ANALYSIS OF TEACHING AND LEARNING ACTIVITIES TO DETERMINE THE QUALITY OF STUDENT LEARNING USING THE K-MEANS CLUSTERING METHOD (CASE STUDY OF SMK MAARIF AL-MIZAN)

Muhammad Abdul Jabbar [1], Fikri Emsa Silmi[2], Ahmad Teguh Satria [3]

Teknologi Rekayasa Multimedia 1, Politeknik Mardira Indonesia 1, Teknologi Rekayasa Multimedia 2, Politeknik Mardira Indonesia 2, Teknologi Rekayasa Multimedia 3, Politeknik Mardira Indonesia 3
E-mail: jabbar@poltekmi.ac.id 1 fikriesma@poltekmi.ac.id.2, ateguh@poltekmi.ac.id 3

ABSTRACT

In the world of education there are often problems how to determine the quality in an educational institution and often institutions ignore it, therefore to find out the quality of education in an institution requires clustering using the K-Means method which will be divided into 3 (three) cluster levels, namely "A" for very good quality levels, "B" for good quality levels and "C" for low education quality levels and to determine the quality of institutions We used data on the average report card of grade X, XI, XII students of all majors for one year.

Keywords : *KMeans, Clustering,* SMK Ma'arif Al-Mizan

1. INTRODUCTION

1.1. Background

The demand for quality school graduates and services is increasingly urgent and competition in the world of work is increasing. One of the impacts of globalization in education is deregulation that allows foreign educational institutions to open schools in Indonesia (Ardhana & Mayasari, 2021)

Therefore, competition between educational institutions and the job market will be increasingly heavy, so that improving the quality of education is something that needs to be improved to anticipate changes - changes that are so rapid and challenges that are getting bigger and more complex.

According to him, schools need to work on how to improve the competitiveness of graduates as well as academic products and other services. (Tanjung et al., 2021)

SMK Ma'arif Al-Mizan Jatiwangi - Majalengka has approximately 1000 students in 3 (three) batches. Because the number of students is very large, an analysis is needed to determine whether the institution has many Smart, Medium and Low IQ students, to determine the quality of the Educational Institution taken from the average report card scores in 2021-2022 classes X, XI, XII All Majors. To create a good teaching and learning atmosphere and obtain optimal results.

The grouping procedures include leveling variations in academic scores in one class consisting of low-achieving to high-achieving students, balancing the number of men and women in the class, equitable distribution of student behavior that needs special attention, placing students based on alphabetical variations in student names and minimizing the number of students who have almost the same name in one class.

Grouping students into classes at SMK Ma'arif Al-Mizan Jatiwangi still uses a manual system. With a manual system, teachers, especially counseling guidance teachers, have difficulty in dividing a large number of students into several classes with certain criteria. This requires a very long turnaround time. Manual grouping results in uneven variation in grades in each class and there are often errors in the form of duplication of student names in different classes. Therefore, software is needed that can help counseling guidance teachers in dividing students into classes in order to facilitate and speed up the process of dividing students into classes according to grouping procedures.

The implementation and testing of student grouping at SMK Ma'arif Al-Mizan Jatiwangi using the K-Means Clustering method is aimed at simplifying and accelerating the process of grouping students and resulting in the division of students into classes with academic grades, student names, student behavior and student gender equally.

1.2. Library Survey

(Feby Tri & Nataliani, n.d.) Conducting a study entitled "Analysis of the Effect of Assessor Assessment on Subject Teacher Performance with K-Means Clustering" in his research conducted research on teacher performance The purpose of this study was to determine teacher performance based on assessment standards by grouping into three clusters, namely very good, good, and good enough. The value of these criteria is processed by the k-means clustering method.

(Aranski & Handoko, n.d.) Conducting a study entitled "DATA MINING IN GROUPING STUDENT IQ SCORES" in his research conducted research on the aim of knowing the characteristics of grouping student IQ score data in data processing using the *K-Means method*.

