Published online on the page : https://ejournal.iainbukittinggi.ac.id/index.php/ijokid/

Knowbase : International Journal of Knowledge in Database

| ISSN (Print) 2798-0758 | ISSN (Online) 2797-7501 |

K-Means Clustering Algorithm To See The Correlation Of Tahfidz Activities With Student's Learning Outcomes

Agus Nur Khomarudin [1,*], *Asri Hidayad* [2]

[1]*Politeknik Pertanian Negeri Payakumbuh*
[2]*Universitas Putra Indonesia "YPTK" Padang*

Article Information

Article History:
Accepted by the Editor: June 30, 2022
Final Revision: June 22, 2022
Published Online: June 30, 2022

Key Word

Data Mining
K-Means Clustering
Learning Outcomes

Correspondence

E-mail: agusnurkhomarudin@gmail.com*

A B S T R A C T

Computer technology is currently used as a tool to support educational activities and assist teachers in quickly, precisely, and accurately processing student data. The Qur'an has been memorized by many students at MAN 1 Bukittinggi as a result of long-running Tahfidz activities, which have also helped them stand out in teaching and learning evaluation meetings held at the end of each semester. In conveying the activities which take place in the field, the Tahfidz teacher has not used real data and certainly will not describe the real state of Tahfidz activity and student memorization. The solution to this problem is applying the Data Mining technique for grouping data of Tahfidz activity and student learning outcomes, in order to discover the correlation between the two and whether there is any effect or not. The data processed in this research is Tahfidz activity data and student learning outcomes data for class XI (eleven) sourced from piles of data, the data used are Tahfidz memorization data, Tahfidz scores, and student grades at MAN 1 Bukittinggi. The data mining development methodology used in this research is Cross-Industry Standart Process for Data Mining (CRIPS-DM). The results of the research by applying the K-Means Clustering algorithm has produced two clusters. The first cluster shows the characteristic that Tahfidz activity doesn't have significant effect on student learning outcomes. The second cluster shows the characteristics that Tahfidz activity has a significant effect on student learning outcomes. The data from the clustering of each cluster is then analyzed and the results of the cluster analysis are used by the school as a consideration in evaluating Tahfidz activity and learning outcomes whether it has been effective or not at MAN 1 Bukittinggi.

1. Introduction

Computer technology is currently used as a tool to support educational activities and assist teachers in quickly, precisely, and accurately processing student data so that the objectives of the work can be accomplished effectively and efficiently by Triyansyah and Fitrianah [1]. There is an algorithm is used to process data, and the k-means clustering technique is particularly useful for grouping data [2]. By using the clustering technique, the k-means clustering algorithm can make it clearer the density and distance of the area in the object space and can find the overall distribution pattern and correlation between attributes. The clustering approach can also effectively be used to distinguish groups or classes of objects [3].

The Tahfidzul Qur'an activity or known as Tahfidz which is being held at MAN 1 Bukittinggi is currently running in a conducive and controlled condition. It is a mandatory activity for all students from ten to twelve grade with different levels of memorization. Tahfidz activity at MAN 1 Bukittinggi has been going on for quite a long time and has produced many students who memorize the Qur'an.

Tahfidz activity also receive attention in teaching and learning evaluation meetings at the end of each semester. Activities for the evaluation of the teaching and learning process, learning administration, and student activity meetings are usually held at the end of the semester. So far, to find out whether Tahfidz activities are running effectively and have an impact on student abilities, such as knowledge, attitudes, emotions, and others, interviews are carried out and ask for direct responses from the Tahfidz teacher. In conveying the activities that take place in the field, the Tahfidz teacher has not used real data to support it, such as grouping Tahfidz student data.

Data processing, such as grouping Tahfidz student data, can provide an analysis of the state of Tahfiz activities for Tahfidz teachers, such as providing an analysis of correlation between Tahfidz activity and students' cognitive abilities through their learning outcomes. The K-Means Clustering algorithm clustering method was used in this study to analyze Tahfiz activity and the score of student learning outcomes. So that the resulting cluster or group can be an analysis for Tahfiz teachers and become input for the school in making decisions.

