Jurnal Teknik Informatika (JUTIF)
Vol. 5, No. 4, sAugust 2024, pp. 559-577

DOI: https://doi.org/10.52436/1.jutif.2024.5.4.2215
p-ISSN: 2723-3863
e-ISSN: 2723-3871

ENHANCING EFFICIENCY IN DETERMINING QURAN LEARNING GROUPS: A WEBSITE-BASED K-MEANS ALGORITHM APPROACH AT NURUL JADID ISLAMIC BOARDING SCHOOL

Ikhwan Abdillah[1], Andi Wijaya[2], Kamil Malik[3]

Department of Informatics Engineering, Faculty of Engineering, Universitas Nurul Jadid, Indonesia
Email: [1]ikhwanabdillah1010@gmail.com, [2]mr.andiwijaya@unuja.ac.id, [3]kamilmalik@unuja.ac.id

(Article received: June 5, 2024; Revision: June 14, 2024; published: August 21, 2024)

Abstract

This research aims to develop a web-based application system using the K-Means algorithm to group students in Quran coaching at the Nurul Jadid Islamic Boarding School in Paiton, Probolinggo. The need for this system is based on the importance of efficiency and accuracy in determining student coaching groups based on their abilities in reading the Quran, including Tajweed, fluency, and memorization scores. This research method involves data analysis from 412 students. The data is processed using the K-Means algorithm to group students into three skill categories: "Good", "Sufficient", and "Poor". The grouping results provide objective and accurate guidance in determining suitable coaching groups for each student. The research results show that the K-Means algorithm is effective in grouping students, thereby improving the efficiency and accuracy of the coaching process. The implementation of web-based technology facilitates access and use of the system by administrators and coaching participants, ensuring that the grouping and coaching processes become faster, more accurate, and more objective. In conclusion, this research successfully develops a more responsive and efficient Quran coaching system, which not only solves specific problems at the Nurul Jadid Islamic Boarding School but also makes a significant contribution to the development of similar systems in other Islamic educational institutions.

Keywords : *K-Means, Student Clustering, Quran Coaching, Data Mining, Website Technology*.

1. INTRODUCTION

"The Islamic Boarding School Nurul Jadid, located in Paiton Probolinggo, plays a crucial role in delivering Islamic education with a focus on Islamic brotherhood [1]. One of the primary educational activities at Nurul Jadid is Quran training for students. A significant component of this training is grouping students based on their skills and individual needs. Currently, this grouping is done manually, requiring deep knowledge of each student's abilities and the varied learning targets of different groups. However, manual methods are susceptible to human errors in assessing students and assigning groups, besides being time-consuming and labor-intensive [2].

To enhance the efficiency and accuracy of Quran training group assignments, this research proposes the use of technology and data analysis methods, particularly data mining to identify relevant patterns and clustering using the K-Means algorithm [3]. This research underscores the need to improve the efficiency and accuracy of group assignments in Quran training [4]. Objectively evaluating student abilities poses challenges due to each student's unique characteristics, requiring considerations such as Quran reading ability, understanding of Tajwid (rules of Quranic recitation), and fluency in reading. Additionally, there are varying expected learning outcomes among training groups; some students may need additional assistance with basic Quranic concepts while others are ready for more advanced lessons [5].

In this scenario, data mining techniques and clustering offer more objective and efficient methods for group assignments [6]. Data analysis can uncover patterns in Quran training test results, which can be used to assess student abilities [7]. Clustering, especially with the K-Means algorithm, allows for homogeneous grouping based on identified data characteristics [8]. The implementation of technology and data analysis methods in this research aims to address the challenges faced by the Islamic Boarding School Nurul Jadid. By leveraging technology, the grouping process can become faster, more accurate, and more objective, thereby enhancing student motivation through personalized training experiences tailored to individual needs [9]. The implementation plan includes using Python for data analysis, the K-Means algorithm, and developing a web interface with PHP for ease of access and use by administrators and training participants [10]. An SQL database will be used to ensure organized data storage that is easily accessible [11].

This research aims to address specific issues at the Islamic Boarding School Nurul Jadid and contribute to the development of a more responsive and efficient training system in similar educational institutions [12]. Furthermore, insights from related studies strengthen the foundation of this research. For instance, Ai Rohmah, Falentino Sembiring, and Adhitia Erfina (2021) highlighted the use of the K-Means algorithm to classify levels of learning barriers in the context of distance learning during the COVID-19 pandemic [13]. Similarly, Dian Permata Sari (2021) applied the K-Means algorithm to determine the spread levels of the COVID-19 pandemic in West Sumatra, clustering regions based on positive cases and virus spread [14]. Moreover, Achmad Dimyati (2023) focused on using K-Means to evaluate academic grades of students at TPQ Darussalamah based on the Laravel Framework [15]. These studies collectively support the adoption of data mining techniques like K-Means for improving educational practices [16]. The integration of K-Means clustering into educational settings has shown promise in various studies, including its application in predicting student academic performance, analyzing students' academic data, and classifying student learning achievements [17]. Such applications illustrate the versatility of K-Means in educational data analysis and its potential to address complex challenges in student assessment and educational management [18]. By leveraging these methodologies, educational institutions can streamline processes, personalize learning experiences, and foster academic success [19]. This research builds on these foundations to propose a data-driven approach for enhancing Quran training at Nurul Jadid and similar institutions.

2. METHOD

The Quran Coaching Group Determination application system was developed using PHP programming language with the Laravel framework and an SQL database. The Knowledge Discovery in Databases (KDD) methodology was employed to analyze data in a structured way to extract new information [20]. Data mining processes involved applying clustering methods to uncover hidden information. The primary goal is to analyze the results of Quran coaching tests conducted by the Religious Affairs Bureau of Nurul Jadid Islamic Boarding School, Paiton, Probolinggo.

This research includes several stages outlining the research design: (a) Data Collection, (b) Understanding K-Means Clustering Literature on Data, (c) Data Processing, (d) Website Development, (e) Data Clustering Steps, and (f) Evaluation of Clustering Results. Figure 1 below illustrates the flowchart for determining clusters using the K-Means algorithm.

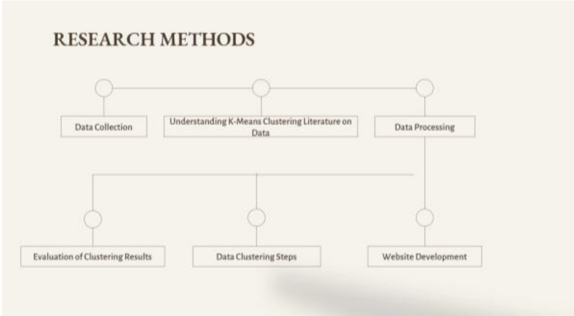

Figure 1. Research Steps

2.1. Data Collection

The data gathered consists of records from Quran coaching tests conducted by the Religious Affairs Bureau at Nurul Jadid Islamic Boarding School, Paiton, Probolinggo. A total of 412 test result records were collected, each containing individual scores for each tested surah. The collected data includes Tajweed, fluency, and memorization scores for each student involved in the Quran coaching activities.

2.2. Understanding K-Means Clustering Literature on Data

Once the data is properly collected, the next step involves understanding the concepts and theories underpinning data grouping using the K-Means Clustering method [21]. The process of grasping literature related to the application of K-Means Clustering on the gathered data entails several steps:

a) **Literature Review**: Conduct searches and reviews of literature and research related to K-Means Clustering by examining articles, books, and scientific publications that discuss the foundational concepts, algorithms, and applications of K-Means Clustering across various fields [22].
b) **Mathematical Concepts**: Explore the mathematical foundations of K-Means Clustering, including the calculations of distances between data points and centroids, as well as the optimization principles that drive the algorithm. This provides a thorough understanding of how the algorithm operates and its efficiency in data grouping [23].
c) **Case Study Analysis**: Analyze case studies and practical implementations of K-Means Clustering in diverse contexts and sectors by reviewing how this method has been successfully used in real-world data analysis, noting its strengths and limitations[9].
d) **Method Evaluation**: Assess the relevance and potential application of the K-Means method for analyzing the collected data by determining whether this method can effectively uncover significant patterns or insights within the data [24].

2.3. Data Processing

In this stage, data is cleaned, transformed, and prepared for clustering. The preprocessing steps include:

a) Removing incomplete or irrelevant data.
b) Normalizing data to ensure uniform scaling across all variables.
c) Addressing any outliers, if necessary.
d) Converting data into a format suitable for the K-Means algorithm.

2.4. Website Development

The next step involves creating the website interface to fulfill the system requirements. This web-based system is built using PHP with the Laravel Framework and an SQL database to facilitate easy access for calculating groupings in Al-Qur'an mentoring. The application of the K-Means Clustering Algorithm within the Laravel Framework includes performing calculations and applying them to the processed data.

