. Formula (3) is used to determine the intra-cluster value:

$$intra\ cluster_j = \frac{1}{n} \sum_{pi=j} d(cj, xi)^2 \qquad (3)$$

The distance between clusters j and k is calculated as the distance between centroid clusters to xi. A good cluster has a low intra-cluster value and a maximum inter-cluster value [22]. Next is to calculate the ratio value with the formula (4):

$$R_{j,k} = \frac{intra\ cluster_j + intra\ cluster_k}{|c_j - c_k|} \qquad (4)$$

After the ratio value is found, then the value of the Davies Bouldin Index can be obtained from the following formula (5):

$$DBI = \frac{1}{M} \sum_{l=1}^{M} R_l \qquad (5)$$

From equation (5), k is the number of clusters used. If the DBI value obtained is smaller (non-negative >=0), then the clustering value obtained is also better [23]. As for how to simplify and speed up data calculations, this research uses tools in the Rapidminer software.
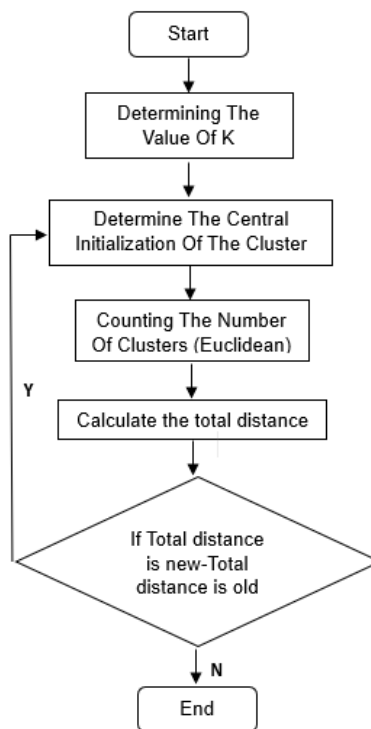

Figure 3. K-Medoids Algorithm Flow Stage

7. Result

After completing all the aforementioned processes, the research will yield results that identify the regions with the lowest and highest levels of education. Furthermore, the study will determine which algorithm demonstrates the highest level of performance.

RESULT AND DISCUSSION

The research follows a pre-designed research method that consists of several stages. The initial step is to determine the topic, which in this case is education in Central Kalimantan. Subsequently, the identified topic is analyzed to identify the existing problems. These problems include suboptimal educational services in certain schools and the unequal distribution of educational assistance among underprivileged students.

The next step involves gathering data from the Ministry of Home Affairs' dukcapil website. Specifically, data on education levels based on villages is obtained, resulting in a dataset consisting of 1,565 entries. Once the dataset is acquired, it undergoes a pre-processing stage, during which missing values are addressed as outlined in the research method. Table I represents the valid data that will be utilized for processing using the K-Means and K-Medoids algorithms, after having undergone the pre-processing stage.

Table 1. Result of Data Pre-Processing

No	Village	Not yet in school	Not finished SD	Finished SD	SMP	SMA	D1, D2	D3	S1	S2	S3
1	Basarang	824	577	890	416	292	13	5	39	0	0
2	Basarang Jaya	467	230	296	250	376	32	9	102	0	0
3	Basungkai	325	199	384	183	111	2	1	18	0	0
4	Batu Nindan	442	257	341	240	309	12	14	46	0	0
5	Batuah	423	211	527	173	126	1	1	15	0	0
6	Bungai Jaya	487	237	442	377	378	19	8	55	0	0
7	Lunuk Ramba	232	142	262	148	210	14	12	52	0	1
8	Maluen	759	473	721	346	326	9	5	80	2	0
...	....	....	......	....	....	.....	......	....	.....	...	...
....	....	....	......	....	....	.....	......	....	.....	...	...
1565	Sabaru	1,757	724	648	768	1,661	67	105	433	25	2

After the preprocessing or data selection stage, the next step is to cluster the data using the K-Means and K-Medoids algorithms. Before data processing, the format of the dataset column is changed first in the data type of the Village attribute where the role is a label. Meanwhile, the level of education from school to college is an integer.