Based on the problem, there needs a solution for it to see the correlation of Tahfidz activity with student learning outcomes. There's a solution that the researcher tried to apply in this research is grouping Tahfidz student data using Data Mining. According to Irfiani and Rani, Data Mining is a way to find an interesting pattern and knowledge in big data. Data mining uses mathematical, statistical, artificial intelligence and machine learning techniques to identify and extract information or useful knowledge taken from various large databases and the method used is the Clustering method [4]. According to Sulastri and Gufroni, Clustering method is a process of grouping data objects that are similar to each other into the same cluster and different from objects in other clusters [5]. Algorithm used in this research to see a correlation of Tahfidz activity with student learning outcomes is K-Means Clustering algorithm.

According to Rosmini, et al. *K-Means* is one of clustering method or non hirarki grouping. The data grouping technique is simple and fast. There are many approaches to create a cluster, among them is making rules that dictate membership in the same group based on similarities among its members. Another approach is to create a function that measures some property of the grouping as a function of some parameter of a clustering [6], and When using the k-means method, data of nominal data type, such as majors, memorizing surah deposit data, semester report cards, students' attitude score data, and students' skill score data, must first be initialized into numeric form [1]. And if data has been numeric form, it will be processed using the tools Rapid Miner.

Research on the application of the k-means clustering algorithm by Gibran, et al entitled Application of Mapping the Quality of Education in Indonesia Using the K-means Method, The output of the application/system made in this research is a map that groups the regions according to the quality of their education [7]. Research by Agus, et al entitled Grouping Data for Scholarship Recipients Using the K-means Clustering Algorithm, this research was conducted to determine which student groups will receive scholarships during a specific semester. The end result of the research is generate a group of potential scholarship recipients [8].

Furthermore, Asroni et al. conducted the following research on the new student admission statistics from the Faculty of Medicine and Nursing in each cluster. (1) Cluster 0 majoring in Nursing, as many as 3795 applicants out of 25000 new student candidates (15%). (2) Cluster 1 majoring in Medical Education, as many as 8296 applicants out of 25000 new student candidates (33%). (3) Cluster 2 majoring in Dentistry, as many as

7427 applicants out of 25000 new student candidates (30%). And (4) cluster 3 majoring in Medical Education, as many as 5482 applicants out of 25000 new student candidates (22%) [9].

2. Method

The stages of the study are visualized in Figure 1, and some of the frameworks that will be used in this research can be explained :

2.1. Describing the Problem

Problem description is the process of identifying a problem in a structured way to facilitate decision-making. This research aims to identify the correlation between Tahfidz activity and the learning outcomes.

2.2. Analyzing the Problem

As a first stage in problem analysis, the researcher examines existing issues in order to comprehend them. The correlation between tahfiz activities and students' learning outcomes was constructed manually, which leads in an inefficient data processing process. As data mining is used, clustering methods employing the K-means clustering algorithm helps in identifying the student groups for tahfiz activities and the groups of learning outcomes. Thus, the correlation between tahfiz activities and learning outcomes will be shown.

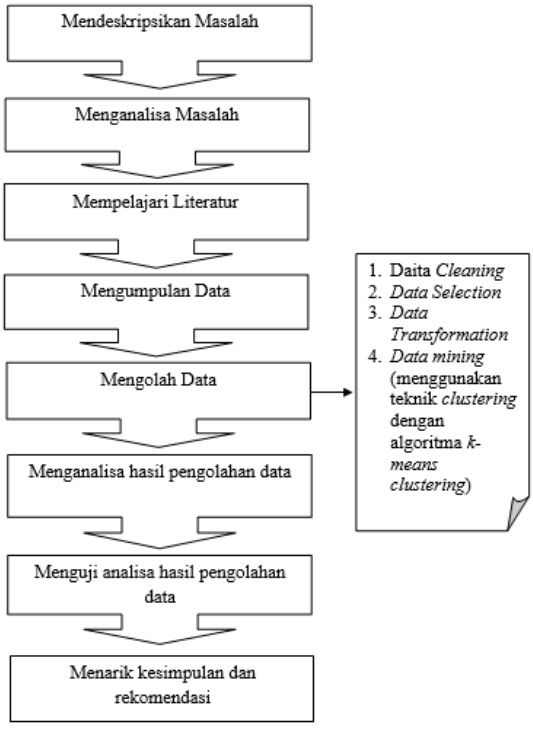

Figure 1. Research Framework

2.3. Study Literature

At this stage, some literatures on theories and concepts that support the problem solving of this research will be discussed. Literature that will be used in this research is in the form of reference books, national journals, and international journals that support in completing this research [10][11]. Those information in the literatures will be analyzed and selected to determine which literature will be used in this study.