2.5. Data Clustering Steps

The next step involves clustering the data using the k-means algorithm. This process includes testing the clustering both manually using Excel with a .xlsx format and using Python in Google Colab to ensure consistent data accuracy. The steps involved are as follows:

a) Select the data to be clustered.
b) Determine the number of clusters (k) to be created.
c) Randomly choose initial cluster centers (centroids).
d) Calculate the distance of each data point to the centroid using the Euclidean Distance formula.

$$D_{(i,j)} = \sqrt{\left(P_{1i} - Q_{1j}\right)^2 + \left(P_{2i} - Q_{2j}\right)^2 + \dots + \left(P_{ki} - Q_{kj}\right)^2}$$

where $D_{(i.j)}$ represents the distance from data point i *to cluster center* j, X_{ki} is the k-th attribute of data point i, and X_{kj} is the k-th attribute of cluster center j.

e) Assign each data point to the cluster with the nearest center.
f) Recalculate the cluster centers based on the mean of the data points in each cluster. The formula to calculate the cluster center is.

$$C_i = \frac{P1 + P2 + P3 + \cdots + Pn}{\sum P}$$

where C_i is the cluster center, Pn *is the* n*-th data point for the* n*-th attribute, and* $\sum P$ is the sum of all data points in the cluster.

g) Repeat the iteration until the cluster assignments remain unchanged.

2.6. Evaluation of Clustering Results

After the clustering process is completed, the results will be evaluated using the Davies-Bouldin Index calculation formula. The Davies-Bouldin Index calculation is based on the principle of maximizing inter-cluster distance while minimizing intra-cluster distance [25]. Therefore, it can be stated that the smaller the Davies-Bouldin Index value, the more optimal the clustering scheme. This evaluation will assist in determining the optimal number of clusters and in evaluating the quality of clustering produced by the K-Means algorithm [26]. The formula for calculating the Davies-Bouldin Index can be presented as follows:

$$DBI = \frac{1}{k}\sum_{i=1}^{k}\{R_i\}$$

where $R_i = max_{j=1,\dots,k,i,\neq j} R_{ij}, R_{ij} = \frac{S_1+S_j}{d_{ij}}$

and $S_1 = \left[\frac{1}{n}\sum_{x\in n_i} d^2(x, v_i)\right]^{\frac{1}{2}}$

In this equation, the symbol k represents the number of clusters, and $R_{i,j}$ is a measure of similarity between the values n_i dan n_j. The symbol S_i represents the dispersion size of the i-th cluster, where $i = 1.2.3, \dots, k$. The symbol d_{ij} indicates the distance between the centroid of cluster i and the centroid of cluster j $\left(d_{ij} = d_{ji}\right)$. The symbol n_i indicates the number of members in cluster i, where $i = 1.2.3, \dots, k$. Finally, v_i represents the centroid value of cluster n_i.

3. ANALYSIS RESULTS

The analysis results of predicting the grouping of Al-Qur'an mentoring using the K-Means algorithm are conducted by considering the tajwid score, fashohah score, and memorization score. Data from 412 students who participated in the Al-Qur'an mentoring test are collected and used in the grouping process.

The analysis results indicate that Al-Qur'an mentoring can be categorized into three different groups based on their proficiency levels in reading Al-Qur'an. These groups are labeled "good," "fairly good," and "less good" according to the characteristics of each group. The following are the results of implementing the K-Means algorithm-based website in determining the Al-Qur'an mentoring groups.

3.1 Data Collection

In this stage, 412 collected data points are utilized. These data encompass assessments from each surah examined in the Al-Qur'an mentoring. Subsequently, the scores from each test are aggregated to obtain the total score of the data. The following presents the results of the Al-Qur'an mentoring tests:

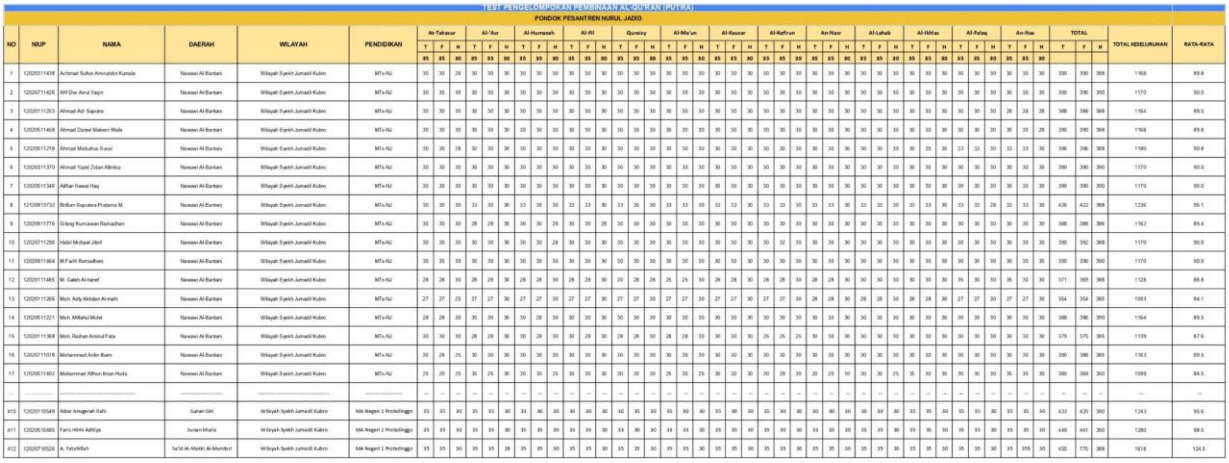

Figure 2. Results of Al-Qur'an Mentoring Tests Data.

In Figure 2, the total number of successfully acquired data points is depicted, while the data to be processed for determining the Al-Qur'an mentoring groups using the K-Means algorithm consist of the total scores of each data point.

3.2 Data Processing

In this stage, data processing is conducted to prepare the data. Data processing involves steps such as data cleansing, normalization, and data labeling or initialization..

a) Data Cleansing

In the data cleansing stage, irrelevant data rows are removed for K-Means calculation. These data rows include columns not utilized in the analysis, such as Region, District, institution, and values from several tested surahs, including Surah At-Takasur, Al-'asr, Al-Humazah, Al-Fil, Quraisy, Al-Ma'un, Al-Kautsar, Al-Kafirun, An-Nasr, Al-Lahab, Al-Ikhlas, Al-Falaq, and An-Nas. This is performed because K-Means calculation only utilizes the total score from all tested surahs.

Table 1. Results of Irrelevant Data Cleansing.

NO	NIUP	NAME	VALUE		
			T	F	H
1	12020311438	Achmad Sulton Amiruddin K	390	390	390
2	12020711426	Afif Dwi Ainul Yaqin	390	390	388
3	12020111253	Ahmad Adi Saputra	388	388	388
4	12020511458	Ahmad Daniel Mateen Wafa	390	390	388
5	12020511218	Ahmad Misbahus Sururi	396	396	388
6	12020311370	Ahmad Yazid Zidan Altintop	390	390	390
7	12020511346	Alifian Nawal Haq	390	390	390
8	12120913732	Brillian Saputera Pratama M.	426	422	388
9	12020911776	Gilang Kurniawan Ramadhan	388	388	386
10	12020711290	Habil Michael Jibril	390	392	388
11	12020911464	M Farih Romadhoni	390	390	390
12	12020111485	M. Galeh Al-hanef	371	369	388
13	12020111266	Moh. Adly Akhdan Al-mahi	354	354	385
14	12020511221	Moh. Miftahul Muhit	388	386	390
15	12020111368	Moh. Raihan Aminul Fata	379	375	385
….	……………	……………………………	…	…	…
408	11720502268	Ahmad Fahmy Kholidy	350	350	380
409	11720902259	Ahmad Imong Budiono	430	408	390
410	11720503113	Ahmad Zaidan Salim	305	310	390
411	11720302242	Aufil Ghulam	335	360	365
412	11820701873	Danil Faizin	335	335	365

b) Data Normalization

Next, data normalization is not performed as a preparatory step before the clustering process using the K-Means algorithm. In this case, it is because the data to be processed already has a uniform scale range for each feature, and there are no features that dominate the clustering process.