1.3. Research Methodology

a. Types of Research

The type of research used in this study is a type of applied research. Applied research is a research to produce something that can be directly applied to solve a problem. In this project, the authors apply the K-Means algorithm to an analysis to determine the quality of education in students. The research data was taken from the object of research, namely SMK Ma'arif Al-Mizan located in Majalengka

b. Data Types and Sources

a) The type of data used is quantitative data. Quantitative data is a type of data that can be calculated, in the form of numbers or nominal. Historical data of sales transactions is a type of quantitative data because it is a number or nominal and can be calculated. More specifically, the data used is matrix data, which is a type of data that has objects and attributes.

b) Data sources used in research are Primary Data and Secondary Data. Primary data sources are data sources obtained directly from the original source and not through intermediary media. The historical data of sales transactions used are obtained directly from the object of research through interviews and documentation. While secondary data is a source of research data obtained indirectly through intermediary media obtained and recorded by other parties. Secondary data is generally in the form of evidence of published historical records or reports. The secondary data that the researcher intended in this study is a data source used to support the completeness of prime data theory

c. Data Collection Methods

Based on the data sources used in this final project research, the data collection methods used by the author are as follows,

a) Interview, is a method of collecting data through direct question and answer activities with people in the object of research. The author conducted interviews with employees and principals of SMK Ma'arif Al-Mizan.

b) Documentation is a method of collecting data by collecting notes or documents - documents the author gets from the interview method, the author collects it into one so that it becomes a document that is ready to be used for research purposes.

c) Literature Study, is the study of scientific papers, scientific books, and other scientific sources that are in accordance with research and have a relationship with the problem under study. The scientific references that the author uses are the sources contained in the bibliography.

d. Data Mining

According to the book Data mining written by Neni Purwati, *Data Mining* is the process of uncovering unexpected relationships in data that are needed to reveal the underlying relationships in *scatteplots* that contain big *data.* according to Efraim Turban (2005), *Data Mining* is a term to describe information contained in a set

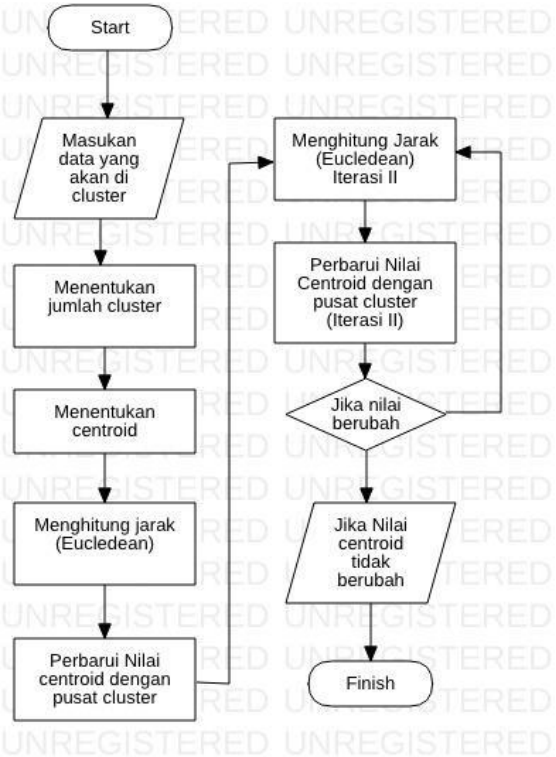
of data. *Data mining* is the process of using static, mathematical, artificial intelligence (*artificial intelligence*) techniques, and machine learning to sort and identify useful information and related knowledge from various large databases. while according to Maimon and Last (2000) *Data Mining* is part of the KDD (*Knowledge Discovery in Databases*) process) which consists of several stages such as data selection, pre-processing, transformation, *data mining*, and evaluation of results. So in thousands of school information databas it is very difficult to select data, for example in determining class IQ.

e. Data Cluster

In the K-means journal written by Yudi Agusta, PD, *Data Clustering* is one of the *Data Mining* methods that is araan, There are two types of data clustering that are often used in data grouping, namely data hierarchy and clustering. in the journal Nur Janna, *Clustering* is also known as *segmentation* and tools for data analysis whose task is to solve the problem of grouping in this al people, events and other objects.