The first dataset processing for clustering involves using the K-Means algorithm. The designed model of the K-Means algorithm in the Rapidminer tool is illustrated in Figure 4. In the figure, the initial cluster values were determined with various K values, including 3 clusters, 4 clusters, 5 clusters, 6 clusters and 7 clusters in 5 experiments.
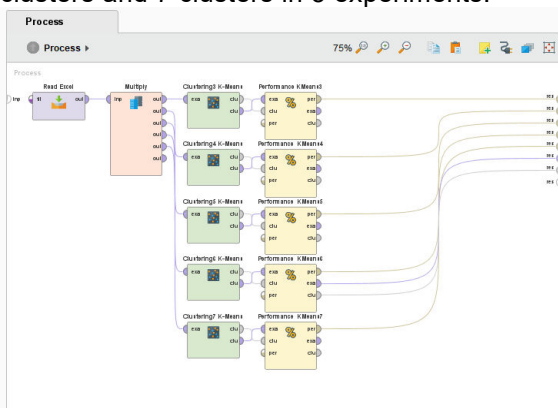

Figure 4. Modeling clustering K Means

Figure 5 is the result of the K-Means model cluster known that K-6 is the cluster with the best DBI values. Figure 5 shows 6 model clusters in villages with the number of educations grouped by similarity in number. Each group consists of cluster 0 as many as 1484 villages, cluster 1 there is 1 village, cluster 2 = 21 villages, cluster 3 = 3 villages, cluster 4 = 56 villages, and cluster 5 = 1 village.
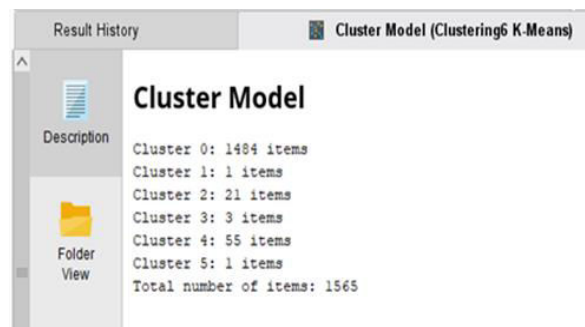

Figure 5. Cluster Model K-Means k=6

The results of the clustering rating visualization are presented in Figure 6. It is observed that the red color holds the top rank within cluster-3. The purple color follows as the second rank in cluster-2. The black color secures the third rank within cluster-5. The orange color

occupies the fourth rank in cluster-4, while the green color represents the fifth rank in cluster-1. Lastly, the blue color represents the last rank in cluster-0. Thus, based on the visualization in Figure 6, it can be explained, as depicted in Figure 5, that cluster-3 holds the first rank with 3 villages. Cluster-2 secures the second rank with 21 villages, while cluster-5 occupies the third rank with 1 village. Cluster-4 is in the fourth position with 55 villages, and cluster-1 holds the fifth rank with 1 village. Finally, cluster-0 encompasses the largest number of villages, with a total of 1484 villages.
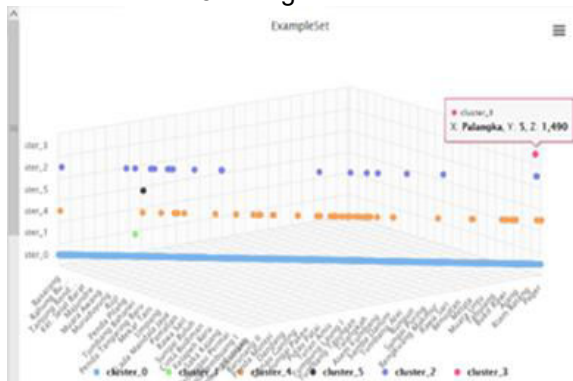

Figure 6. Result K-Means data cluster ranking k=6

Table 2. Results of Performance K-Means

Cluster K-Means	Davies Bouldin Index
3	-0.636
4	-0.684
5	-0.711
6	-0.439
7	-0.571

The clustering modeling and ranking of each K-Means cluster have been revealed in Figure 5 and Figure 6. The performance of the clusters is described in Table 2, which presents the results of performance testing aimed at determining the accuracy of the clustering process. The cluster results were evaluated for accuracy using the Davies Bouldin Index (DBI). By conducting DBI performance testing for different values of k (k=3, k=4, k=5, k=6, and k=7), it is evident that the optimal accuracy is achieved with a group of 6 clusters. Notably, the K-Means algorithm achieves the best performance with a cluster value of -0.439. This finding indicates that, among all the performance test results for the K-Means algorithm, the result with k=6 closely aligns with the expected value (non-negative >=0) according to the DBI theory.