3.3 Website Development

The next step in this research is the development of a website foundation using PHP programming language, Laravel framework, and SQL database. The purpose of this website is to present the processed data in an easily understandable and usable format. This website is also expected to facilitate administrators in accessing the data generated from the k-means algorithm calculations. Below are the appearance and features of the k-means algorithm website.

a) Dashboard

Figure 3. Dashboard Interface Display.

b) Variable Management

Figure 4. Variable Management Interface Display.

c) Data Management

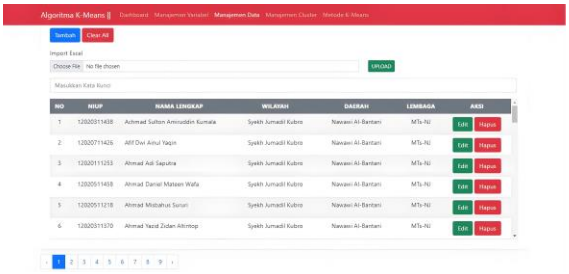

Figure 5. Data Management Interface Display.

d) Cluster Management

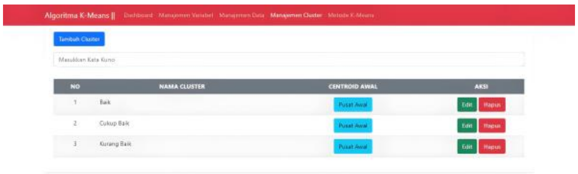

Figure 6. Cluster Management Interface Display.

e) Initial Centroid Determination

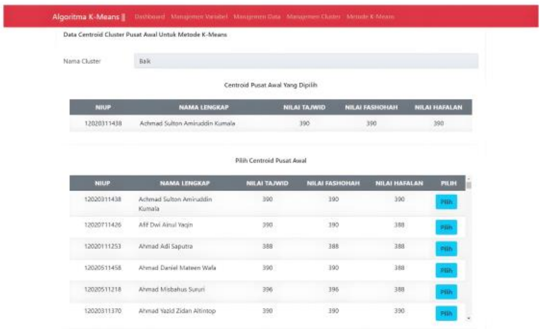

Figure 7. Initial Centroid Determination Interface Display.

f) K-Means Algorithm Iteration

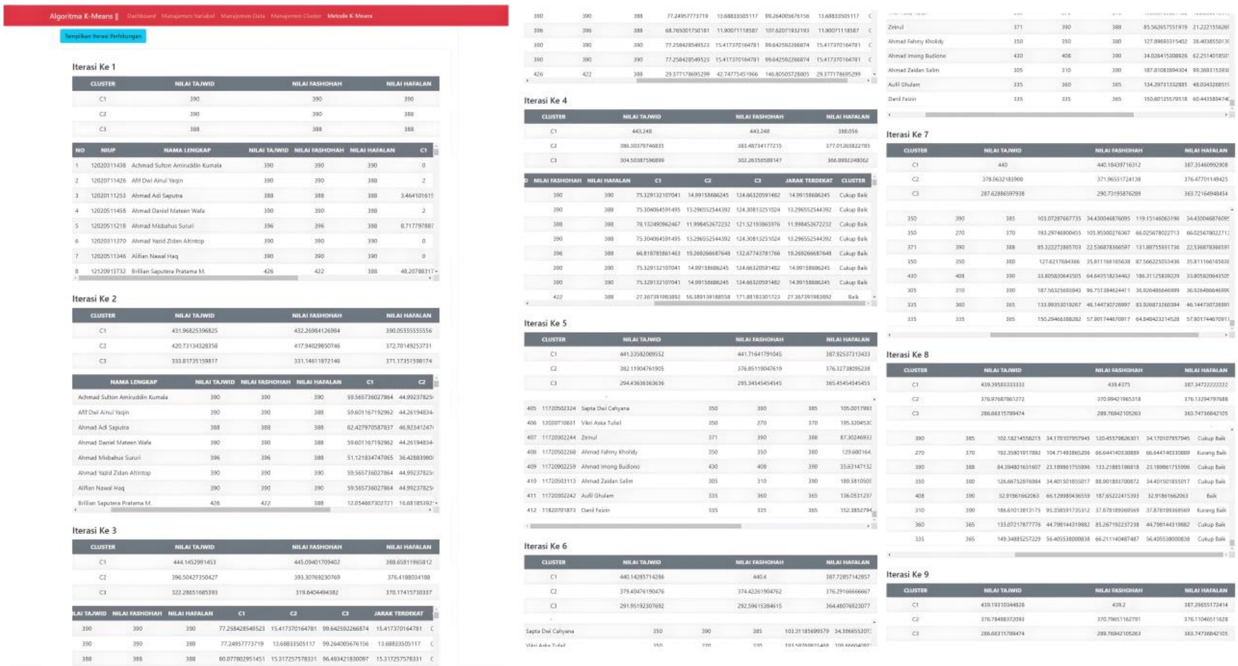

Figure 8. K-Means Iteration Calculation Interface Display.

g) Clustering Results

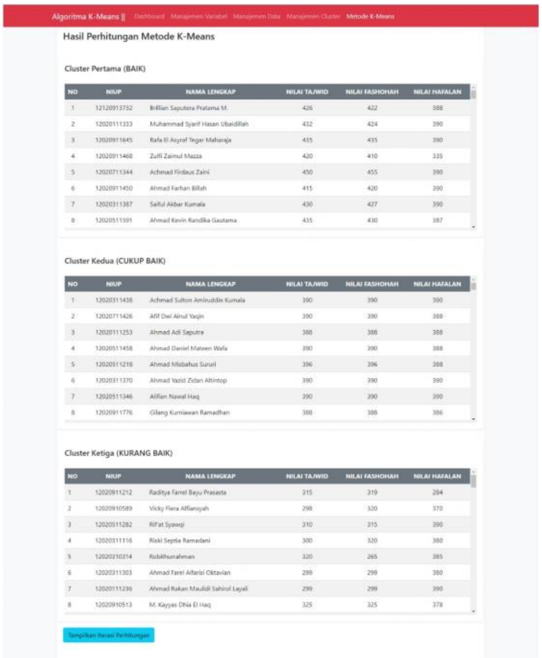

Figure 9. K-Means Iteration Calculation Interface Display.

3.4 Clustering Data Calculation

In this study, the clustering process was conducted based on the data related to Tajwid Score, Fashohah Score, and Memorization Score. After successfully processing the data, it was determined that the objects would be grouped into 3 clusters. The choice of $k = 3$ or $c = 3$ for clustering was driven by the analysis indicating that the data could be effectively categorized into three distinct groups. This decision was informed by the distribution and patterns observed in the scores, which suggested three identifiable clusters with varying levels of Tajwid proficiency, Fashohah clarity, and Memorization skills. By selecting $k = 3$, the study aimed to provide a meaningful segmentation of the data, enabling clearer insights into the relationships between the variables and facilitating a more insightful interpretation of the results. Thus, $k = 3$ was pivotal in achieving a structured understanding of how these scores correlate and differ across the clustered groups. The following is the data clustering process for each calculation model.

a) Determining Initial Cluster Centers

In the initial cluster, the cluster centers are randomly determined. In the first trial (Iteration 1), 3 d ata points are randomly selected as the initial center points for calculating the distance from all cluster groups to be formed.

Number of Clusters = 3 (Good, Fairly Good, Less Good)
Number of Data = 412 Data
Number of Variables = 3 (Tajwid Score, Fashohah Score, Memorization Score)

Table 2. Initial Data of First Iteration Center

NIUP	NAME	T	F	H	CLUSTER
12020311438	Achmad Sulton Amiruddin K	390	390	390	C1
12020711426	Afif Dwi Ainul Yaqin	390	390	388	C2
12020111253	Ahmad Adi Saputra	388	388	388	C3

From Table 2, it can be seen that each data has a cluster center value.

C1 = (390 390 390)
C2 = (390 390 388)
C3 = (388 388 388)

b) Calculation of Centroid Distances

In this phase, the distances between each data point and the cluster centers are determined using Euclidean Distance. The distance calculation from each data point to the first cluster center is outlined below:

For Cluster 1 (1):

$$\sqrt{((390-390)^2 + (390-390)^2 + (390-390)^2)}$$

For Cluster 2 (2):

$$\sqrt{((390-390)^2 + (390-390)^2 + (388-390)^2)}$$

For Cluster 3 (3):

$$\sqrt{((388-390)^2 + (388-390)^2 + (388-390)^2)}$$

This process continues for all data points and all clusters. After obtaining these distances in the first iteration, they are recorded in Table 3.

Table 3. Calculation Results in the First Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	0	2	3,464101615
2	Afif Dwi Ainul Yaqin	2	0	2,828427125
3	Ahmad Adi Saputra	3,464101615	2,828427125	0
…	…………………	…	…	…
410	Ahmad Zaidan Salim	116,7433082	113,9166362	113,9166362
411	Aufil Ghulam	66,73829485	64,20280368	64,20280368
412	Danil Faizin	81,11103501	78,40280607	78,40280607

Next, the second iteration involves calculating new centroid positions. This is achieved by finding the mean of the data within each cluster. The means for each attribute (Tajwid, Fashohah, and Hafalan) are computed and presented as the new centroids in Table 4.

Table 4. New Centroids in the Second Iteration

NEW CENTROIDS	T	F	H
C1	431,968253968254	432,269841269841	390,055555555556
C2	420,731343283582	417,940298507463	372,701492537313
C3	333,817351598174	331,146118721461	371,173515981735

After computing the new centroids, the distances from every data point to these centroids are re-evaluated utilizing the Euclidean Distance method. This step replicates the methodology described in the initial iteration.

The second iteration commences with computing the Euclidean Distance between each data point and the updated centroids of every cluster. This procedure closely resembles the approach undertaken in the First Iteration.

For each data point, its distance to the first cluster center (C1) is computed as follows:

$$C1\ (1) = \sqrt{((390-431{,}968253968254)^2 + (390-432{,}269841269841)^2 + (390-390{,}055555555556)^2)}$$

$$C1\ (2) = \sqrt{((390-431{,}968253968254)^2 + (390-432{,}269841269841)^2 + (388-390{,}055555555556)^2)}$$

$$C1\ (3) = \sqrt{((388-431{,}968253968254)^2 + (388-432{,}269841269841)^2 + (388-390{,}055555555556)^2)}$$

And so on until C1(412).