f. Methodology K-Means

This research uses data mining. In this industrial era 4.0 where the development of data into big data is needed, a method is needed to process big data to be useful for our lives, one method is data mining. Where in data mining is a way to automate processes to find patterns from large-scale data collections. This research uses data mining. In this industrial era 4.0 where the development of data into big data is needed, a method is needed to process big data to be useful for our lives, one method is data mining. Where in data mining is a way to automate processes to find patterns from large-scale data collections. (Happy Siregar et al., 2021) (Happy Siregar et al., 2021)



Gambar 1. Flowchart diagram K-Means

(Source : Happy Siregar, 2021)

Data mining is used to process big data in databases so as to produce new information that is useful for business strategies. Broadly speaking, data mining is divided into two main categories, namely Descriptive Mining and Predictive Mining, where descriptive mining is used to determine data characteristics, while predictive mining is used to find data patterns. (Darmi et al., 2016)

done with stages that refer to the stages of the K-Means method, K-Means is one of the methods in data mining. Where in this method is carried out data grouping (clustering) with a partition system and modeling without

supervision [4]. In this method, the grouping of data is carried out into several groups, and each group has similar characteristics and has characteristics with other groups (Indrayani & Irbiani, 2019)

The K-Means algorithm performs point-based clustering (centroid) by determining three parameters, namely the number of clusters, cluster initialization and system distance adjusted to the object of study.

The steps taken to form clustering in the K-Means method are:

a. Specifies the number of clusters (k) for the number of clusters from the existing dataset.
b. Determining k as Centroid, usually done randomly (random).
c. Calculate data distance with centroid using distance formula using Euclidean formula (equation 1) Distance:

$$d_{euclidean(x,y)} = \sqrt{\sum_i (x_i - y_i)^2} \text{ ..... (1)}$$

where d is the document point, xi is the criterion data and μj is the centroid on the j-th cluster.

d. Group the data based on proximity to the centroid and then update the new centroid value with the location from the center of the cluster using equation 2:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \text{ ..... (2)}$$

Where μj(t+1) is the new centroid on the (t+1) iteration and Nsj is the amount of data on the sj cluster.

e. Perform steps 2 through 4 until none of the cluster members have changed.

2. DISCUSSION

a. Prepare the collected data

Table.1. Student grade data

No	Nama	Kelas	Kepribadian/Sikap	Nilai Rata UTS (Semester 1)	Nilai Rata UAS(Semester 1)	Nilai Rata UTS (Semester 2)	Ujian Sekolah
1	Adi Ahmad Rifai	X TKJ 1	87	70	78	80	0
2	Anang Ma'ruf	X TKJ 1	54	80	78	88	0
3	Ayika Feri Nurfazri	X TKJ 1	84	60	57	70	0
4	Dilan Abdul Muhyi	X TKJ 1	60	80	77	89	0
5	Dimas Riqma Nurfajar	X TKJ 1	80	87	79	69	0
6	Herlin Hermubayu	X TKJ 1	87	54	96	45	0
7	Ikmal Maulana	X TKJ 1	54	84	76	72	0
8	Mughni Rizfa Fairuz	X TKJ 1	84	60	57	70	0
9	Maman Wasdiman	X TKJ 1	71	80	77	89	0
10	Muhamad Dedi Rohiman	X TKJ 1	54	87	79	69	0
11	Muhamad Riki Rijmadin	X TKJ 1	84	54	96	45	0
12	Muhamad Syamsul	X TKJ 1	54	84	76	72	0
13	Muhammad Dede Aryanto	X TKJ 1	84	71	85,5	50	0
14	Muhammad Nawal Ramadani	X TKJ 1	60	54	80	45	0
15	Nana Sulanjana	X TKJ 1	80	84	76	72	0
16	Nurhaermawan	X TKJ 1	87	54	69	57	0
17	Desi Agustian	X TKJ 1	54	84	45	77	0
18	Defani Reuzista	X TKJ 1	84	60	72	79	0
19	Dinda Zalima	X TKJ 1	71	80	70	96	0
20	Lutfiatur Rosyidah. F	X TKJ 1	54	87	89	76	0
21	Putri Febrianti	X TKJ 1	96	54	69	85,5	0
22	Siti Hafidoh	X TKJ 1	70	84	45	50	0
23	Vinandini Vibriani	X TKJ 1	80	71	69	79	0
24	Ziha Najhatul Aeni	X TKJ 1	60	54	45	42	0

b. Determining centroid value

Determine the center of the cluster randomly, creating the *initial centroid* or center point of the *cluster*, suppose in the following table.