2.4. Data Collection

In collecting the data, researcher using direct observations at the research site to make research problem clearly identified. Then, researcher also using interview as part of data collection. The interview will be conduct to the teachers that in the charge of Quran recitation task to obtain information about the correlation

between Tahfidz activities and learning outcomes as well as other required data. The sample of this data is the students of MAN 1 Bukittinggi enrolled in even semester of 2018/2019.

2.5. *Data Processing*

After the required data is collected, the next stage is to process the data with some steps :

2.5.1. *Data cleaning*

To remove the incosistent and inrelevan datas.

2.5.2. *Data selection*

To take the corresponding data for analysis. In this research, there are only some attribute data will be used; students' NISN, students' name, students' major, reciting data, students' semester score, students' attitude score, and students' skills score.

2.5.3. *Data transformation*

The data will be processed using k-means clustering algorithm, before processing the data, letter-based data such as students' major, reciting data, students' semester grade, students' attitude grade, and students' skills grade will be transformed into number.

2.5.4. *Data mining*

The data can be processed using k-means clustering algorithm after the data transformation [8].

2.5.5. *Analyzing the result*

In this stage, researcher analyzing the data in the form of cluster correlation of Tahfidz activity with students' score that have been processed with k-means clustering algorithm. This data anlysis consist of some cluster, such as analysis of total students per major, analysis of students' recitation data, analysis of students' grades. This analysis can be used to decide the prospective recipients of underprivileged scholarship in MAN 1 Bukittinggi.

2.5.6. *Testing the results of the data analysis*

The results that have been processed manually with clustering technique with k-means clustering algorithm will be analyzed and compared with the result that processed with *Rapid Miner* [12]. Data that will be analyzed is sample data and all of the data that includes iteration testing and cluster results.

2.5.7. *Conclusion and recommendation*

In this stage, researcher can draw a conclusion on the research problem based on the result of the data analysis. After that, researcher will give some recommendation for this research to help improve this research in the future.

3. Results and Discussion

In this step, the data which has been collected will be analyzed using *K-Means Clustering* method. This analysis aims to gain knowledge to see the correlation between tahfiz activities and student learning outcomes. Result of this prediction will be used by MAN 1 Bukittinggi to evaluate Thafidz score with student learning outcomes. Based on the previous discussion, a flow chart of the system to be designed can be formed as shown at Figure 2.

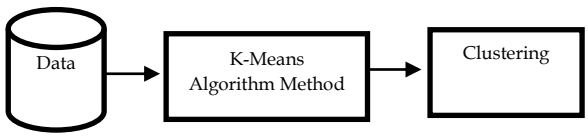

Figure 2. Designed System Architecture

The data used in this research is data that we got from MAN 1 Bukittinggi. The researcher asked the dean of students by hand in a research letter which obtained from school. So that they without any objections is willing to submit the data needed by the researchers, such as tahfidz score, total of memorizing, and learning outcomes for the 2020-2021 school year. The following data provided by MAN 1 Bukittinggi can be seen in Table 1.

Table 1. Preliminary Data

No	Name	Score	Average
1.	Zulfitra	98	88,83
2.	Muhammad Wahyudi Arafah	84	85,38
3.	Amirul Mukminin	84	85,48
4.	Olivia Levan's	62	87,23
5.	Annisa Yulia Dasman	95	87,18
6.	Dzurriyyatul'afifah	95	87,73
7.	Afifah Athohirah	79	83,83
8.	Amamatur Rusydiah Mufid	90	83,83
9.	Khairani	90	84,48
10.	Jerri Dwi Alkarizmi	64	80,78
11.	Yusril Yamanda	95	84,8
12.	Fikri Estiawan	76	85,88
13.	Nadia Yustika R.	95	84,73
14.	Alhusna Melfi	92	87,68
15.	Zahra Zakiah Raihanisa	79	84,72
16.	Andini	92	86,14
17.	Haura Fadhila Dzamri	79	85,25
18.	Farzana Quratul Aini	94	85,86
19.	Rangga	79	82,28
20.	Yaumil Khair Djumardi	92	89,31
21.	Feri Gunawan	84	83,39
22.	Fauziah Az Zahro	95	87,89
23.	Muhammad Farhab Shiddiq	64	85,53
24	Wilza Lolisya Shandori	79	84,44
25	Bogy Adventage	84	85,22
26	Habibatul Khairiah	95	88,81
27	Miftahus Sa'adah	84	90,28
28	Annisa Fadhillah	98	84,81
29	Nahda Navilah Daulai	59	87,28
30	Nurmisbah Suduri	95	87,69
31	Mufidah Asri Azizah	75	85,5
32	Wardah Hafifah	64	82,81
33	Megita Irza Mulyani	96	86,31
34	Rizki Ananta	84	85,06
35	Yolanda Agustina	84	82,47
36	Yudi Alfikri	64	80,08
37	Abdurrazaq	92	81,61
38	Najla Amatullah	64	84,53
39	M. Imam Faisal	84	81,11
40	Fadhlan Habib	79	82,67
41	M. Alfattah	64	82,92

