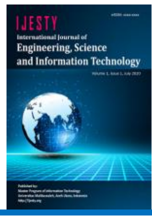
International Journal of Engineering, Science and InformationTechnology
Volume 5, No. 1 (2025) pp. 167-176
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i1.673
Research Paper, Short Communication, Review, Technical Paper



Comparative Analysis of K-Means and K-Medoids to Determine Study Programs

Salamah*, Dahlan Abdullah, Nurdin

Department of Information Technology, Faculty of Engineering, Malikussaleh University, Aceh, Indonesia

**Corresponding author Email: salamah.227110201003@mhs.unimal.ac.id*

The manuscript was received on June 25 2024, revised on August 28 2024, and accepted on December 15 2024, date of publication January 9 2025

Abstract

Education is the main foundation for the advancement of civilization. A high level of education in society is directly proportional to that civilization's progress. Higher education is vital in shaping quality human resources and contributing to community and national development. In today's era of information and technology, data processing and analysis are key to understanding the development of study programs in higher education institutions. Clustering techniques identify patterns and relationships in large and complex datasets, which are crucial in determining study programs at educational institutions. This research compares two popular clustering methods, K-Means and K-Medoids, to assess study programs. The data consists of odd semester grades of 87 students in the third year of high school with five variables. The cluster information is based on the minimum academic criteria of 18 study programs representing 7 faculties in Malikussaleh University and is grouped into 5 clusters. The evaluation of clusters is conducted using the Davies-Bouldin Index (DBI). The result of the study indicates that the K-Means algorithm has 5 clusters with cluster members of 31, 5, 13, 26 and 17, and a DBI value of 1,19010. Meanwhile, the K-Medoids algorithm has 5 clusters with cluster members of 33, 15, 17, 17 and 5, and a DBI value of 1,27833. Based on the DBI value, the K-Means algorithm demonstrates better cluster quality than the K-Medoids algorithm.

Keywords: *Education, Clustering, K-Means, K-Medoids, Davies-Bouldin Index (DBI).*

1. Introduction

Education is the foundation of a civilization's progress. The higher the level of education in a society, the more advanced the civilization that is formed. Conversely, civilization can regress and become backward if a society neglects education. Education is passing on knowledge, skills, and habits from one generation to the next [1]. Higher education plays a vital role in shaping quality human resources, which contribute to the development of society and the nation [2]. In the current era of information and technology, data processing and analysis are key to understanding the development of study programs in higher education institutions. One of the commonly used techniques for data analysis is clustering, which effectively identifies patterns and relationships in large and complex datasets. In the context of determining study programs in educational institutions, selecting the appropriate clustering algorithm is crucial to optimizing resource allocation and maximizing students learning experiences [3] [4] [5] [6]. With the availability of various clustering methods, this research focuses on comparing two popular clustering methods, namely K-Means and K-Medoids, to determine study programs. Research conducted by Nurdin et al. designed an information system model to map and cluster fisheries products from 10 fishing ports on Aceh's northern coast using the K-Means algorithm and web-based GIS. This study resulted in two groups of fisheries products: Group 1, superior fish, and Group 2, common fish [7]. The K-Means clustering algorithm can also be used to cluster provinces in Indonesia based on COVID-19 pandemic risk using COVID-19 data from the Task Force Team as of April 19, 2020 [8] [9] [10].

2. Literature Review

2.1. Data Mining

Data mining is a field of study that explores techniques for extracting knowledge or discovering hidden patterns in data. Therefore, it is often also called Knowledge Discovery in Database (KDD) [11]. Data mining involves the analysis of data to identify significant relationships and infer previously unknown information using contemporary methods that benefit data owners. This technology extracts predictive information hidden in databases, presenting a considerable potential for companies to utilize data warehouses. In general, data mining is divided into two main categories [12]:

1. Descriptive
   Data mining techniques that reveal essential characteristics of the data in a database, such as clustering, association, and sequential mining.
2. Predictive

Data mining techniques that discover patterns from data to predict future outcomes using methods such as classification. Data mining is described as extracting knowledge from large datasets, also known as Knowledge Discovery in Database (KDD).

2.2. Clustering

Clustering is one of the well-known techniques in the field of data mining. In the context of data mining, clustering refers to grouping a number of data points or objects into clusters. Clustering aims to group data so that each cluster contains highly similar data and is distinct from objects in other clusters [13]. Clustering is also often referred to as segmentation. This method aims to identify groups in a case based on the similarity of attributes among these groups. Clustering works by separating a data group based on individual characteristics, where the object can be people, events, or other elements. These objects are grouped so that several interconnected levels are formed between clusters. The strengths and weaknesses among members of different clusters can be seen through the level of similarity or difference within the members included in one cluster [14].