The next stage is to cluster the datasets using the K-Medoids Algorithm. Dataset modeling using the K-Medoids algorithm is shown in Figure 7. K-Medoids design modeling is done in the RapidMiner the same as K-Means modeling, the difference lies only in the use of algorithm tools. The determination of the value of K also starts from K=3 to K =7, with 5 experiments on the RapidMiner tool.
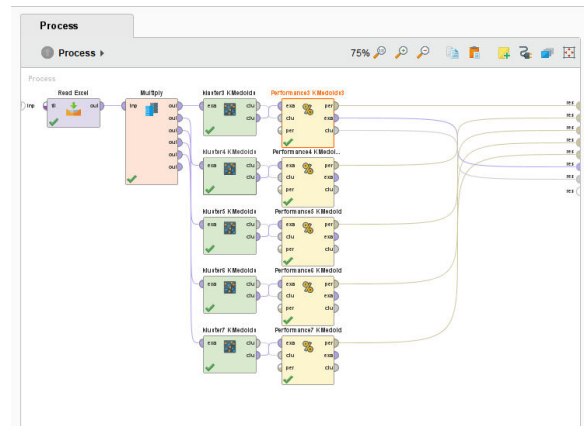

Figure 7. Modeling clustering K-Medoids

After completing the modeling stage of the K-Medoids algorithm, a K=3 cluster model was obtained, which demonstrated the highest accuracy based on the performance test. The cluster results of the K=3 K-Medoids model are illustrated in Figure 8. It is worth mentioning that cluster-0 comprises 22 villages, cluster-1 includes 79 villages, and cluster-2 encompasses 1464 villages.
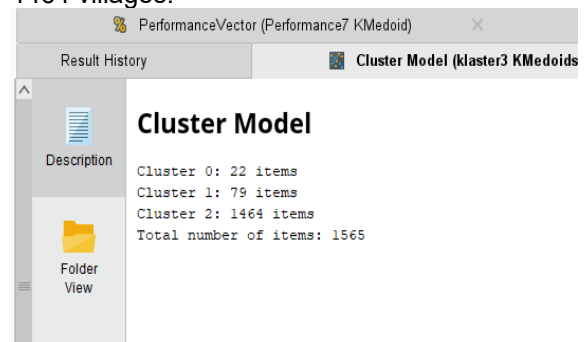

Figure 8. Cluster Model K-Medoids k=3

Figure 9 displays the visualization of the ranking results from the K-Medoids algorithm clustering. It is evident that the orange color occupies the first rank, followed by the green color in the second rank, and the blue color in the third rank. Therefore, based on the visualizations in Figure 8 and Figure 9, it can be concluded that the orange color represents Cluster-2, indicating that Cluster-2 holds the highest rank with a total of 1464 villages. The green color represents Cluster-0, securing the

second rank with 22 villages. Lastly, the blue color represents Cluster-1, positioned in the third rank with 79 villages.
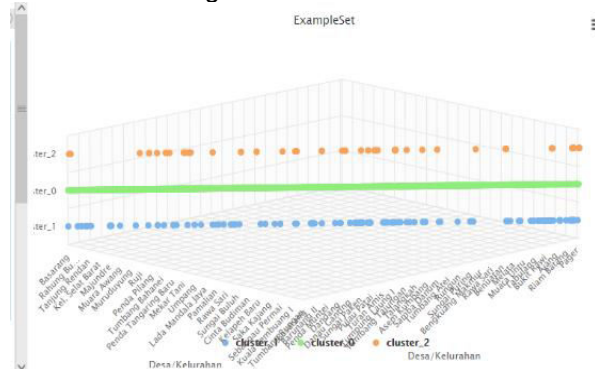

Figure 9. Result K-Medoids data cluster of ranking k=6

Next, we proceed to the accuracy assessment stage of the K-Medoids algorithm for each pre-determined cluster. Table 3 presents the clustering results for K=3, K=4, K=5, K=6, and K=7, with varying accuracy values arranged in ascending order. The grouping of clusters using the K-Medoids algorithm has provided insights into their

respective performance values. Among the five experiments conducted, the DBI accuracy value reveals that the optimal performance is achieved in the village clustering with K=3, attaining an accuracy value of -0.866.

The subsequent stage involves comparing the performance of each cluster between the K-Means and K-Medoids algorithms. This comparison is conducted to determine the algorithm that yields the best performance. Additionally, the comparison aims to assess the accuracy of the clustering process implemented by both the K-Means and K-Medoids algorithms.