No	Name	Score	Average
42	Fadila Khairunnisa	76	83,19
43	Dwi Fitri R.	95	84,11
44	Teguh Sabrian T.	84	79,83
45	Muhammad Akthar	84	80,86

The sample data that has been obtained will then be processed using the K-Means Clustering method in order to get the results in the form of grouping the expected data previously. The algorithms contained in K-Means are [3]:

3.1.1. Determining the total of cluster

In this study, the researchers clustered the data into 2 clusters, namely C0 and C1. C0 for the unaffected group, C1 for the influential group.

3.1.2. Determining the center point of the initial centeroid cluster

In determining the score initial centroid, it can be taken randomly from the sample data used in the study. This initial centroid is the center point of the first cluster. The centroid for C0 is taken from the 3rd sample data and C1 is taken from the 42nd sample data.

Table 2. Center Point / Initial Centroid

No	Name	Score	Average
3	Amirul Mukminin	84	85,48
42	Fadila Khairunnisa	76	83,19

3.1.3. Calculating distance of initial centroid

The formula used to calculate the distance between the centroid point and the point of each object is to use Euclidion Distance. The initial Centroid calculation manually with the formula:

$$D(m,n)=\sqrt{(x_{1m}-x_{1n})^2+(x_{2m}-x_{2n})^2+.....+(x_{km}-x_{kn})^2} \quad .....(1)$$

Calculation of the distance from the 1st data to the Cluster.

$$DC0=\sqrt{(98-84)^2+(88{,}83-85{,}48)^2}=14{,}395$$
$$DC1=\sqrt{(98-76)^2+(88{,}83-84{,}11)^2}=22{,}711$$

Calculation of the distance from the 2nd data to the Cluster.

$$DC0=\sqrt{(84-84)^2+(85{,}38-85{,}48)^2}=0{,}100$$
$$DC1=\sqrt{(84-76)^2+(85{,}38-84{,}11)^2}=8{,}294$$

Calculation of the distance from the 3rd data to the Cluster.

$$DC0=\sqrt{(84-84)^2+(85{,}48-85{,}48)^2}=0$$
$$DC1=\sqrt{(84-76)^2+(85{,}48-84{,}11)^2}=8{,}321$$

Calculation of the distance from the 4th data to the Cluster.

$$DC0 = \sqrt{(62-84)^2 + (87{,}23-85{,}48)^2} = 22{,}069$$
$$DC1 = \sqrt{(62-76)^2 + (87{,}23-84{,}11)^2} = 14{,}571$$

Calculation of the distance from the 5th data to the Cluster.

$$DC0 = \sqrt{(98-84)^2 + (87{,}18-85{,}48)^2} = 11{,}131$$
$$DC1 = \sqrt{(95-76)^2 + (87{,}18-84{,}11)^2} = 19{,}414$$

After calculating the dataset from the 1,2,3,4,5...45 data distance with the Centroid data obtained, then the allocation or placing of the Cluster members is based on the smallest value. The allocation of cluster members is grouped by giving the code "C0" if the number calculated from the closest distance is in the first group, the code "C1" is given if the number is in the second group. The following is a form of grouping cluster members based on the closest distance as outlined in Table 3.