Then, the distance from each data to the second cluster center (C2) is computed using a similar formula:

$$C2\ (1) = \sqrt{((390-420{,}731343283582)^2 + (390-417{,}940298507463)^2 + (390-372{,}701492537313)^2)}$$

$$C2\ (2) = \sqrt{((390-420{,}731343283582)^2 + (390-417{,}940298507463)^2 + (388-372{,}701492537313)^2)}$$

$$C2\ (3) = \sqrt{((388-420{,}731343283582)^2 + (388-417{,}940298507463)^2 + (388-372{,}701492537313)^2)}$$

And so on until C2(412).

Finally, the distance from each data to the third cluster center (C3) is calculated as follows:

$$C3\ (1) = \sqrt{((390-333{,}817351598174)^2 + (390-331{,}146118721461)^2 + (390-371{,}173515981735)^2)}$$

$$C3\ (2) = \sqrt{((390-333{,}817351598174)^2 + (390-331{,}146118721461)^2 + (388-371{,}173515981735)^2)}$$

$$C3\ (3) = \sqrt{((388-333{,}817351598174)^2 + (388-331{,}146118721461)^2 + (388-371{,}173515981735)^2)}$$

And so on until C3(412).

The results of these distance calculations are recorded in Table 5. Next, because the results of the Second Iteration are not the same as the First Iteration, the process continues with the Third Iteration. This process involves calculating the positions of the new centroids based on the average of the data in each cluster, as done in the First and Second Iterations.

Table 5. Calculation Results in the Second Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	59,56573603	44,99237826	83,51470424
2	Afif Dwi Ainul Yaqin	59,60116719	44,26194834	83,08670103
3	Ahmad Adi Saputra	62,42797059	46,92341248	80,31969726
…	…………………	…	…	…
410	Ahmad Zaidan Salim	176,2692676	159,1982733	40,39844787
411	Aufil Ghulam	123,5052756	103,7605628	29,53061838
412	Danil Faizin	139,6139146	119,533549	7,37315114

In the third iteration, to find the new centroids, the average value of the data grouped in the second iteration is taken. These new centroids are then recorded in Table 6.

Table 6. New Centroids in the Third Iteration

NEW CENTROIDS	T	F	H
C1	444,145299145299	445,094017094017	388,65811965812
C2	396,504273504274	393,307692307692	376,418803418803
C3	322,286516853933	319,640449438202	370,174157303371

After computing the new centroids, the distances from every data point to these centroids are re-evaluated utilizing the Euclidean Distance method. This step repeats the methodology outlined in the previous iteration.

This iteration begins with computing the Euclidean Distance between each data point and the newest centroids of each cluster. This procedure is similar to the approach taken in the previous iteration.

For each data point, its distance to the first cluster center (C1) is computed as follows:

$$C1\,(1) = \sqrt{((390-444{,}145299145299)^2 + (390-445{,}094017094017)^2 + (390-388{,}65811965812)^2)}$$

$$C1\,(2) = \sqrt{((390-444{,}145299145299)^2 + (390-445{,}094017094017)^2 + (388-388{,}65811965812)^2)}$$

$$C1\,(3) = \sqrt{((388-444{,}145299145299)^2 + (388-445{,}094017094017)^2 + (388-388{,}65811965812)^2)}$$

And so on until C1(412).

Then, the distance from each data to the second cluster center (C2) is computed using a similar formula:

$$C2\,(1) = \sqrt{((390-396{,}504273504274)^2 + (390-393{,}307692307692)^2 + (390-376{,}418803418803)^2)}$$

$$C2\,(2) = \sqrt{((390-396{,}504273504274)^2 + (390-393{,}307692307692)^2 + (388-376{,}418803418803)^2)}$$

$$C2\,(3) = \sqrt{((388-396{,}504273504274)^2 + (388-393{,}307692307692)^2 + (388-376{,}418803418803)^2)}$$

And so on until C2(412).

Finally, the distance from each data to the third cluster center (C3) is calculated as follows:

$$C3\,(1) = \sqrt{((390-322{,}286516853933)^2 + (390-319{,}640449438202)^2 + (390-370{,}174157303371)^2)}$$

$$C3\,(2) = \sqrt{((390-322{,}286516853933)^2 + (390-319{,}640449438202)^2 + (388-370{,}174157303371)^2)}$$

$$C3\,(3) = \sqrt{((388-322{,}286516853933)^2 + (388-319{,}640449438202)^2 + (388-370{,}174157303371)^2)}$$

And so on until C3(412).

From these calculations in the third iteration, the distances from each data point to each cluster are obtained as listed in Table 7.

Table 7. Calculation Results in the Third Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	77,25842855	15,41737016	99,64259227
2	Afif Dwi Ainul Yaqin	77,24957774	13,68833505	99,26400568
3	Ahmad Adi Saputra	80,07780295	15,31725758	96,48342183
…	…………………	…	…	…
410	Ahmad Zaidan Salim	193,942281	124,4895681	28,01474556
411	Aufil Ghulam	140,404397	70,87007195	42,62977691
412	Danil Faizin	156,8218593	85,51579803	20,59903762

Next, because the result of the third iteration differs from the second iteration, the process continues with the fourth iteration. The positions of the new centroids in the fourth iteration are calculated by taking the average of the data included in each group or centroid as in the previous iteration.

In the fourth iteration, to find the new centroids, the average value of the data grouped in the third iteration is taken. These new centroids are then recorded in Table 8.

Table 8. New Centroids in the Fourth Iteration

NEW CENTROIDS	T	F	H
C1	443,248	443,248	388,056
C2	386,303797468354	383,487341772152	377,012658227848
C3	304,503875968992	302,263565891473	366,899224806202

After computing the new centroids, the distances from every data point to these centroids are re-evaluated utilizing the Euclidean Distance method. This step repeats the methodology outlined in the previous iteration.

This iteration begins with computing the Euclidean Distance between each data point and the newest centroids of each cluster. This procedure is similar to the approach taken in the previous iteration.

For each data point, its distance to the first cluster center (C1) is computed as follows:

C1 (1) = $\sqrt{((390-443{,}248)^2+(390-443{,}248)^2+(390-388{,}056)^2)}$

C1 (2) = $\sqrt{((390-443{,}248)^2+(390-443{,}248)^2+(388-388{,}056)^2)}$

C1 (3) = $\sqrt{((388-443{,}248)^2+(388-443{,}248)^2+(388-388{,}056)^2)}$

And so on until C1(412).

Then, the distance from each data to the second cluster center (C2) is computed using a similar formula:

C2 (1) = $\sqrt{((390-386{,}303797468354)^2+(390-383{,}487341772152)^2+(390-377{,}012658227848)^2)}$

C2 (2) = $\sqrt{((390-386{,}303797468354)^2+(390-383{,}487341772152)^2+(388-377{,}012658227848)^2)}$

C2 (3) = $\sqrt{((388-386{,}303797468354)^2+(388-383{,}487341772152)^2+(388-377{,}012658227848)^2)}$

And so on until C2(412).

Finally, the distance from each data to the third cluster center (C3) is calculated as follows:

C3 (1) = $\sqrt{((390-304{,}503875968992)^2+(390-302{,}263565891473)^2+(390-366{,}899224806202)^2)}$

C3 (2) = $\sqrt{((390-304{,}503875968992)^2+(390-302{,}263565891473)^2+(388-366{,}899224806202)^2)}$

C3 (3) = $\sqrt{((388-304{,}503875968992)^2+(388-302{,}263565891473)^2+(388-366{,}899224806202)^2)}$

And so on until C3(412).

From these calculations in the fourth iteration, the distances from each data point to each cluster are obtained as listed in Table 9.

Table 9. Calculation Results in the Fourth Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	75,32913211	14,99158686	124,6632059
2	Afif Dwi Ainul Yaqin	75,30406459	13,29655254	124,3081325
3	Ahmad Adi Saputra	78,13249096	11,99845267	121,5219387
…	…………………	…	…	…
410	Ahmad Zaidan Salim	192,0190567	110,3601737	24,36687026
411	Aufil Ghulam	138,4898485	57,68915683	65,32316939
412	Danil Faizin	154,8122674	71,60590691	44,78051757

Furthermore, since the results of the fourth iteration differ from the third iteration, the process continues with the fifth iteration. The positions of the new centroids in the fifth iteration are calculated by taking the average of the data included in each group or centroid as in the previous iteration.

In the fifth iteration, to find the new centroids, the average value of the data grouped in the fourth iteration is taken. These new centroids are then recorded in Table 10.

Table 10. New Centroids in the Fifth Iteration

NEW CENTROIDS	T	F	H
C1	441,335820895522	441,716417910448	387,925373134328
C2	382,119047619048	376,851190476191	376,327380952381
C3	294,436363636364	295,345454545455	365,454545454545

After computing the new centroids, the distances from every data point to these centroids are re-evaluated utilizing the Euclidean Distance method. This step repeats the methodology outlined in the previous iteration. This iteration begins with computing the Euclidean Distance between each data point and the newest centroids of each cluster. This procedure is similar to the approach taken in the previous iteration.