Table 2. Centroid

No	Nama	Kelas	Kepribadian/Sikap	Nilai Rata UTS (Semester 1)	Nilai Rata UAS(Semester 1)	Nilai Rata UTS (Semester 2)	Ujian Sekolah	Cluster
1	Adi Ahmad Rifai	X TKJ 1	87	70	78	80	0	C1
4	Dilan Abdul Muhyi	X TKJ 1	60	80	77	89	0	C2
9	Maman Wasdiman	X TKJ 1	71	70	77	70	0	C3

c. Calculating Distances using Euclidean formulas

The distance of each BI data to each centroid at. Suppose the data used is a student named Adi Ahmad Rifai who has a value (87,70,78,80,0), then the distance between the data and each centroid is:

C1=

$$d\,(xi-\mu i)=\sqrt{\sum}\quad((87-87)^2+(70-70)^2+(78-78)^2+(80-80)^2$$

$$d\,(xi-\mu i)=0$$

C2=

$$d\,(xi-\mu i)=\sqrt{\sum}\quad((87-60)^2+(70-80)^2+(78-77)^2+(80-89)^2$$

$$d\,(xi-\mu i)=30.18$$

C3=

$$d\,(xi-\mu i)=\sqrt{\sum}\quad((87-71)^2+(70-70)^2+(78-78)^2+(80-80)^2$$

$$d\,(xi-\mu i)=18.89$$

So for the closest result using the Euclidean formula as follows is:
C1 (Adi Ahmed Rifai),
K2 (Ang Ma'Aroof, Dilan Abdul Mahi, Akmal Maulana),
C3 (Ayika Feri N, Dimas Riqma N, Herlin Hermabayu, Mughni Rizfa F, Maman Wasdiman, M Dedi Rohiman):

Table 3. Iteration Results 1

No	Nama	Kelas	Kepribadian/Sikap	Nilai Rata UTS (Semester 1)	Nilai Rata UAS(Semester 1)	Nilai Rata UTS (Semester 2)	Ujian Sekolah	C0	C1	C2	Nilai Minimal
1	Adi Ahmad Rifai	X TKJ 1	87	70	78	80	0	0,00	30,18	18,89	0,00
2	Anang Ma'ruf	X TKJ 1	54	80	78	88	0	35,40	6,16	26,72	6,16
3	Ayika Feri Nurfazri	X TKJ 1	84	60	57	70	0	25,50	41,68	25,87	25,50
4	Dilan Abdul Muhyi	X TKJ 1	60	80	77	89	0	30,18	0,00	24,12	0,00
5	Dimas Riqma Nurfajar	X TKJ 1	80	87	79	69	0	21,45	29,21	19,36	19,36
6	Herlin Hermubayu	X TKJ 1	50	54	50	45	0	60,28	58,66	45,29	45,29
7	Ikmal Maulana	X TKJ 1	54	84	76	72	0	36,78	18,49	22,14	18,49
8	Mughni Rizfa Fairuz	X TKJ 1	50	60	57	70	0	44,83	35,51	30,68	30,68
9	Maman Wasdiman	X TKJ 1	71	70	77	70	0	18,89	24,12	0,00	0,00
10	Muhamad Dedi Rohiman	X TKJ 1	54	63	56	59	0	45,42	40,82	30,00	30,00

d. Calculate the average value of existing data on the same cluster to get a new *centroid position*. After all students have calculated the distance to all *their centroids* and determined their *cluster membership, the new* centroid *position is calculated*. Table 2 is the result of data renewal / processing values using the *centroid* values above by calculating the average value in each *cluster*.