2.3. Study Programs

Selecting a study program for prospective students is a step in focusing them on a specific concentration area. This process aims for each individual to delve further into lessons that align with the established concentration. Additionally, the determination of a study program aims to guide each individual to develop their skills and interests. Thus, it is hoped that prospective students can optimize their potential in the chosen study program concentration [15]. This selection process aims to ensure that students select programs that match their potential and desires, allowing them to enjoy the learning process more effectively [16].

2.4. K-Means Algorithm

The K-Means Algorithm is a non-hierarchical data clustering technique that separates data into clusters. This process involves grouping data with similar features into one group, while data with different characteristics are grouped into different groups. This technique allows for data clustering based on similarities in certain features or attributes, thus facilitating the analysis and understanding of complex data structures [17]. The K-Means algorithm is a distance-based clustering method that divides data into several clusters and operates only on numerical attributes. K-Means is a type of partitioning clustering that separates data into distinct sub-regions. It is known for its simplicity and ability to cluster large datasets and handle outliers [18] quickly. K-Means is used for initializing parameters due to its simplicity and performance with large datasets compared to hierarchical clustering. Performance evaluation is done by analyzing various image categories as a case study [19]. The steps of the K-Means method used to determine the number of clusters and data placement within clusters include [20]:

1. Determine the number of clusters (k).
2. Initialize *centroid*
   Initially, randomly select k data points as the cluster centres.
3. Calculate Distance.
   Compute the Distance between data points and cluster centres using Euclidian Distance. The Euclidian Distance formula is as follows:
   D(X,Y) =
   $$\sqrt{(x1 - yi)^2 + (xi - yi)^2}$$
4. Cluster data.
   Assign data points to the nearest cluster centre. New clusters are formed if all data points have been assigned to the nearest cluster.
5. Determine new centroids (k)
   C(i) =
   $$\frac{x_1 + x_2 + x_3 + \cdots + x_n}{\sum x}$$
   C is the new Centroid calculated based on the data in the formed clusters.
6. Iterate
   Repeat determining cluster centres until no further changes occur in the centroids.

These steps form the basis of the K-Means Clustering process for grouping data into appropriate clusters.

2.5. K-Medoids Algorithm

The K-Medoids algorithm is a partitioning clustering method that aims to find k-clusters among data that best represent the objects in a dataset [21]. The K-Medoids algorithm is used to overcome the weaknesses of the K-Means algorithm, which is highly sensitive to outliers. Outliers have characteristics or locations that are very distant from most other data. The presence of outlier objects can distort the mean value of a cluster if it is included. Using K-Medoids is expected to be more resistant to the influence of outliers because it uses data objects as representatives (medoids) as the cluster centre, thus being more stable against extreme values [22]. Another advantage is that even if the order of objects in the dataset is changed, the clustering results remain consistent and unchanged. This is because the clustering process results in the K-Medoids algorithm do not depend on the order in which the dataset is entered. This advantage provides stability to the algorithm, providing consistent results regardless of how the data is arranged or entered into the algorithm [23]. The following are the steps in determining the K-Medoids method [24]:

1. Data Preparation: Prepare data samples for cluster centres in the required clusters.
2. Initialization: Determine random medoid values for each cluster.
3. Distance Calculation: Calculate the Euclidean Distance between each data object and the cluster medoid using the equation:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (X_i - Y_j)^2}$$

Where:
d (x,y) is the Distance between object i and object j,
X_i is the first attribute value of object *i*,
Y_j is the first attribute value of the object *j*,
n is the number of attributes used.

4. Medoid Selection: Randomly select data objects in each cluster that will become the new medoid to produce the least total cost in distance calculation.
5. New Medoid Determination: Calculate the Number of differences to determine the new distance value and subtract the number of old distance values. If the difference is negative, it will form a new object in the medoid, and steps 2-4 will be repeated until no changes in the medoid are produced to get clusters and the members in each cluster.

2.6. Cluster Evaluation

The Davies-Bouldin Index (DBI) method, introduced by David L. Davies and Donald W. Bouldin, evaluates clusters. DBI assesses cluster quality based on the compactness and separation between clusters. This index measures the validity of clustering results by considering cohesion, which describes the overall relation of data to the cluster centre and separation, which refers to the Distance between cluster centres. The purpose of using DBI is to maximize the Distance between clusters and minimize the Distance between points within a cluster. When the inter-cluster Distance is maximized, the differences between clusters become more apparent.