For each data point, its distance to the first cluster center (C1) is computed as follows:

C1 (1) = $\sqrt{((390-441{,}335820895522)^2+(390-441{,}716417910448)^2+(390-387{,}925373134328)^2)}$

C1 (2) = $\sqrt{((390-441{,}335820895522)^2+(390-441{,}716417910448)^2+(388-387{,}925373134328)^2)}$

C1 (3) = $\sqrt{((388-441{,}335820895522)^2+(388-441{,}716417910448)^2+(388-387{,}925373134328)^2)}$

And so on until C1(412).

Then, the distance from each data to the second cluster center (C2) is computed using a similar formula:

C2 (1) = $\sqrt{((390-382{,}119047619048)^2+(390-376{,}851190476191)^2+(390-376{,}327380952381)^2)}$

C2 (2) = $\sqrt{((390-382{,}119047619048)^2+(390-376{,}851190476191)^2+(388-376{,}327380952381)^2)}$

C2 (3) = $\sqrt{((388-382{,}119047619048)^2+(388-376{,}851190476191)^2+(388-376{,}327380952381)^2)}$

And so on until C2(412).

Finally, the distance from each data to the third cluster center (C3) is calculated as follows:

C3 (1) = $\sqrt{((390-294{,}436363636364)^2+(390-295{,}345454545455)^2+(390-365{,}454545454545)^2)}$

C3 (2) = $\sqrt{((390-294{,}436363636364)^2+(390-295{,}345454545455)^2+(388-365{,}454545454545)^2)}$

C3 (3) = $\sqrt{((388-294{,}436363636364)^2+(388-295{,}345454545455)^2+(388-365{,}454545454545)^2)}$

And so on until C3(412).

From these calculations in the fifth iteration, the distances from each data point to each cluster are obtained as listed in Table 11.

Table 11. Calculation Results in the Fifth Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	72,89896066	20,54120527	136,7273598
2	Afif Dwi Ainul Yaqin	72,86947206	19,26786542	136,3825102
3	Ahmad Adi Saputra	75,69787918	17,17939435	133,5938485
…	…………………	…	…	…
410	Ahmad Zaidan Salim	189,581051	102,9726647	30,47663785
411	Aufil Ghulam	136,0531238	51,30766832	76,32709515
412	Danil Faizin	152,3852794	64,03152624	56,72828379

Furthermore, since the results of the fifth iteration differ from the fourth iteration, the process continues with the sixth iteration. The positions of the new centroids in the sixth iteration are calculated by taking the average of the data included in each group or centroid as in the previous iteration.

In the sixth iteration, to find the new centroids, the average value of the data grouped in the fifth iteration is taken. These new centroids are then recorded in Table 12.

Table 12. New Centroids in the Sixth Iteration

NEW CENTROIDS	T	F	H
C1	440,142857142857	440,4	387,728571428571
C2	379,404761904762	374,422619047619	376,291666666667
C3	291,951923076923	292,596153846154	364,480769230769

After computing the new centroids, the distances from every data point to these centroids are re-evaluated utilizing the Euclidean Distance method. This step repeats the methodology outlined in the previous iteration. This iteration begins with computing the Euclidean Distance between each data point and the newest centroids of each cluster. This procedure is similar to the approach taken in the previous iteration.

For each data point, its distance to the first cluster center (C1) is computed as follows:

C1 (1) = $\sqrt{((390-440{,}142857142857)^2+(390-440{,}4)^2+(390-387{,}728571428571)^2)}$

C1 (2) = $\sqrt{((390-440{,}142857142857)^2+(390-440{,}4)^2+(388-387{,}728571428571)^2)}$

C1 (3) = $\sqrt{((388-440{,}142857142857)^2+(388-440{,}4)^2+(388-387{,}728571428571)^2)}$

And so on until C1(412).

Then, the distance from each data to the second cluster center (C2) is computed using a similar formula:

C2 (1) = $\sqrt{((390-379{,}404761904762)^2+(390-374{,}422619047619)^2+(390-376{,}291666666667)^2)}$

C2 (2) = $\sqrt{((390-379{,}404761904762)^2+(390-374{,}422619047619)^2+(388-376{,}291666666667)^2)}$

C2 (3) = $\sqrt{((388-379{,}404761904762)^2+(388-374{,}422619047619)^2+(388-376{,}291666666667)^2)}$

And so on until C2(412).

Finally, the distance from each data to the third cluster center (C3) is calculated as follows:

C3 (1) = $\sqrt{((390-291{,}951923076923)^2+(390-292{,}596153846154)^2+(390-364{,}480769230769)^2)}$

C3 (2) = $\sqrt{((390-291{,}951923076923)^2+(390-292{,}596153846154)^2+(388-364{,}480769230769)^2)}$

C3 (3) = $\sqrt{((388-291{,}951923076923)^2+(388-292{,}596153846154)^2+(388-364{,}480769230769)^2)}$

And so on until C3(412).

From these calculations in the sixth iteration, the distances from each data point to each cluster are obtained as listed in Table 13.

Table 13. Calculation Results in the Sixth Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	71,13104463	23,29876113	140,5423985
2	Afif Dwi Ainul Yaqin	71,09528673	22,18104905	140,1930414
3	Ahmad Adi Saputra	73,92368514	19,88236558	137,4055354
…	…………………	…	…	…
410	Ahmad Zaidan Salim	187,8108389	99,36931539	33,53173588
411	Aufil Ghulam	134,2973133	48,03432685	79,97927859
412	Danil Faizin	150,6013558	60,44358947	60,42758225

Furthermore, since the results of the sixth iteration differ from the fifth iteration, the process continues with the seventh iteration. The positions of the new centroids in the seventh iteration are calculated by taking the average of the data included in each group or centroid as in the previous iteration.

In the seventh iteration, to find the new centroids, the average value of the data grouped in the sixth iteration is taken. These new centroids are then recorded in Table 14.

Table 14. New Centroids in the Seventh Iteration

NEW CENTROIDS	T	F	H
C1	440	440,184397163121	387,354609929078
C2	378,063218390805	371,965517241379	376,477011494253
C3	287,628865979381	290,731958762887	363,721649484536

After computing the new centroids, the distances from every data point to these centroids are re-evaluated utilizing the Euclidean Distance method. This step repeats the methodology outlined in the previous iteration. This iteration begins with computing the Euclidean Distance between each data point and the newest centroids of each cluster. This procedure is similar to the approach taken in the previous iteration.

For each data point, its distance to the first cluster center (C1) is computed as follows:

C1 (1) = $\sqrt{((390-440)^2+(390-440{,}184397163121)^2+(390-387{,}354609929078)^2)}$

C1 (2) = $\sqrt{((390-440)^2+(390-440{,}184397163121)^2+(388-387{,}354609929078)^2)}$

C1 (3) = $\sqrt{((388-440)^2+(388-440{,}184397163121)^2+(388-387{,}354609929078)^2)}$

And so on until C1(412).

Then, the distance from each data to the second cluster center (C2) is computed using a similar formula:

C2 (1) = $\sqrt{((390-378{,}063218390805)^2+(390-371{,}965517241379)^2+(390-376{,}477011494253)^2)}$

C2 (2) = $\sqrt{((390-378{,}063218390805)^2+(390-371{,}965517241379)^2+(388-376{,}477011494253)^2)}$

C2 (3) = $\sqrt{((388-378{,}063218390805)^2+(388-371{,}965517241379)^2+(388-376{,}477011494253)^2)}$

And so on until C2(412).

Finally, the distance from each data to the third cluster center (C3) is calculated as follows:

C3 (1) = $\sqrt{((390-287{,}628865979381)^2+(390-290{,}731958762887)^2+(390-363{,}721649484536)^2)}$

C3 (2) = $\sqrt{((390-287{,}628865979381)^2+(390-290{,}731958762887)^2+(388-363{,}721649484536)^2)}$

C3 (3) = $\sqrt{((388-287{,}628865979381)^2+(388-290{,}731958762887)^2+(388-363{,}721649484536)^2)}$

And so on until C3(412).

From these calculations in the seventh iteration, the distances from each data point to each cluster are obtained as listed in Table 15.

Table 15. Calculation Results in the Seventh Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	70,89056219	25,50687244	144,9984303
2	Afif Dwi Ainul Yaqin	70,84412641	24,50527673	144,6493394
3	Ahmad Adi Saputra	73,67243606	22,10483047	141,8621679
…	…………………	…	…	…
410	Ahmad Zaidan Salim	187,5632569	96,75138462	36,92648665
411	Aufil Ghulam	133,9935302	46,14473073	83,92687326
412	Danil Faizin	150,2946639	57,90174467	64,84842321

Furthermore, since the results of the seventh iteration differ from the sixth iteration, the process continues with the eighth iteration. The positions of the new centroids in the eighth iteration are calculated by taking the average of the data included in each group or centroid as in the previous iteration.

In the eighth iteration, to find the new centroids, the average value of the data grouped in the seventh iteration is taken. These new centroids are then recorded in Table 16.