Table 3. Grouping of Cluster Members Based on Nearest Distance

No	Name	DC0	DC1	Cluster
1.	Zulfitra	14,395	22,711	C0
2.	Muhammad Wahyudi Arafah	0,100	8,294	C0
3.	Amirul Mukminin	0	8,321	C0
4.	Olivia Levan's	22,069	14,571	C1
5.	Annisa Yulia Dasman	11,131	19,414	C0
6.	Dzurriyyatul'afifah	11,228	19,535	C0
7.	Afifah Athohirah	5,265	3,068	C1
8.	Amamatur Rusydiah Mufid	6,223	14,015	C0
9.	Khairani	6,083	14,059	C0
10.	Jerri Dwi Alkarizmi	20,545	12,240	C1
11.	Yusril Yamanda	11,021	19,068	C0
12.	Fikri Estiawan	8,010	2,690	C1
13.	Nadia Yustika R.	11,026	19,062	C0
14.	Alhusna Melfi	8,297	16,618	C0
15.	Zahra Zakiah Raihanisa	5,057	3,368	C1
16.	Andini	8,027	16,270	C0
17.	Haura Fadhila Dzamri	5,005	3,639	C1
18.	Farzana Quratul Aini	10,007	18,197	C0
19.	Rangga	5,936	3,135	C1
20.	Yaumil Khair Djumardi	8,870	17,131	C0
21.	Feri Gunawan	2,090	8,002	C0
22.	Fauziah Az Zahro	11,261	19,573	C0
23.	Muhammad Farhab Shiddiq	20,000	12,226	C1
24	Wilza Lolisya Shandori	5,107	3,250	C1
25	Bogy Adventage	0,260	8,254	C0
26	Habibatul Khairiah	11,493	19,814	C0
27	Miftahus Sa'adah	4,800	10,690	C0
28	Annisa Fadhillah	14,016	22,060	C0
29	Nahda Navilah Daulai	25,065	17,485	C1

30	Nurmisbah Suduri	11,220	19,526	C0
31	Mufidah Asri Azizah	9,000	2,517	C1
32	Wardah Hafifah	20,177	12,006	C1
33	Megita Irza Mulyani	12,029	20,242	C0
34	Rizki Ananta	0,420	8,216	C0
35	Yolanda Agustina	3,010	8,032	C0
36	Yudi Alfikri	20,716	12,396	C1
37	Abdurrazaq	8,887	16,078	C0
38	Najla Amatullah	20,023	12,075	C1
39	M. Imam Faisal	4,370	8,266	C0
40	Fadhlan Habib	5,736	3,045	C1
41	M. Alfattah	20,163	12,003	C1
42	Fadila Khairunnisa	8,321	0	C1
43	Dwi Fitri R.	11,085	19,022	C0
44	Teguh Sabrian T.	5,650	8,677	C0
45	Muhammad Akthar	4,620	8,332	C0

Based on the results of the first iteration, the closest dataset calculations have been grouped (1) DC0 = Calculation result with *Centroid* 0. And (2) DC1 = Calculation result with *Centroid* 1.

In the first iteration grouping in Cluster C0, there are 28 items in the no effect category can be seen in Table 4.

Table 4. Members of Cluster C0 with No Influence

No	Name	DC0	DC1	Cluster
1.	Zulfitra	14,395	22,711	C0
2.	Muhammad Wahyudi Arafah	0,100	8,294	C0
3.	Amirul Mukminin	0	8,321	C0
4.	Annisa Yulia Dasman	11,131	19,414	C0
5.	Dzurriyyatul'afifah	11,228	19,535	C0
6.	Amamatur Rusydiah Mufid	6,223	14,015	C0
7.	Khairani	6,083	14,059	C0
8.	Yusril Yamanda	11,021	19,068	C0
9.	Nadia Yustika R.	11,026	19,062	C0
10.	Alhusna Melfi	8,297	16,618	C0
11.	Andini	8,027	16,270	C0
12.	Farzana Quratul Aini	10,007	18,197	C0
13.	Yaumil Khair Djumardi	8,870	17,131	C0
14.	Feri Gunawan	2,090	8,002	C0
15.	Fauziah Az Zahro	11,261	19,573	C0
16.	Bogy Adventage	0,260	8,254	C0
17.	Habibatul Khairiah	11,493	19,814	C0
18.	Miftahus Sa'adah	4,800	10,690	C0
19.	Annisa Fadhillah	14,016	22,060	C0
20.	Nurmisbah Suduri	11,220	19,526	C0
21.	Megita Irza Mulyani	12,029	20,242	C0
22.	Rizki Ananta	0,420	8,216	C0
23.	Yolanda Agustina	3,010	8,032	C0
24	Abdurrazaq	8,887	16,078	C0
25	M. Imam Faisal	4,370	8,266	C0
26	Dwi Fitri R.	11,085	19,022	C0
27	Teguh Sabrian T.	5,650	8,677	C0
28	Muhammad Akthar	4,620	8,332	C0