In contrast, minimal intra-cluster Distance indicates a high degree of similarity among the characteristics within that cluster [25]. The DBI calculation begins with finding the Sum of Squares Within the Cluster (SSW), which indicates the cohesion matrix within the *I* cluster. The SSW value is obtained using the following equation [26]:

$$SSW = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)$$

Where:

SSW = Sum of Square Within Cluster
m_i = number of data points in the i cluster
c_i = Centroid of the i cluster
$d(x_j, c_i)$ = Distance from the j data point to the *i* cluster centroid.

The next step in the DBI calculation is to use the above equation to determine the separation between clusters. This equation is obtained by calculating the Sum of Squares Between Clusters (SSB), which indicates the Distance between clusters. The SSB value is obtained using the following equation:

$$SSB_{ij} = d_{ij}$$

Where:
SSB = Sum of Square Between Cluster
d_{ij} = Euclidian Distance between data points i and j

The following step is to find the ratio of SSW and SSB calculation as in the following equation:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

The final step in the DBI calculation is to use the equation below to obtain the DBI value. This equation is obtained by calculating the DBI as the ratio of SSW and SSB. The equation is as follows:

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

Where k is the number of clusters being calculated.

3. Research Methods

The following steps were undertaken in the research:

1. Identification of Research Objectives: The research objectives are to conduct a comparative analysis of the K-Means and K-Medoids algorithms to determine the study program.
2. Literature Study: Conducting a literature study to understand the basic concepts of the K-Means and K-Medoids algorithms, theories related to the research, and previous studies that have been conducted using the K-Means and K-Medoids methods as references that will be used in this study. After conducting a literature study, it is continued to the next stage.
3. Collecting data: At this stage, the data collection process is carried out; the data collected is in the form of grade data from 3rd-grade students of SMA Sukma Bangsa Lhokseumawe totalling 87 student data and study program data at Malikussaleh University in the form of minimum academic limit values of 18 study programs representing seven existing faculties.
4. Data analysis: Before the data is analyzed using the K-Means algorithm with K-Medoids, the student value data that has been obtained is grouped into five variables consisting of Language Value, Logic Value, Science Value, Practical Value and Social Value, which will then be used in the analysis process. While cluster data is determined from the minimum academic limit value data from the study program consisting of 18 study programs representing each faculty. Then, the data is analyzed using the K-Means and K-Medoids algorithms to obtain results for each.
5. Cluster result evaluation: After the analysis is carried out using the algorithms and the results are obtained from each cluster, the analysis results are evaluated using the Davies Bouldin Index (DBI) formula to determine the best algorithm to use to determine the study program. From the results of the cluster evaluation, the best algorithm can be determined based on the DBI value obtained.
6. Analysis and Results: Conducting analysis for cluster results and results of Davies Bouldin Index (DBI) values obtained and formulating conclusions based on the data analysis results that have been carried out to determine the more effective algorithm of the two algorithms used.

Based on the above description, to facilitate understanding of the research flow, the flowchart of the research can be seen in the image below:

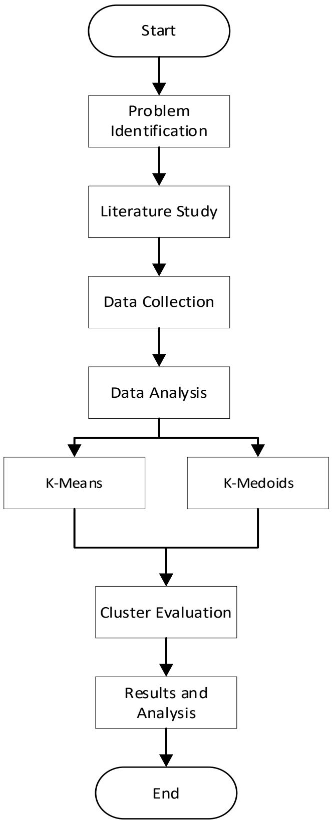


Fig 1. Research Stages

4. Result and Discussions

4.1. Dataset

The dataset used is the odd semester academic grade of students from the third year at Senior High School of Sukma Bangsa Lhokseumawe. The dataset includes 87 data points from the 3rd-grade students Sukma Bangsa Lhokseumawe High School. The number of data points used is 87, representing students from the third year at Sukma Bangsa Lhokseumawe High School, with five variables: Language Score (K1), Logic Score (K2), Science Score (K3), Practical Score (K4), and Social Score (K5). The complete data can be seen in the table below:

Table 1. *Dataset*

No	Name	(K_1)	(K_2)	(K_3)	(K_4)	(K_5)
1	A R	93.35	95.03	95.66	94.88	94.19
2	A RS	96.95	98.16	97.43	94.60	95.39
3	A N P	97.39	94.24	93.71	97.12	94.89
4	A S	95.61	93.04	93.01	93.98	92.29
…..	…..	…..	…..	…..	…..	…..
84	S K	97.65	96.50	94.33	95.65	97.89
85	S A T	95.08	92.00	92.59	95.55	95.70
86	T M A S	96.61	96.34	95.50	95.76	98.11
87	U A	95.37	94.95	93.97	95.23	97.77