Table 16. New Centroids in the Eighth Iteration

NEW CENTROIDS	T	F	H
C1	439,395833333333	439,4375	387,347222222222
C2	376,976878612717	370,994219653179	376,132947976879
C3	286,663157894737	289,768421052632	363,747368421053

After computing the new centroids, the distances from every data point to these centroids are re-evaluated utilizing the Euclidean Distance method. This step repeats the methodology outlined in the previous iteration. This iteration begins with computing the Euclidean Distance between each data point and the newest centroids of each cluster. This procedure is similar to the approach taken in the previous iteration.

For each data point, its distance to the first cluster center (C1) is computed as follows:

C1 (1) = $\sqrt{((390-439{,}395833333333)^2+(390-439{,}4375)^2+(390-387{,}347222222222)^2)}$

C1 (2) = $\sqrt{((390-439{,}395833333333)^2+(390-439{,}4375)^2+(388-387{,}347222222222)^2)}$

C1 (3) = $\sqrt{((388-439{,}395833333333)^2+(388-439{,}4375)^2+(388-387{,}347222222222)^2)}$

And so on until C1(412).

Then, the distance from each data to the second cluster center (C2) is computed using a similar formula:

C2 (1) = $\sqrt{((390-376{,}976878612717)^2 + (390-370{,}994219653179)^2 + (390-376{,}132947976879)^2)}$

C2 (2) = $\sqrt{((390-376{,}976878612717)^2 + (390-370{,}994219653179)^2 + (388-376{,}132947976879)^2)}$

C2 (3) = $\sqrt{((388-376{,}976878612717)^2 + (388-370{,}994219653179)^2 + (388-376{,}132947976879)^2)}$

And so on until C2(412).

Finally, the distance from each data to the third cluster center (C3) is calculated as follows:

C3 (1) = $\sqrt{((390-286{,}663157894737)^2 + (390-289{,}768421052632)^2 + (390-363{,}747368421053)^2)}$

C3 (2) = $\sqrt{((390-286{,}663157894737)^2 + (390-289{,}768421052632)^2 + (388-363{,}747368421053)^2)}$

C3 (3) = $\sqrt{((388-286{,}663157894737)^2 + (388-289{,}768421052632)^2 + (388-363{,}747368421053)^2)}$

And so on until C3(412).

From these calculations in the eighth iteration, the distances from each data point to each cluster are obtained as listed in Table 17.

Table 17. Calculation Results in the Eighth Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	69,93605642	26,89082574	146,3354811
2	Afif Dwi Ainul Yaqin	69,88877503	25,91617836	145,9899397
3	Ahmad Adi Saputra	72,71708334	23,48473321	143,2019162
…	…………………	…	…	…
410	Ahmad Zaidan Salim	186,6101381	95,35859174	37,87818937
411	Aufil Ghulam	133,0721788	44,79814432	85,26719224
412	Danil Faizin	149,3488526	56,405538	66,21114049

Furthermore, since the results of the eighth iteration differ from the seventh iteration, the process continues with the ninth iteration. The positions of the new centroids in the ninth iteration are calculated by taking the average of the data included in each group or centroid as in the previous iteration.

In the ninth iteration, to find the new centroids, the average value of the data grouped in the eighth iteration is taken. These new centroids are then recorded in Table 18.

Table 18. New Centroids in the Ninth Iteration

NEW CENTROIDS	T	F	H
C1	439,193103448276	439,2	387,296551724138
C2	376,78488372093	370,796511627907	376,110465116279
C3	286,663157894737	289,768421052632	363,747368421053

After computing the new centroids, the distances from every data point to these centroids are re-evaluated utilizing the Euclidean Distance method. This step repeats the methodology outlined in the previous iteration. This iteration begins with computing the Euclidean Distance between each data point and the newest centroids of each cluster. This procedure is similar to the approach taken in the previous iteration.

For each data point, its distance to the first cluster center (C1) is computed as follows:

C1 (1) = $\sqrt{((390-439{,}193103448276)^2 + (390-439{,}2)^2 + (390-387{,}296551724138)^2)}$

C1 (2) = $\sqrt{((390-439{,}193103448276)^2 + (390-439{,}2)^2 + (388-387{,}296551724138)^2)}$

C1 (3) = $\sqrt{((388-439{,}193103448276)^2 + (388-439{,}2)^2 + (388-387{,}296551724138)^2)}$

And so on until C1(412).

Then, the distance from each data to the second cluster center (C2) is computed using a similar formula:

C2 (1) = $\sqrt{((390-376{,}78488372093)^2 + (390-370{,}796511627907)^2 + (390-376{,}110465116279)^2)}$

C2 (2) = $\sqrt{((390-376{,}78488372093)^2 + (390-370{,}796511627907)^2 + (388-376{,}110465116279)^2)}$

C2 (3) = $\sqrt{((388 - 376{,}78488372093)^2 + (388 - 370{,}796511627907)^2 + (388 - 376{,}110465116279)^2)}$

And so on until C2(412).

Finally, the distance from each data to the third cluster center (C3) is calculated as follows:

C3 (1) = $\sqrt{((390 - 286{,}663157894737)^2 + (390 - 289{,}768421052632)^2 + (390 - 363{,}747368421053)^2)}$

C3 (2) = $\sqrt{((390 - 286{,}663157894737)^2 + (390 - 289{,}768421052632)^2 + (388 - 363{,}747368421053)^2)}$

C3 (3) = $\sqrt{((388 - 286{,}663157894737)^2 + (388 - 289{,}768421052632)^2 + (388 - 363{,}747368421053)^2)}$

And so on until C3(412).

From these calculations in the ninth iteration, the distances from each data point to each cluster are obtained as listed in Table 19.

Table 19. Calculation Results in the Ninth Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	69,62693487	27,13544625	146,3354811
2	Afif Dwi Ainul Yaqin	69,57798694	26,16819259	145,9899397
3	Ahmad Adi Saputra	72,40627514	23,72972577	143,2019162
…	…………………	…	…	…
410	Ahmad Zaidan Salim	186,3001279	95,09050708	37,87818937
411	Aufil Ghulam	132,7628676	44,56437597	85,26719224
412	Danil Faizin	149,0334829	56,13206915	66,21114049

Furthermore, since the results of the ninth iteration differ from the eighth iteration, the process continues with the tenth iteration. The positions of the new centroids in the tenth iteration are calculated by taking the average of the data included in each group or centroid as in the previous iteration.

In the tenth iteration, to find the new centroids, the average value of the data grouped in the ninth iteration is taken. These new centroids are then recorded in Table 20.

Table 20. New Centroids in the Tenth Iteration

NEW CENTROIDS	T	F	H
C1	439,082191780822	438,86301369863	387,315068493151
C2	376,514619883041	370,684210526316	376,029239766082
C3	286,663157894737	289,768421052632	363,747368421053

After computing the new centroids, the distances from every data point to these centroids are re-evaluated utilizing the Euclidean Distance method. This step repeats the methodology outlined in the previous iteration. This iteration begins with computing the Euclidean Distance between each data point and the newest centroids of each cluster. This procedure is similar to the approach taken in the previous iteration.

For each data point, its distance to the first cluster center (C1) is computed as follows:

C1 (1) = $\sqrt{((390 - 439{,}193103448276)^2 + (390 - 439{,}2)^2 + (390 - 387{,}296551724138)^2)}$

C1 (2) = $\sqrt{((390 - 439{,}193103448276)^2 + (390 - 439{,}2)^2 + (388 - 387{,}296551724138)^2)}$

C1 (3) = $\sqrt{((388 - 439{,}193103448276)^2 + (388 - 439{,}2)^2 + (388 - 387{,}296551724138)^2)}$

And so on until C1(412).

Then, the distance from each data to the second cluster center (C2) is computed using a similar formula:

C2 (1) = $\sqrt{((390 - 376{,}78488372093)^2 + (390 - 370{,}796511627907)^2 + (390 - 376{,}110465116279)^2)}$

C2 (2) = $\sqrt{((390 - 376{,}78488372093)^2 + (390 - 370{,}796511627907)^2 + (388 - 376{,}110465116279)^2)}$

C2 (3) = $\sqrt{((388 - 376{,}78488372093)^2 + (388 - 370{,}796511627907)^2 + (388 - 376{,}110465116279)^2)}$

And so on until C2(412).

Finally, the distance from each data to the third cluster center (C3) is calculated as follows:

$$C3\,(1) = \sqrt{((390 - 286{,}663157894737)^2 + (390 - 289{,}768421052632)^2 + (390 - 363{,}747368421053)^2)}$$

$$C3\,(2) = \sqrt{((390 - 286{,}663157894737)^2 + (390 - 289{,}768421052632)^2 + (388 - 363{,}747368421053)^2)}$$

$$C3\,(3) = \sqrt{((388 - 286{,}663157894737)^2 + (388 - 289{,}768421052632)^2 + (388 - 363{,}747368421053)^2)}$$

And so on until C3(412).

From these calculations in the tenth iteration, the distances from each data point to each cluster are obtained as listed in Table 21.