Cluster C1 consist of 17 items as *Cluster* with influential categories, which will influence students to increase their Tahfidz activity and learning outcomes, from grouping Cluster 1 as shown in Table 5.

Table 5. C1 Cluster Members with Influential Categories

No	Name	DC0	DC1	Cluster
1	Olivia Levan's	22,069	14,571	C1
2	Afifah Athohirah	5,265	3,068	C1
3	Jerri Dwi Alkarizmi	20,545	12,240	C1
4	Fikri Estiawan	8,010	2,690	C1
5	Zahra Zakiah Raihanisa	5,057	3,368	C1
6	Haura Fadhila Dzamri	5,005	3,639	C1
7	Rangga	5,936	3,135	C1
8	Muhammad Farhab Shiddiq	20,000	12,226	C1
9	Wilza Lolisya Shandori	5,107	3,250	C1
10	Nahda Navilah Daulai	25,065	17,485	C1
11	Mufidah Asri Azizah	9,000	2,517	C1
12	Wardah Hafifah	20,177	12,006	C1
13	Yudi Alfikri	20,716	12,396	C1
14	Najla Amatullah	20,023	12,075	C1
15	Fadhlan Habib	5,736	3,045	C1
16	M. Alfattah	20,163	12,003	C1
17	Fadila Khairunnisa	8,321	0	C1

Each cluster member has been grouped, then an iteration will be carried out by generating a new Centroid score based on the sum of the Cluster members divided by the total Cluster members, with the following equation:

$$\mu = \frac{\sum_{i-1}^{n} x_i}{n}; i = 1,2,3....n \quad \text{..... (2)}$$

Table 6. Counting to get new Centroid score

Cluster	Score	Average
C0	98+84+84+95+95+90+90+95+95+92+9294+92+84+95+84+95+84+98+95+96+8484+92+84+95+84+84/28	88,83+85,38+85,48+87,18+87,73+83,83+84,48+84,8+84,73+87,68+86,14+85,86+89,31+83,39+87,89+85,22+88,81+90,28+84,81+87,69+86,31+85,06+82,47+81,61+81,11+84,11+79,83+80,86/28
C0	62+79+64+76+79+79+79+64+79+59+75+64+64+64+79+64+76/17	87,23+83,83+80,78+85,88+84,72+85,25+82,28+85,53+84,44+87,28+85,5+82,81+80,08+84,53+82,67+82,92+83,19/17

From the results of the summation calculations in Table 6, it is found that the new Cluster center can be seen in Table 7 as follows:

Table 7. New Centroid Second Iteration

Cluster	Score	Average
C0	90,5	85,389
C1	70,941	84,054

After getting the result with new Centroid, then do the calculation of the distance between new Centroid point and the point of each dataset object again which is the dataset in Table 1. To get the results of the second iteration, reuse the Euclidion Distance formula to calculate the distance of the Cluster with the new Centroid that has been obtained from the calculation of the first iteration.

After calculating the dataset from the 1,2,3,4,5…45 data distance with the new Centroid data obtained, then the allocation or placing of the Cluster members is based on the smallest value. The next iteration does the calculation again using the new Centroid point in the 2nd iteration. The results of the calculation of the 2nd iteration can be seen in Table 8.