4.2. Manual K-Means Calculation

1. Determine the number of clusters (k): In this research, the number of clusters (k) is set to 5vb clusters.
2. Determine the cluster centres (initial centroids): Four centroids are randomly selected, corresponding to the number of clusters determined. The initial centroids can be seen in the table below:

Table 2. Initial Centroid

Data	Centroid 1	K_1	K_2	K_3	K_4	K_5
Data 3	C1	97.39	94.24	93.71	97.12	94.89
Data 8	C2	97.26	98.40	97.88	96.63	96.86
Data 56	C3	96.83	96.30	96.97	95.80	95.29
Data 58	C4	94.87	94.61	93.79	97.30	94.45
Data 85	C5	95.08	92.00	92.59	95.55	95.70

3. Calculate the Distance for each variable to the determined cluster centres: This is done using the Euclidean Distance formula for each data point to the cluster centre.
The distance calculations for the first data point to each Centroid are as follows:

$$D(d_1, c_1) = \sqrt{(93.35-97.39)^2 + (\dots\dots\dots\dots\dots.)^2 + (94.19-94.89)^2}$$
$$= \sqrt{16{,}34181 + 0{,}62410 + 3{,}80250 + 5{,}01760 + 0{,}49000}$$
$$= 5{,}126101$$

$$D(d_1, c_2) = \sqrt{(93.35-97.26)^2 + (\dots\dots\dots\dots\dots.)^2 + (94.19-96.86)^2}$$
$$= \sqrt{15{,}32723 + 11{,}35690 + 4{,}92840 + 3{,}06259 + 7{,}12890}$$
$$= 6{,}46560$$

$$D(d_1, c_3) = \sqrt{(93.35-96.83)^2 + (\dots\dots\dots\dots\dots.)^2 + (94.19-95.29)^2}$$
$$= \sqrt{12{,}14523 + 1{,}61290 + 1{,}71610 + 0{,}84640 + 1{,}21000}$$
$$= 4{,}18696$$

$$D(d_1, c_4) = \sqrt{(93.35-94.87)^2 + (\dots\dots\dots\dots\dots.)^2 + (94.19-94.45)^2}$$
$$= \sqrt{2{,}32563 + 0{,}17640 + 3{,}49690 + 5{,}85640 + 0{,}06760}$$
$$= 3{,}45296$$

$$D(d_1, c_5) = \sqrt{(93.35-95.08)^2 + (\dots\dots\dots\dots\dots.)^2 + (94.19-95.70)^2}$$
$$= \sqrt{3{,}01023 + 9{,}18090 + 9{,}42490 + 0{,}44890 + 2{,}28010}$$
$$= 4{,}93407$$

After the distance calculations are completed, the next step is to determine the nearest Distance and assign the data points to their respective clusters based on the previously determined cluster centres. The results of the overall distance calculations can be seen in the table below:

Table 3. Results of Iteration 1

No	C1	C2	C3	C4	C5	Distance	Results
1	5.12601	6.46560	4.18696	3.45296	4.93407	3.45296	C4
2	5.99954	2.57613	2.26632	6.19375	8.11641	2.26632	C3
3	0.00000	6.23151	4.13333	2.58981	3.83635	0.00000	C1
4	4.66026	9.11508	6.33373	4.39377	3.95339	3.95339	C5
5	3.65581	6.39425	3.69684	2.54108	4.21922	2.54108	C4
……	……	……	……	……	……	……	……
……	……	……	……	……	……	……	……
……	……	……	……	……	……	……	……
83	4.27159	4.84234	3.69569	4.61634	5.10621	3.69569	C3
84	4.08923	4.28788	3.80322	5.11353	5.88970	3.80322	C3
85	3.83635	8.72975	6.40059	3.59461	0.00000	0.00000	C5
86	4.52069	3.55667	3.18824	4.97110	5.95792	3.18824	C3
87	4.06307	5.79231	4.40788	3.96299	3.88308	3.88308	C5

In the third iteration, the process stopped because the results of the centroid calculations in the third iteration showed no change in cluster centres. Therefore, the clustering results from the third iteration are as follows. The results of the clusters can be seen in the table below:

Table 4. K-Means Clustering

No	C1	C2	C3	C4	C5	Distances	Results
1	3.93219	6.48017	3.84318	2.44921	5.73346	2.44921	C4
2	4.29740	3.15081	1.75282	5.50588	10.03384	1.75282	C3
3	2.41908	6.08882	4.74042	3.01944	6.62114	2.41908	C1
4	5.28513	9.19119	6.44737	3.02185	4.43010	3.02185	C4
5	3.21963	6.39220	3.78412	1.22644	5.56764	1.22644	C4
……	……	……	……	……		……	……
……	……	……	……	……		……	……
……	……	……	……	……		……	……
83	2.22694	5.07258	3.13961	3.93910	7.34998	2.22694	C1
84	2.52313	4.44428	3.66190	4.94359	8.54238	2.52313	C1
85	4.20993	8.70438	6.61121	3.25863	3.05341	3.05341	C5
86	2.38069	3.74259	3.03982	4.81872	8.40356	2.38069	C1
87	2.25540	5.86628	4.12470	3.60698	6.15741	2.25540	C1

The clustering results from the table above can be seen in Table 5 below:

Table 5. K-Means Clustering Results

No	Cluster	Amount of Data	DBI Value
1	C1	31	1.19010
2	C2	5	
3	C3	13	
4	C4	26	
5	C5	12	

The results of clustering using the K-Means algorithm consist of 5 clusters, with members in each cluster being Cluster 1 having 31 members, cluster 2 having five members, cluster 3 having 13 members, cluster 4 having 27 members, and Cluster 5 having 12 members with a resulting DBI value of 1.19010. The graph of clustering results with the K-Means algorithm can be seen in the figure 2 below:

Fig 2. K-Means Clustering Results Graph

4.3. Manual Calculation of K-Medoids

1. Determine the number of clusters (k): In this research, the number of clusters (k) is set to 5 clusters.
2. Determine the cluster centres (initial medoids): Four medoids are randomly selected, which can be seen in the table below:

Table 6. Initial Medoids

Data	Medoid 1	K_1	K_2	K_3	K_4	K_5
Data 3	C1	97.39	94.24	93.71	97.12	94.89
Data 8	C2	97.26	98.40	97.88	96.63	96.86
Data 56	C3	96.83	96.30	96.97	95.80	95.29
Data 58	C4	94.87	94.61	93.79	97.30	94.45
Data 85	C5	95.08	92.00	92.59	95.55	95.70

3. Calculate the minimum Distance using the Euclidean Distance formula for each data point to the cluster centers. The distance calculations for the first data point to each medoid are as follows:

$$D(d_1, c_1) = \sqrt{(93.35 - 97.39)^2 + (\dots\dots\dots\dots\dots)^2 + (94.19 - 94.89)^2}$$
$$= \sqrt{16{,}34181 + 0{,}62410 + 3{,}80250 + 5{,}01760 + 0{,}49000}$$
$$= 5{,}126101$$

$$D(d_1, c_2) = \sqrt{(93.35 - 97.26)^2 + (\dots\dots\dots\dots\dots)^2 + (94.19 - 96.86)^2}$$
$$= \sqrt{15{,}32723 + 11{,}35690 + 4{,}92840 + 3{,}06259 + 7{,}12890}$$
$$= 6{,}46560$$

$$D(d_1, c_3) = \sqrt{(93.35 - 96.83)^2 + (\dots\dots\dots\dots\dots)^2 + (94.19 - 95.29)^2}$$
$$= \sqrt{12{,}14523 + 1{,}61290 + 1{,}71610 + 0{,}84640 + 1{,}21000}$$
$$= 4{,}18696$$

$$D(d_1, c_4) = \sqrt{(93.35 - 94.87)^2 + (\dots\dots\dots\dots\dots)^2 + (94.19 - 94.45)^2}$$
$$= \sqrt{2{,}32563 + 0{,}17640 + 3{,}49690 + 5{,}85640 + 0{,}06760}$$
$$= 3{,}45296$$

$$D(d_1, c_5) = \sqrt{(93.35 - 95.08)^2 + (\dots\dots\dots\dots\dots)^2 + (94.19 - 95.70)^2}$$
$$= \sqrt{3{,}01023 + 9{,}18090 + 9{,}42490 + 0{,}44890 + 2{,}28010}$$
$$= 4{,}93407$$

The results of the overall calculations can be seen in the table below:

Table 7. Results of Iteration 1

No	C1	C2	C3	C4	C5	Distance	Results
1	5.12601	6.46560	4.18696	3.45296	4.93407	3.45296	C4
2	5.99954	2.57613	2.26632	6.19375	8.11641	2.26632	C3
3	0.00000	6.23151	4.13333	2.58981	3.83635	0.00000	C1
4	4.66026	9.11508	6.33373	4.39377	3.95339	3.95339	C5
……	……	……	……	……	……	……	……
84	4.08923	4.28788	3.80322	5.11353	5.88970	3.80322	C3
85	3.83635	8.72975	6.40059	3.59461	0.00000	0.00000	C5
86	4.52069	3.55667	3.18824	4.97110	5.95792	3.18824	C3
87	4.06307	5.79231	4.40788	3.96299	3.88308	3.88308	C5