Table 21. Calculation Results in the Tenth Iteration

NO	NAME	C1	C2	C3
1	Achmad Sulton Amiruddin K	69,30991643	27,38863526	146,3354811
2	Afif Dwi Ainul Yaqin	69,26127915	26,4245019	145,9899397
3	Ahmad Adi Saputra	72,08956659	23,98019229	143,2019162
…	…………………	…	…	…
410	Ahmad Zaidan Salim	185,9863417	94,82666506	37,87818937
411	Aufil Ghulam	132,4780731	44,26353068	85,26719224
412	Danil Faizin	148,7232011	55,84326884	66,21114049

Furthermore, since the results of the tenth iteration are the same as the ninth iteration, it is not necessary to perform calculations for the eleventh iteration, or it is sufficient to stop at the tenth iteration.

c) Nearest Neighbor Search

In this stage, the search for the minimum distance is performed on the values of each cluster result. For example, in the first data, if the smallest value among the 3 clusters is C1, then the first data belongs to the C1 group. If the smallest value among the 3 clusters is C2, then the first data belongs to the C2 group. If the smallest value among the 3 clusters is C3, then the first data belongs to the C3 group. Please refer to the following table:

Table 22. Results of Nearest Neighbor Search and Cluster Group in the First Iteration

NAME	C1	C2	C3	JARAK TERDEKAT	CLUSTER
Achmad Sulton Amiruddin K	0	2	3,4641 01615	0	C1
Afif Dwi Ainul Yaqin	2	0	2,8284 27125	0	C2
Ahmad Adi Saputra	3,4641 01615	2,8284 27125	0	0	C3
…………..	…	…	…	………	………
Ahmad Zaidan Salim	116,72 61753	116,74 33082	113,91 66362	113,9166362	C3
Aufil Ghulam	67,453 68782	66,738 29485	64,202 80368	64,20280368	C3
Danil Faizin	81,700 67319	81,111 03501	78,402 80607	78,40280607	C3

In the first iteration, the number of grouped data in each cluster was as follows: C1 had 126 grouped data, C2 had 67 grouped data, and C3 had 219 grouped data. In the second iteration, the numbers changed to: C1 with 117 grouped data, C2 with 117 grouped data, and C3 with 178 grouped. Please refer to the following table:

Table 23. Results of Nearest Distance Search and Cluster Groups in the Second Iteration

NAME	C1	C2	C3	JARAK TERDEKAT	CLUSTER
Achmad Sulton Amiruddin K	59,565 73603	44,992 37826	83,514 70424	44,99237826	C2
Afif Dwi Ainul Yaqin	59,601 16719	44,261 94834	83,086 70103	44,26194834	C2
Ahmad Adi Saputra	62,427 97059	46,923 41248	80,319 69726	46,92341248	C2
…………..	…	…	…	………	………
Ahmad Zaidan Salim	176,26 92676	159,19 82733	40,398 44787	40,39844787	C3
Aufil Ghulam	123,50 52756	103,76 05628	29,530 61838	29,53061838	C3
Danil Faizin	139,61 39146	119,53 3549	7,3731 5114	7,37315114	C3

Since the number of grouped data in each cluster in the second iteration differed from the first iteration, a search for the nearest distance (minimum) was conducted for each cluster's results in the third iteration. The results were: C1 with 125 grouped data, C2 with 158 grouped data, and C3 with 129 grouped data. Please refer to the following table:

Table 24. Results of Nearest Distance Search and Cluster Groups in the Third Iteration

NAMA	C1	C2	C3	JARAK TERDEKAT	CLUSTER
Achmad Sulton Amiruddin K	77,258 42855	15,417 37016	99,642 59227	15,41737016	C2
Afif Dwi Ainul Yaqin	77,249 57774	13,688 33505	99,264 00568	13,68833505	C2
Ahmad Adi Saputra	80,077 80295	15,317 25758	96,483 42183	15,31725758	C2
…………..	…	…	…	………	………

Ahmad Zaidan Salim	193,94 2281	124,48 95681	28,014 74556	28,01474556	C3
Aufil Ghulam	140,40 4397	70,870 07195	42,629 77691	42,62977691	C3
Danil Faizin	156,82 18593	85,515 79803	20,599 03762	20,59903762	C3

This iterative process continued through the fourth, fifth, sixth iterations, and so on, where the number of grouped data in each cluster could change in each iteration. This demonstrates an iterative process in finding the optimal clusters.

Finally, in the tenth iteration, the number of grouped data in each cluster remained the same as in the ninth iteration, concluding the iterative process with the following results: C1 with 146 grouped data, C2 with 171 grouped data, and C3 with 95 grouped data. Please refer to the following table:

Table 25. Results of Nearest Distance Search and Cluster Groups in the Tenth Iteration

NAME	C1	C2	C3	JARAK TERDEKAT	CLUSTER
Achmad Sulton Amiruddin K	69,309 91643	27,388 63526	146,33 54811	27,38863526	C2
Afif Dwi Ainul Yaqin	69,261 27915	26,424 5019	145,98 99397	26,4245019	C2
Ahmad Adi Saputra	72,089 56659	23,980 19229	143,20 19162	23,98019229	C2
....	...	...	...	....	....
Ahmad Zaidan Salim	185,98 63417	94,826 66506	37,878 18937	37,87818937	C3
Aufil Ghulam	132,47 80731	44,263 53068	85,267 19224	44,26353068	C2
Danil Faizin	148,72 32011	55,843 26884	66,211 14049	55,84326884	C2

3.5 Evaluation of Clustering Results

The clustering process for the Al-Qur'an learning groups using the k-means algorithm resulted in an appropriate number of grouped data in each cluster after the tenth iteration. From the calculation of 412 students' data, the following groups were formed:

a) The first cluster (C1), representing the group with a good rating, consists of 146 students.
b) The second cluster (C2), representing the group with a fairly good rating, consists of 171 students.
c) The third cluster (C3), representing the group with a poor rating, consists of 95 students.

Table 26. First Cluster Group (Good)

NO	NAME	VALUE			CLUSTER
		T	F	H	
1	Brillian Saputera Pratama M.	426	422	388	C1
2	Muhammad Syarif Hasan Ubaidillah	432	424	390	C1
3	Rafa El Asyraf Tegar Maharaja	435	435	390	C1
4	Zulfi Zaimul Mazza	420	410	335	C1
5	Achmad Firdaus Zaini	450	455	390	C1
6	Ahmad Farhan Billah	415	420	390	C1
7	Saiful Akbar Kumala	430	427	390	C1
8	Ahmad Kevin Randika Gautama	435	430	387	C1
9	Ahmad Sofiyan Hidayatullah	448	445	390	C1
10	Saiful Anwar Fatahillah	440	430	390	C1
...	....	...	...	...	....
145	Mochammad Rizal Rofiqi	425	415	385	C1
146	Ahmad Imong Budiono	430	408	390	C1

Table 27. Second Cluster Group (Fairly Good)

NO	NAME	VALUE			CLUSTER
		T	F	H	
1	Achmad Sulton Amiruddin K	390	390	388	C2
2	Afif Dwi Ainul Yaqin	390	390	390	C2
3	Ahmad Adi Saputra	388	388	388	C2
4	Ahmad Daniel Mateen Wafa	390	390	388	C2
5	Ahmad Misbahus Sururi	396	396	388	C2
6	Ahmad Yazid Zidan Altintop	390	390	390	C2
7	Alifian Nawal Haq	390	390	390	C2
8	Gilang Kurniawan Ramadhan	388	388	386	C2
9	Habil Michael Jibril	390	392	388	C2
10	M Farih Romadhoni	390	390	390	C2
...	....	...	...	...	....
170	Aufil Ghulam	335	360	365	C2
171	Danil Faizin	335	335	365	C2

Table 32. Third Cluster Group (Poor)

NO	NAME	VALUE			CLUSTER
		T	F	H	
1	Raditya Farrel Bayu Prasasta	315	319	284	C3
2	Vicky Fiera Alfiansyah	298	320	370	C3
3	Rif'at Syawqi	310	315	390	C3
4	Riski Septia Ramadani	300	320	380	C3
5	Robithurrahman	320	265	385	C3
6	Ahmad Farel Alfarisi Oktavian	299	299	380	C3
7	Ahmad Rakan Maulidi Sahirol Layali	299	299	390	C3
8	M. Kayyas Dhia El Haq	325	325	378	C3
9	Mochammad Iqbal Matlubi	325	325	390	C3
10	Mohammad Rizal	325	325	390	C3
...	....	...	...	...	....
94	Vikri Aska Tufail	350	270	370	C3
95	Ahmad Zaidan Salim	305	310	390	C3

4. DISCUSSION

This study successfully developed a web-based application system using the K-Means algorithm to group students in the Qur'an learning activities at Nurul Jadid Islamic Boarding School, Paiton Probolinggo. The results show that the use of the K-Means algorithm is effective in clustering students, thus improving efficiency and accuracy in the learning process. In this section, the authors will further discuss the findings of this study, compare them with similar research, and provide interpretation and implications of the results obtained.