Table 8. Calculation of Distance and Data Grouping of the 2nd Iteration

No	Name	DC0	DC1	Cluster
1.	Zulfitra	8,252	27,477	C0
2.	Muhammad Wahyudi Arafah	6,500	13,126	C0
3.	Amirul Mukminin	6,501	13,137	C0
4.	Olivia Levan's	28,559	9,488	C1
5.	Annisa Yulia Dasman	4,843	24,261	C0
6.	Dzurriyyatul'afifah	5,073	24,338	C0
7.	Afifah Athohirah	11,605	8,062	C1
8.	Amamatur Rusydiah Mufid	1,637	19,060	C0
9.	Khairani	1,037	19,064	C0
10.	Jerri Dwi Alkarizmi	26,898	7,674	C1
11.	Yusril Yamanda	4,538	24,071	C0
12.	Fikri Estiawan	14,508	5,378	C1
13.	Nadia Yustika R.	4,548	24,068	C0
14.	Alhusna Melfi	2,738	21,369	C0
15.	Zahra Zakiah Raihanisa	11,519	8,086	C1
16.	Andini	1,677	21,162	C0
17.	Haura Fadhila Dzamri	11,501	8,147	C1
18.	Farzana Quratul Aini	3,532	23,130	C0
19.	Rangga	11,913	8,252	C1
20.	Yaumil Khair Djumardi	4,198	21,705	C0
21.	Feri Gunawan	6,800	13,076	C0
22.	Fauziah Az Zahro	5,148	24,363	C0
23.	Muhammad Farhab Shiddiq	26,500	7,096	C1
24	Wilza Lolisya Shandori	11,539	8,068	C1
25	Bogy Adventage	6,502	13,111	C0
26	Habibatul Khairiah	5,653	24,525	C0
27	Miftahus Sa'adah	8,135	14,467	C0
28	Annisa Fadhillah	7,522	27,070	C0
29	Nahda Navilah Daulai	31,557	12,369	C1
30	Nurmisbah Suduri	5,054	24,332	C0
31	Mufidah Asri Azizah	15,500	4,309	C1
32	Wardah Hafifah	26,625	7,052	C1
33	Megita Irza Mulyani	5,577	25,160	C0
34	Rizki Ananta	6,508	13,098	C0
35	Yolanda Agustina	7,125	13,155	C0

36	Yudi Alfikri	27,027	7,998	C1
37	Abdurrazaq	4,066	21,200	C0
38	Najla Amatullah	26,514	6,957	C1
39	M. Imam Faisal	7,782	13,387	C0
40	Fadhlan Habib	11,817	8,177	C1
41	M. Alfattah	26,615	7,033	C1
42	Fadila Khairunnisa	14,666	5,132	C1
43	Dwi Fitri R.	4,678	24,059	C0
44	Teguh Sabrian T.	8,553	13,725	C0
45	Muhammad Akthar	7,922	13,444	C0

After the calculations in the second iteration are carried out and the same Cluster members are obtained as the first iteration as in Table 3. then the iteration process is stopped. The K-Means algorithm that has been used in the above calculation is an algorithm that partitions data into the same clusters and data that have dissimilarities are in other clusters. The analysis of the results of the grouping of 2 Cluster members obtained can be seen in Table 9.

Table 9. C0 Cluster Members with No Influence

No	Name	DC0	DC1	Cluster
1.	Zulfitra	8,252	27,477	C0
2.	Muhammad Wahyudi Arafah	6,500	13,126	C0
3.	Amirul Mukminin	6,501	13,137	C0
4.	Annisa Yulia Dasman	4,843	24,261	C0
5.	Dzurriyyatul'afifah	5,073	24,338	C0
6.	Amamatur Rusydiah Mufid	1,637	19,060	C0
7.	Khairani	1,037	19,064	C0
8.	Yusril Yamanda	4,538	24,071	C0
9.	Nadia Yustika R.	4,548	24,068	C0
10.	Alhusna Melfi	2,738	21,369	C0
11.	Andini	1,677	21,162	C0
12.	Farzana Quratul Aini	3,532	23,130	C0
13.	Yaumil Khair Djumardi	4,198	21,705	C0
14.	Feri Gunawan	6,800	13,076	C0
15.	Fauziah Az Zahro	5,148	24,363	C0
16.	Bogy Adventage	6,502	13,111	C0
17.	Habibatul Khairiah	5,653	24,525	C0
18.	Miftahus Sa'adah	8,135	14,467	C0
19.	Annisa Fadhillah	7,522	27,070	C0
20.	Nurmisbah Suduri	5,054	24,332	C0
21.	Megita Irza Mulyani	5,577	25,160	C0
22.	Rizki Ananta	6,508	13,098	C0
23.	Yolanda Agustina	7,125	13,155	C0
24	Abdurrazaq	4,066	21,200	C0
25	M. Imam Faisal	7,782	13,387	C0
26	Dwi Fitri R.	4,678	24,059	C0
27	Teguh Sabrian T.	8,553	13,725	C0
28	Muhammad Akthar	7,922	13,444	C0

C0 consists of 28 items as a cluster with no influence category which will be maintained later, because in the "C0" cluster, students whose Tahfidz activity will not affect their learning outcomes. Therefore, in the

"C0" cluster , it can be concluded that the correlation between Tahfidz activities and student learning outcomes does not affect, for more details, it can be seen in Table 10.