The total Distance from the results of the first iteration is 223,00955

The process stopped in the third iteration because the total deviation value S in the third iteration was calculated to be 0. Therefore, the clustering results from the previous iteration and the second iteration are obtained. The clustering results can be seen in the table below:

Table 8. K-Medoids Clustering

No	C1	C2	C3	C4	C5	Distances	Results
1	2.38722	3.94161	4.82892	4.85502	6.35613	2.38722	C1
2	4.55715	2.35347	4.75841	8.64712	2.94942	2.35347	C2
3	3.00372	3.89419	1.54465	5.74280	6.23559	1.54465	C3
4	4.09791	6.17247	5.37640	3.79429	9.19782	3.79429	C4
5	1.65682	3.46792	3.71382	4.67577	6.40022	1.65682	C1
……	……	……	……	……		……	……
83	3.11724	3.40862	3.01533	5.80665	4.73586	3.01533	C3
84	4.08044	3.57118	2.68444	7.10816	4.20690	2.68444	C3
85	3.93434	6.21522	4.45724	2.41731	8.58812	2.41731	C4
86	3.81696	3.04167	3.06020	7.02831	3.33170	3.04167	C2
87	3.01116	4.13631	3.17737	4.79307	5.54751	3.01116	C1

The clustering results from the table above can be seen in the following table:

Table 9. K-Medoids Clustering Results

No	Cluster	Amount of Data	DBI Value
1	C1	33	1.27833
2	C2	15	
3	C3	17	
4	C4	17	
5	C5	5	

Meanwhile, the results of clustering using the K-Medoids algorithm consist of 5 clusters, with members in each cluster being cluster 1 having 33 members, cluster 2 having 15 members, cluster 3 having 17 members, cluster 4 having 17 members and cluster 5 having 5 members with a resulting DBI value of 1.27833.

.

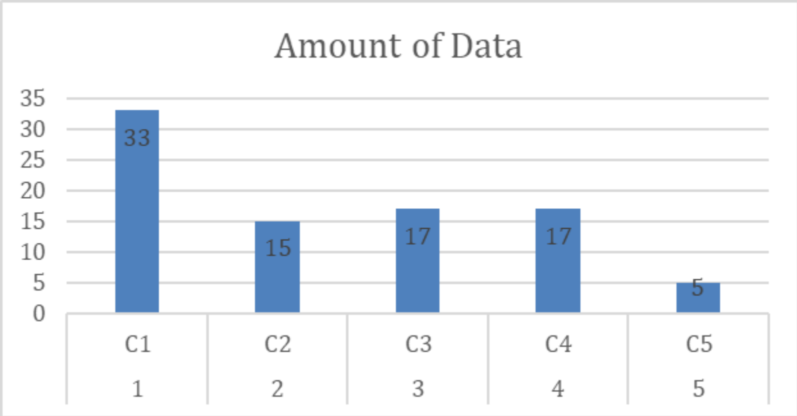


Fig 3. K-Medoids Clustering Results Graph

4.4 Comparison Results of K-Means and K-Medoids

The comparison results obtained from both algorithms can be seen in the table below:

Table 10. Results of K-Means and K-Medoids Algorithm

	Amount of Data					DBI
	C1	**C2**	**C3**	**C4**	**C5**	**Value**
K-Means	31	5	13	26	12	1,19010
K-Medoids	33	15	17	17	5	1,27833

The table above shows that the K-Means algorithm produces clusters with sizes 31, 5, 13, 26, and 12, and K-Medoids produces cluster sizes 33, 25, 17, 17, and 5. Meanwhile, K-medoids have a lower DBI value (1,27833) than K-Means (1,19010). This indicates that K-Means results in clusters with better quality. Based on the results, it can be concluded that although the number of clusters varies, K-Means consistently outperforms K-Medoids in terms of cluster quality measured by DBI. The comparison graphs for both algorithms can be seen in the following figure:

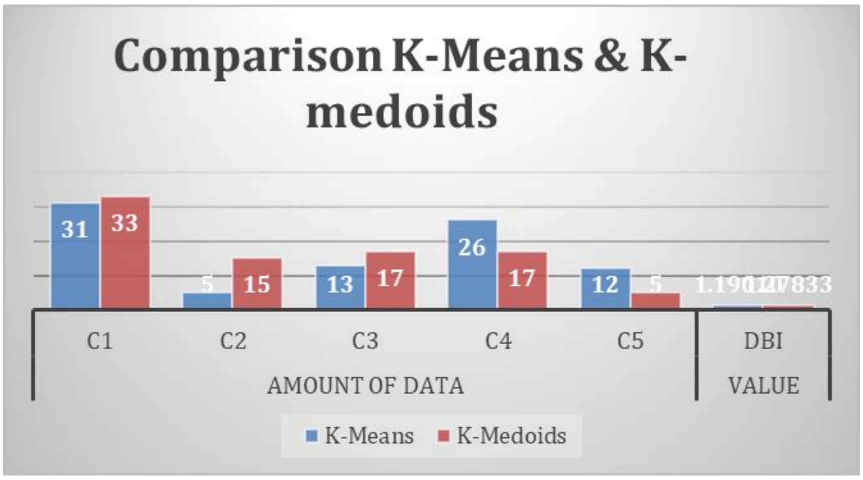


Fig 4. Comparison Graph of K-Means and K-Medoids Results

5

This comparison shows that the K-Means and K-Medoids algorithms are practical in determining study programs. Both algorithms can cluster student data into clusters with relatively similar sizes and compositions. However, the cluster evaluation results show that the K-Means algorithm produces better cluster quality than K-Medoids, as indicated by a lower Davies-Bouldin Index (DBI), with a DBI value of 1,19010 for K-Means and 1,27833 for K-Medoids. The performance of both algorithms shows some measurable differences. The K-Means algorithm produces clusters of 31, 5, 13, 26, and 12, while K-Medoids produces clusters of 33, 15, 17, 17 and 5. However, the DBI value shows that K-Means performs better in terms of clustering quality, with a DBI value of 1,19010 compared to 1,27833 for K-Medoids. Therefore, in this research, the K-Means algorithm is superior to K-Medoids.

5. Conclusion

From the research conducted using both the K-Means and K-Medoids algorithm, the following conclusions can be drawn:

1. The K-Means and K-Medoids algorithms are practical in determining study programs, as they can cluster student data into similar sizes and compositions. However, cluster evaluation results show that K-Means algorithm produces clusters of better quality than K-Medoids, as indicated by a lower Davies-Bouldin Index (DBI). The DBI values are 1,19010 for K-Means and 1,27833 for K-Medoids.
2. The performance of the K-Means and K-Medoids algorithm in the context of determining study programs reveals measurable differences. The K-Means algorithm produces clusters of 31, 5, 13, 26, and 12, while K-Medoids results in clusters of 33, 15, 17, 17 and 5. However, the DBI values indicate that K-Means perform better in clustering quality, with a DBI value of 1,19010 compared to K-Medoids 1,27833. Therefore, in this research, the K-Means algorithm produces better clusters than the K-Medoids algorithm.
3. Overall, it can be concluded that the K-Means algorithm is more recommended for use in this research based on the quality of the clusters produced, although both algorithms demonstrate good effectiveness.

References

[1] S. Anwar, T. Suprapti, G. Dwilestari, I. Ali, P. Studi Rekayasa Perangkat Lunak Jln Perjuangan No, and B. Kesambi Kota Cirebon, “Pengelompokkan HAsil Belajar Siswa Dengan Metode Clustering K-Means Program Studi Sistem Informasi Jln Perjuangan No 10B Kesambi Kota Cirebon 4),” *J. Sist. Inf. dan Teknol. Informasi)*, vol. 4, no. 2, pp. 60–72, 2022.

[2] Rahmawati and T. Arifin, “Penerapan Algoritma K-Means Untuk Pengelompokan Siswa Lolos Snmptn Di Sman 8 Bandung,” *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 184–190, 2020, doi: 10.51977/jti.v2i2.271.

[3] J. Faran and R. T. Aldisa, “Perbandingan Algoritma K-Means dan K-Medoids Dalam Pengelompokan Kelas Untuk Mahasiswa Baru Program Magister,” *J. Inf. Syst. Res.*, vol. 5, no. 2, pp. 509–519, 2024, doi: 10.47065/josh.v5i2.4753.

[4] D. Riyan Rizaldi, A. Doyan, Z. Fatimah, M. Zaenudin, and M. Zaini, "Strategies to Improve Teacher Ability in Using The Madrasah E-Learning Application During the COVID-19 Pandemic," *Int. J. Eng. Sci. Inf. Technol.*, vol. 1, no. 2, 2021, doi: 10.52088/ijesty.v1i2.47.

[5] E. Purnamasari, “Prediksi Perkembangan Nilai Impor Komoditas Utama,” *J. Inf. dan Teknol. Vol.*, vol. 5, no. 1, pp. 165–172, 2023, doi: 10.37034/jidt.v5i1.271.