One of the main findings of this study is the effectiveness of the K-Means algorithm in grouping students based on their Qur'anic reading abilities, including fluency in Tajwid and memorization scores. This algorithm successfully clustered students into three skill categories: "Good," "Fair," and "Poor." This clustering provides an objective and accurate guide in determining the appropriate learning groups for each student. These results are consistent with the research conducted by Achmad Dimyati (2023), which also found that the K-Means algorithm was effective in evaluating academic scores of students at TPQ Darussalamah. The table below compares the results of this study with several similar studies:

Table 33. Comparison of Research Results.

STUDY	CONTEXT	METHOD	MAIN FINDINGS
This study	Grouping students based on Qur'an abilities at Nurul Jadid Islamic Boarding School	K-Means	Effective in clustering students into three skill categories: "Good," "Fair," and "Poor."
Achmad Dimyati (2023)	Evaluating academic scores of students at TPQ Darussalamah	K-Means	The K-Means algorithm is effective in clustering students based on academic scores. These findings support our study's conclusion that K-Means can be used for educational classification.
Ai Rohmah, Falentino Sembiring, and Adhitia Erfina (2021)	Classifying levels of learning obstacles in distance learning during the COVID-19 pandemic	K-Means	The K-Means algorithm helps identify levels of learning obstacles faced by students. This study highlights the flexibility of K-Means in various educational contexts, similar to our research.
Dian Permata Sari (2021)	Determining the spread levels of the COVID-19 pandemic in West Sumatra	K-Means	The K-Means algorithm is effective in clustering regions based on positive cases and virus spread. This study shows that K-Means can be used in epidemiological data analysis, which is different from educational contexts but demonstrates the algorithm's flexibility.

The implementation of technology and data analysis methods in this study aims to address the challenges faced by Nurul Jadid Islamic Boarding School. By leveraging technology, the clustering process becomes faster, more accurate, and more objective, thus increasing student motivation through personalized and tailored learning experiences. This study shows that with proper clustering, students can receive training that matches their abilities, which in turn improves learning outcomes.

The development of this web-based system also facilitates access and use by administrators and participants. This system not only benefits Nurul Jadid Islamic Boarding School but also has the potential to be applied in other Islamic educational institutions. With this system, the Qur'an learning process can be conducted more efficiently, allowing teachers to focus more on teaching rather than administrative tasks.

Although the results of this study demonstrate success in clustering students, several limitations should be noted. First, this study only used data from one institution, so the results may not be generalizable to other institutions without further adjustments. Second, this system relies on the quality of the input data; therefore, it is important to ensure that the data used is accurate and complete.

For future research, it is recommended to test this system in various Islamic educational institutions with different characteristics to validate its reliability. Additionally, integrating more advanced automatic assessment systems can further enhance the accuracy of clustering.

5. CONCLUSION

This study demonstrates that the use of the K-Means algorithm is highly effective in grouping students based on their ability to read the Al-Qur'an. The research analyzed 412 students' data, evaluated in terms of tajwid (pronunciation), fashohah (fluency), and memorization. This data was then processed and clustered using the K-Means algorithm.

The results show that students can be categorized into three levels of proficiency: "Good", "Fairly Good", and "Poor". These categories provide clear guidance for determining the appropriate training group for each student, ultimately enhancing the efficiency and accuracy of the training process in accordance with the established curriculum.

This clustering method enables more precise training by grouping students based on similarities in their Al-Qur'an reading abilities. The use of web-based technology also facilitates easy access and utilization of the system by administrators, making the process faster, more accurate, and objective.

Overall, this study not only addresses the challenges faced by Pondok Pesantren Nurul Jadid in organizing Al-Qur'an training groups but also contributes to the development of a more responsive and efficient training system for similar institutions in the future.

REFERENCES

[1] M. Syafiih and N. Hatima Indah Arifin, "Sistem Informasi Monitoring Target Capaian Pembinaan Al-Qur'an Di Wilayah Pondok Pesantren Nurul Jadid Berbasis Web," *Jar's*, vol. 1, no. 2, p. 23, 2023, [Online]. Available: https://www.ejournalwiraraja.com/index.php/JARS

[2] 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.

[3] S. Santi and H. Februariyanti, "Implementation of Clustering on Tweet Uploading Side Effects of Covid-19 Post Vaccination Using K-Means Algorithm," *J. Tek. Inform.*, vol. 4, no. 4, pp. 779–786, 2023, doi: 10.52436/1.jutif.2023.4.4.704.

[4] S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.

[5] D. Adillah, N. Manurung, and ..., "Implementation of K-means Clustering Analysis to Determine Barriers to Online Learning Case Study: Swasta Yapendak Tinjowan Junior High School," *J. Tek. Inform. …*, vol. 3, no. 3, pp. 519–525, 2022, [Online]. Available: http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/189%0Ahttp://jutif.if.unsoed.ac.id/index.php/jurnal/article/download/189/100

[6] I. Soliani and S. Juanita, "Grouping the Prevalence of Disease Cases By Age in Bandung City Hospitals Using K-Means," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1647–1654, 2022, doi: 10.20884/1.jutif.2022.3.6.430.

[7] M. Syafiih *et al.*, "Pembuatan Website Biro Pendidikan Nurul Jadid dengan Sistem Framework sebagai Media Informasi Pendidikan di Pesantren," *GUYUB J. Community Engagem.*, vol. 2, no. 1, pp. 157–169, 2021, doi: 10.33650/guyub.v2i1.2120.

[8] N. Mirantika, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat," *Nuansa Inform.*, vol. 15, no. 2, pp. 92–98, 2021, doi: 10.25134/nuansa.v15i2.4321.

[9] Esmi Nur Fitri *et al.*, "Performance of K-Means Clustering and Knn Classifier in Fish Feed Seller Determination Models," *J. Tek. Inform.*, vol. 4, no. 3, pp. 485–491, 2023, doi: 10.52436/1.jutif.2023.4.3.725.

[10] Z. Subecz, "Web-development with Laravel framework," *Gradus*, vol. 8, no. 1, pp. 211–218, 2021, doi: 10.47833/2021.1.csc.006.

[11] Z. D. Montero, "Customer Grouping for Customer Relationship Management Optimization with the K-Means Algorithm," *J. Comput. Scine Inf. Technol.*, pp. 98–105, 2022, doi: 10.35134/jcsitech.v8i4.46.

[12] S. Likmi and J. Barat, "APPLICATION OF DATA MINING TO MEASURE THE EFFECTIVENESS OF THE ISLAMIC BOARDING SCHOOL ' S INDEPENDENT CURRICULUM BASED ON LEARNING ACHIEVEMENT USING THE CLUSTERING METHOD," vol. 3, no. 5, pp. 514–519, 2024, doi: 10.58344/jws.v3i5.595.

[13] A. Rohmah, F. Sembiring, and A. Erfina, "Implementasi Algoritma K-Means Clustering Analysis untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung)," *Semin. Nas. Sist. Inf. dan Manaj. Inform.*, pp. 290–298, 2021, [Online]. Available: https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/32

[14] D. P. Sari, "Implementasi Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Sumatera Barat," *Comput. Based Inf. Syst. J.*, vol. 9, no. 1, pp. 50–56, 2021, doi: 10.33884/cbis.v9i1.3646.

[15] A. Dimyati, "E-ISSN : XX-XX-XX ejurnal . itsnupekalongan . ac . id / sandtree E-ISSN : XX-XX-XX," vol. 1, no. 1, pp. 39–54, 2023.

[16] R. Anggara[1], S. Defit, B. Hendrik[3], and F. I. Komputer, "Implementasi K-Means Clustering Dalam Analisa Soal Ujian CBT Universitas Baiturrahmah," vol. 5, no. 2, pp. 577–586, 2024.

[17] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://doi.org/10.58602/jima-ilkom.v3i1.26

[18] M. R. Palevi and Z. Indra, "Implementasi Algoritma K-Means Clustering Dengan Pendekatan Active Learning Pada Siswa SMA Untuk Menentukan Jurusan Ke Perguruan Tinggi," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 23, no. 1, p. 26, 2024, doi: 10.53513/jis.v23i1.9553.

[19] R. Novita, A. N. Khomarudin, R. Aulia, J. Jamaluddin, A. Yuditihwa, and A. Ayuri, “Penerapan Algoritma K-Means dan Analisisnya untuk Menentukan Kebijakan Strategis Penyelesaian Studi Mahasiswa,” *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 22, no. 2, p. 401, 2023, doi: 10.53513/jis.v22i2.8461.

[20] E. Ramadanti and M. Muslih, “Penerapan Data Mining Algoritma K-Means Clustering Pada Populasi Ayam Petelur Di Indonesia,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 1, pp. 1–7, 2022, doi: 10.36341/rabit.v7i1.2155.

[21] K. P. Sinaga and M. S. Yang, “Unsupervised K-means clustering algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[22] M. M. Hidayat, “Data Mining Data mining,” *Min. Massive Datasets*, vol. 2, no. January 2013, pp. 5–20, 2015, [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part

[23] B. Juli, “Jurnal Kajian dan Terapan Matematika,” vol. 8, pp. 114–128, 2023.

[24] Z. Nabila, A. R. Isnain, Permata, and Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

[25] M. R. Muttaqin and M. Defriani, “Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 121–129, 2020, doi: 10.33096/ilkom.v12i2.542.121-129.

[26] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.