Table 10. C1 Cluster Members with Influential Categories

No	Name	DC0	DC1	Cluster
1.	Olivia Levan's	28,559	9,488	C1
2.	Afifah Athohirah	11,605	8,062	C1
3.	Jerri Dwi Alkarizmi	26,898	7,674	C1
4.	Fikri Estiawan	14,508	5,378	C1
5.	Zahra Zakiah Raihanisa	11,519	8,086	C1
6.	Haura Fadhila Dzamri	11,501	8,147	C1
7.	Rangga	11,913	8,252	C1
8.	Muhammad Farhab Shiddiq	26,500	7,096	C1
9.	Wilza Lolisya Shandori	11,539	8,068	C1
10.	Nahda Navilah Daulai	31,557	12,369	C1
11.	Mufidah Asri Azizah	15,500	4,309	C1
12.	Wardah Hafifah	26,625	7,052	C1
13.	Yudi Alfikri	27,027	7,998	C1
14.	Najla Amatullah	26,514	6,957	C1
15.	Fadhlan Habib	11,817	8,177	C1
16.	M. Alfattah	26,615	7,033	C1
17.	Fadila Khairunnisa	14,666	5,132	C1

C1 consists of 17 items as Clusters with influential categories which later in Cluster "C1" students whose Tahfidz scores are less or have not reached the target so that the number of memorization is further increased so that these students get scores that reach the desired target. More details can be seen in Table 10.

4. Conclusion

In this research, the results of calculations using the K-Means Clustering method produced two clusters with distinct characteristics. Cluster 1 displays the results of 28 student data points with no influence, while Cluster 2 displays the results of 17 student data points with influential on student learning outcomes. There are 45 students when all tests are added together. The results of this cluster are then taken into account by the Madrasah when evaluating Tahfidz activities the following year. Based on the comparison with real data of Tahfidz activities with learning outcomes, there are two students whose accuracy level is with the K-Means Clustering calculation, resulting in a 95.56 % comparison result.

References

[1] D. Triyansyah and D. Fitrianah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.

[2] J. O. Ong, "Implementasi Algotritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no, no. juni, pp. 10–20, 2013.

[3] Agus Nur Khomarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: https://ilmukomputer.org/category/datamining/.

[4] E. Irfiani and S. S. Rani, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," *J. Sist. dan Teknol. Inf.*, vol. 6, no. 4, p. 161, 2018, doi: 10.26418/justin.v6i4.29024.

[5] H. Sulastri and A. I. Gufroni, "The Application of Data Mining in the Grouping of Thalassemia Patients," *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017.

[6] R. Rosmini, A. Fadlil, and S. Sunardi, "Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," *It J. Res. Dev.*, vol. 3, no. 1, pp. 22–31, 2018, doi: 10.25299/itjrd.2018.vol3(1).1773.

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

[8] A. Nur Khomarudin, S. Zakir, R. Novita, Endrawati, M. Zahiri Bin Awang Mat, and E. Maiyana, "K-Mean Clustering Algorithm in Grouping Prospective Scholarship Recipients," *J. Phys. Conf. Ser.*, vol. 1779, no. 1, 2021, doi: 10.1088/1742-6596/1779/1/012007.

[9] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.

[10] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "Comprehensive Review of K-Means Clustering Algorithms," *Int. J. Adv. Sci. Res. Eng.*, vol. 07, no. 08, pp. 64–69, 2021, doi: 10.31695/ijasre.2021.34050.

[11] W. Purba, S. Tamba, and J. Saragih, "The effect of mining data k-means clustering toward students profile model drop out potential," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, 2018, doi: 10.1088/1742-6596/1007/1/012049.

[12] S. Hameetha Begum, "Data Mining Tools and Trends – An Overview," *Int. J. Emerg. Res. Manag. &Technology*, vol. ISSN, no. February, pp. 2278–9359, 2013.