[6] W. Febriani, G. W. Nurcahyo, and S. Sumijan, “Diagnosa Penyakit Rubella Menggunakan Metode Fuzzy Tsukamoto,” *J. Sistim Inf. dan Teknol.*, 2019, doi: 10.35134/jsisfotek.v1i3.4.

[7] Nurdin, Taufiq, Fajriana, and M. Z. Ulhaq, “Mapping System Model and Clustering of Fishery Products using K-Means Algorithm with Web GIS Approach,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 3, pp. 738–749, 2023.

[8] 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.*, 2021, doi: 10.1007/s11135-021-01176-w.

[9] M. Mahfut, “Identification and Efforts to Control Infection Odontoglossum ringspot virus (ORSV) on Orchid,” *Int. J. Eng. Sci. Inf. Technol.*, vol. 1, no. 1, 2021, doi: 10.52088/ijesty.v1i1.38.

[10] D. Abdullah, “Potensi teknologi informasi dan komunikasi dalam peningkatan mutu pembelajaran di kelas,” 2009.

[11] Y. Prihati, Suwarno, and A. Dharmawan, “Implementasi Algoritma K-Means Untuk Pemetaan Prestasi Akademik Siswa Disekolah Dasar Terang Bagi Bangsa Pati,” *Kinabalu*, vol. 11, no. 2, pp. 50–57, 2019.

[12] N. Noviyanto, “Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia,” *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 183–188, 2020, doi: 10.31294/p.v22i2.8808.

[13] W. Lestari, “Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari),” *Simkom*, vol. 4, no. 2, pp. 35–48, 2019, doi: 10.51717/simkom.v4i2.37.

[14] Yulia and M. Silalahi, “Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means,” *Indones. J. Comput. Sci.*, vol. 10, no. 1, 2021.

[15] S. W. Nengsih, I. Alfian, D. Aji, and S. Anwar, “Analisis Pengelompokan Penentuan Jurusan Siswa Sma Menggunakan Metode K-Means Clustering,” *J. Ilm. Betrik (Besemah Teknol. Inf. dan Komputer)*, no. 03, pp. 242–248, 2021.

[16] Y. Renatalia, M. Asfi, and R. Fahrudin, “Program Studi Menggunakan Metode,” vol. 10, no. 2, pp. 148–160, 2020.

[17] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, “Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food,” *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022.

[18] D. Abdullah and C. I. Erliana, “Model of Ict Goods Inventory Clustering Application Using K-Means Method,” *İlköğretim Online*, vol. 20, no. 1, pp. 1128–1132, 2021, doi: 10.17051/ilkonline.2021.01.116.

[19] 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, doi: 10.35957/jatisi.v9i1.1516.

[20] M. Faisal, N. Nurdin, F. Fajriana, and Z. Fitri, "Information and Communication Technology Competencies Clustering For Students For Vocational High School Students Using K-Means Clustering Algorithm," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 3, pp. 111–120, 2022, doi: 10.52088/ijesty.v2i3.318.

[21] W. Kurniawan, A. Rifai, ; Windu Gata, and D. Gunawan, “Analisis Algoritma K-Medoids Clustering Dalam Menentukan Pemesanan Hotel,” *Swabumi (Suara Wawasan Sukabumi) Ilmu Komputer, Manajemen, dan Sos.*, vol. 8, no. 2, pp. 182–187, 2020.

[22] B. Wira, A. E. Budianto, and A. S. Wiguna, “Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiwa Baru Tahun 2018 Di Universitas Kanjuruhan Malang,” *RAINSTEK J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 53–68, 2019, doi: 10.21067/jtst.v1i3.3046.

[23] B. Riyanto, “Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokkan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan),” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 562–568, 2019, doi: 10.30865/komik.v3i1.1659.

[24] G. B. Kaligis and S. Yulianto, “Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokkan

Kinerja Pegawai," *IT-Explore J. Penerapan Teknol. Inf. dan Komun.*, vol. 1, no. 3, pp. 179–193, 2022, doi: 10.24246/itexplore.v1i3.2022.pp179-193.

[25] M. Herviany, S. Putri Delima, T. Nurhidayah, and K. Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021, doi: 10.57152/malcom.v1i1.60.

[26] F. Fathurrahman, S. Harini, and R. Kusumawati, "Evaluasi Clustering K-Means Dan K-Medoid Pada Persebaran Covid-19 Di Indonesia Dengan Metode Davies-Bouldin Index (Dbi)," *J. Mnemon.*, vol. 6, no. 2, pp. 117–128, 2023, doi: 10.36040/mnemonic.v6i2.6642.