Table. 4. Iteration II Centroid Results

Cluster	Kepribadian/Sikap	Nilai Rata UTS (Semester 1)	Nilai Rata UAS(Semester 1)	Nilai Rata UTS (Semester 2)	Ujian Sekolah
C1	87	70	78	80	0
C2	56	81,33	77	83	
C3	64,83	65,67	62,67	63,83	

e. Calculates Iteration Distance II to ensure no members change

C1=

$$d\,(xi-\mu i)=\sqrt{\sum}\quad((87-87)^2+(70-70)^2+(78-78)^2+(80-80)^2$$

$$d\,(xi-\mu i)=0$$

C2=

$$d\,(xi-\mu i)=\sqrt{\sum}\quad((87-56)^2+(70-81)^2+(78-77)^2+(80-83)^2$$

$$d\,(xi-\mu i)=30.18$$

C3=

$$d\,(xi-\mu i)=\sqrt{\sum}\quad((87-64)^2+(70-65)^2+(78-62)^2+(80-63)^2$$

$$d\,(xi-\mu i)=18.89$$

Table 5. Iteration II Results

No	Nama	Kelas	Kepribadian/Sikap	Nilai Rata UTS (Semester 1)	Nilai Rata UAS(Semester 1)	Nilai Rata UTS (Semester 2)	Ujian Sekolah	C0	C1	C2	Nilai Minimal
1	Adi Ahmad Rifai	X TKJ 1	87	70	78	80	0	0,00	33,16	33,79	0,00
2	Anang Ma'ruf	X TKJ 1	54	80	78	88	0	35,40	5,64	21,67	5,64
3	Ayika Feri Nurfazri	X TKJ 1	84	60	57	70	0	25,50	42,52	32,67	25,50
4	Dilan Abdul Muhyi	X TKJ 1	60	80	77	89	0	30,18	7,33	31,28	7,33
5	Dimas Riqma Nurfajar	X TKJ 1	80	87	79	69	0	21,45	28,43	29,52	21,45
6	Herlin Hermubayu	X TKJ 1	50	54	50	45	0	60,28	54,37	26,42	26,42
7	Ikmal Maulana	X TKJ 1	54	84	76	72	0	36,78	11,54	17,95	11,54
8	Mughni Rizfa Fairuz	X TKJ 1	50	60	57	70	0	44,83	32,56	17,33	17,33
9	Maman Wasdiman	X TKJ 1	71	70	77	70	0	18,89	22,86	13,87	13,87
10	Muhamad Dedi Rohiman	X TKJ 1	54	63	56	59	0	45,42	36,84	128,52	36,84

f. Repeat Step to d to e until the centroid value does not change, until the 4th iteration.

Here's the final result of iteration 4 with unchanged values:

C1 : (Adi Ahmad, Ayika Feri, Dimas Riga, Mother Wasdiman)

(Ang Ma'Rouf, Dilan Abdul Mahi, Akmal Maulana)

C3 : (Herlin Hermabayu, Mughni Rizfa, M Dedi Rohiman)