Table 3. Results of Performance K-Medoids

Cluster K-Medoids	Davies Bouldin Index
3	-0.866
4	-1.274
5	-1.963
6	-2.301
7	-2.229

Table 4. **Davies Bouldin Index value comparison**

Cluster	K-Means	K-Medoids
3	-0.636	-0.866
4	-0.684	-1.274
5	-0.711	-1.936
6	-0.439	-2.301
7	-0.571	-2.229
Smallest DBI Value	-0.439	-0.866

The test results for each cluster using both the K-Means and K-Medoids algorithms are displayed in Table 4, which presents the outcomes of the Davies Bouldin Index test for each cluster. This table was created to compare the performance values of the K-Means and K-Medoids algorithms. Notably, the K=6 cluster in K-Means exhibits the smallest DBI value of -0.439, whereas the smallest DBI value in K-Medoids is observed in cluster K=3 with a value of -0.866. The smaller DBI value in K-Means (k=6) indicates that clustering using the K-Means algorithm yields superior and optimal results compared to the K-Medoids method in clustering education levels by the village. Additionally, after employing the Rapidminer tool to perform clustering with both algorithms, it was

observed that the K-Means algorithm processes the dataset faster than the K-Medoids algorithm.

CONCLUSION

After clustering the education level data of 1,565 villages using the K-Means and K-Medoids algorithms, the performance of the clusters was evaluated using the Davies Bouldin Index. The results showed that the K-Means algorithm formed 6 clusters with a DBI value of -0.439, while the K-Medoids algorithm formed 3 clusters with a DBI value of -0.866. Consequently, the K-Means algorithm demonstrated more optimal results in clustering the education levels in Central Kalimantan province based on the DBI performance. This is evident from the lower DBI value achieved by

the K-Means algorithm compared to the DBI value of the K-Medoids algorithm. A lower DBI value indicates better quality in clustering. Therefore, the K-Means algorithm was selected as the preferred method for identifying villages with the lowest and highest education levels.

The K-Means algorithm resulted in the formation of 6 clusters. Cluster_0 consisted of 1484 villages, cluster_1 had 1 village, cluster_2 contained 21 villages, cluster_3 encompassed 3 villages, cluster_4 comprised 56 villages, and cluster_5 included 1 village. By visualizing the data using the K-Means algorithm, it was observed that there were villages with both low and high education levels. Villages with low education levels were primarily located in Cluster_0, including Basarang, Rahung Bungai, Tanjung Rendan, Majundre, Muruduyung, Ruji, Penda Pilang, Tumbang Bahanei, Umpang, Pamalian, Saka Kajang, Melata, Muara Untu, Dandang, Continenttan, and numerous others, totaling 1484 villages. On the other hand, villages with high education levels were found in cluster_3, comprising 3 villages, namely Bukit Tunggal, Menteng, and Palangka Raya. This research serves as a valuable recommendation for the government and the Ministry of Education and Culture regarding the distribution of educational assistance, including support for students and the provision of school facilities in villages with the lowest education levels. The aim is to address the educational disparities within these communities and reduce the educational backwardness that they face.

REFERENCES

[1] N. Nurahman, A. Purwanto, and S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, 2022, doi: 10.30812/matrik.v21i2.1411.

[2] A. S. M. Noor, M. Younas, and M. Arshad, "A review on cloud based knowledge management in higher education institutions," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 6, 2019, doi: 10.11591/ijece.v9i6.pp5420-5427.

[3] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.

[4] A. Akram *et al.*, "Comparative review on information and communication technology issues in education sector of developed and developing countries: A case study about pakistan," *Bull. Electr. Eng. Informatics*, vol. 10, no. 6, 2021, doi: 10.11591/eei.v10i6.3245.

[5] E. Horne, H. Tibble, A. Sheikh, and A. Tsanas, "Challenges of clustering multimodal clinical data: Review of applications in asthma subtyping," *JMIR Medical Informatics*, vol. 8, no. 5. 2020. doi: 10.2196/16452.

[6] A. Hermawati, S. Jumini, M. Astuti, F. Ismail, and R. Rahim, "Unsupervised Data Mining with K-Medoids Method in Mapping Areas of Student and Teacher Ratio in Indonesia," *TEM J.*, vol. 9, no. 4, 2020, doi: 10.18421/TEM94-37.

[7] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo, "PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS DALAM KLASTERISASI PRODUK ASURANSI PERUSAHAAN NASIONAL," *J. Ilm. FIFO*, vol. 11, no. 2, 2019, doi: 10.22441/fifo.2019.v11i2.010.

[8] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "PERBANDINGAN ALGORITMA K-MEANS DENGAN K-MEDOIDS PADA PENGELOMPOKAN ARMADA KENDARAAN TRUK BERDASARKAN PRODUKTIVITAS," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 2, 2021, doi: 10.29100/jipi.v6i2.2008.

[9] R. Ananda, A. R. Dewi, and N. Nurlaili, "A COMPARISON OF CLUSTERING BY IMPUTATION AND SPECIAL CLUSTERING ALGORITHMS ON THE REAL INCOMPLETE DATA," *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, 2020, doi: 10.21609/jiki.v13i2.818.

[10] T. Hardiani, "ANALISIS CLUSTERING KASUS COVID 19 DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI | 157," vol. 11, pp. 156–165, 2022.

[11] K. Khomsatun, D. Ikhsan, M. Ali, and K. Kursini, "SISTEM PENGAMBILAN KEPUTUSAN PEMILIHAN LAHAN TANAM DI KABUPATEN WONOSOBO DENGAN K-MEANS CLUSTERING DAN TOPSIS," *J. Nas. Pendidik. Tek. Inform.*, vol. 9, no. 1, 2020, doi: 10.23887/janapati.v9i1.23073.

[12] M. Minarni, E. I. Sari, A. Syahrani, and P. Mandarani, "Klasterisasi Penyakit Menggunakan Algoritma K-Medoids pada Dinas Kesehatan Kabupaten Agam Provinsi Sumatra Barat Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI," *J. Nas. Pendidik. Tek. Inform.*,

vol. 10, no. 3, 2021, doi: 10.23887/janapati.v10i3.34904.

[13] S. Ramadhani, D. Azzahra, and T. Z, "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, doi: 10.31849/digitalzone.v13i1.9292.

[14] N. Dwi Aulia, Diana; Nurhaman, "Klasterisasi Pendidikan Masyarakat untuk mengetahui Daerah dengan Pendidikan Terendah menggunakan Algoritma K-Means," vol. 5, no. 1, pp. 38–44, 2023.

[15] R. Kurniawan, M. M. M. Mukarrobin, and M. Mahradianur, "Klasterisasi Tingkat Pendidikan Di Dki Jakarta Pada Tingkat Kecamatan Menggunakan Algoritma K-Means," *Technol. J. Ilm.*, vol. 12, no. 4, p. 234, 2021, doi: 10.31602/tji.v12i4.5633.

[16] W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia," *Ilk. J. Ilm.*, vol. 13, no. 1, 2021, doi: 10.33096/ilkom.v13i1.763.31-35.

[17] R. A. Malik, S. Defit, and Y. Yuhandri, "Comparison of K-Means Clustering Algorithm with Fuzzy C-Means In Measuring Satisfaction Level Of Television Da'wah Surau TV," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 3, no. 1, 2018, doi: 10.36341/rabit.v3i1.387.

[18] K. Ariasa, I. G. A. Gunadi, and I. M. Candiasa, "Optimasi Algoritma Klaster Dinamis pada K-Means dalam Pengelompokkan Kinerja Akademik Mahasiswa (Studi Kasus: Universitas Pendidikan Ganesha)," *J. Nas. Pendidik. Tek. Inform. JANAPATI*, vol. 9, no. 2, 2020.

[19] N. Qona'ah, A. R. Devi, and I. M. G. M. Dana, "Laboratory Clustering using K-Means, K-Medoids, and Model-Based Clustering," *Indones. J. Appl. Stat.*, vol. 3, no. 1, 2020, doi: 10.13057/ijas.v3i1.40823.

[20] C. Oktarina, K. A. Notodiputro, and I. Indahwati, "COMPARISON OF K-MEANS CLUSTERING METHOD AND K-MEDOIDS ON TWITTER DATA," *Indones. J. Stat. Its Appl.*, vol. 4, no. 1, 2020, doi: 10.29244/ijsa.v4i1.599.

[21] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, 2021.

[22] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, 2021, doi: 10.18421/TEM103-13.

[23] N. R. Shamsuddin and N. I. Mahat, "Comparison Between k-Means and k-Medoids for Mixed Variables Clustering," in *Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017)*, 2019. doi: 10.1007/978-981-13-7279-7_37.