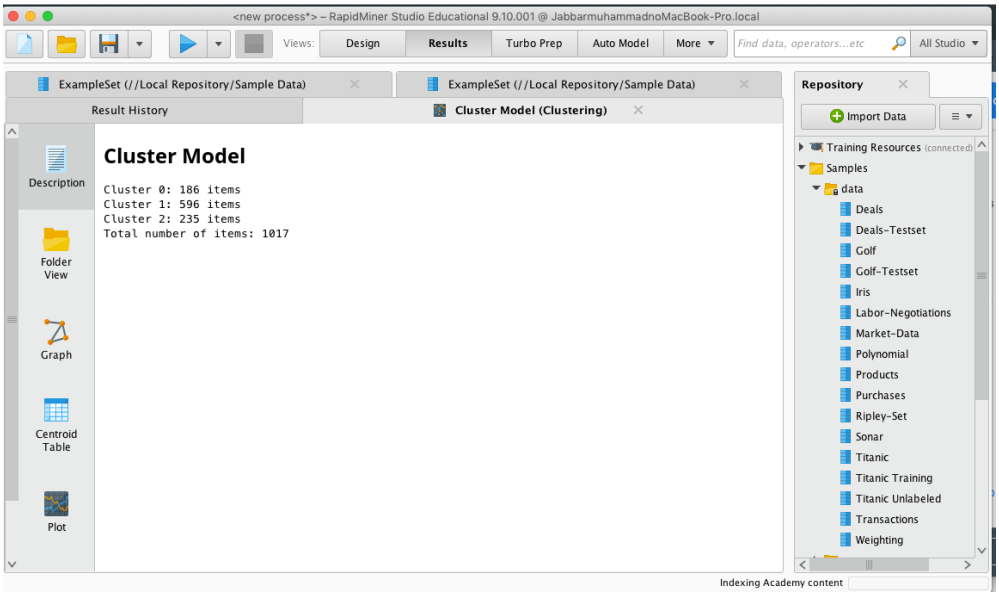
Table 6. Iteration IV Results

No	Nama	Kelas	Kepribadian/Sikap	Nilai Rata UTS (Semester 1)	Nilai Rata UAS(Semester 1)	Nilai Rata UTS (Semester 2)	Ujian Sekolah	C1	C2	C3	Nilai Minimal
1	Adi Ahmad Rifai	X TKJ 1	87	70	78	80	0	11,53	33,16	49,37	11,53
2	Anang Ma'ruf	X TKJ 1	54	80	78	88	0	32,34	5,64	43,68	5,64
3	Ayika Feri Nurfazri	X TKJ 1	84	60	57	70	0	20,09	42,52	34,92	20,09
4	Dilan Abdul Muhyi	X TKJ 1	60	80	77	89	0	28,05	7,33	44,62	7,33
5	Dimas Riqma Nurfajar	X TKJ 1	80	87	79	69	0	16,81	28,43	48,32	16,81
6	Herlin Hermubayu	X TKJ 1	50	54	50	45	0	50,05	54,37	14,65	14,65
7	Ikmal Maulana	X TKJ 1	54	84	76	72	0	29,38	11,54	36,02	11,54
8	Mughni Rizfa Fairuz	X TKJ 1	50	60	57	70	0	36,35	32,56	12,41	12,41
9	Maman Wasdiman	X TKJ 1	71	70	77	70	0	10,79	22,86	34,14	10,79
10	Muhamad Dedi Rohiman	X TKJ 1	54	63	56	59	0	35,14	36,84	5,19	5,19

3. CONCLUSION

Based on the description that has been stated in the previous chapters, several conclusions and suggestions can be drawn as follows:

a. Conclusion The Clustering *method with the* K-Means *algorithm* can be used to group student data based on the average score of report cards each semester from the data collected in the table above students of SMK Ma'arif Al-Mizan with the number of students with a low IQ of 186 students, medium 596 and high 235, it can be concluded that the quality of education at SMK Ma'arif Al-Mizan has not reached good quality, Here are the results of the rapid miner maximum of 10 repetitions,



Picture. 2 Results of Clustering using Rapidminer

b. Suggestion

From these conclusions, it is necessary to conduct a deeper analysis so that students can get the best grades so that students can compete / continue to the university level.

BOOK

Aranski, A. W., & Handoko, K. (n.d.). *DATA MINING IN GROUPING STUDENTS' IQ SCORES*.

Ardhana, no. R., & Mayasari, D. (2021). Jargon Santri Asrama Ibnu Sina Pandak Peshendran Darul Uloom Peterongan Jambang. In *Journal OF Education Research* (Vol. 2, Issue 2).

Darmi, Y., Setiawan, A., Bali, J., Kampung Bali, K., Teluk Segara, K., &; Bengkulu, K. (2016). APPLICATION OF K-MEANS CLUSTERING METHOD IN GROUPING PRODUCT SALES. In *Media Infotama Journal* (Vol. 12, Issue 2).

Efraim *Turban*, dkk. 2005. *Decision Support Systems and. Intelligent Systems*

Feby Tri, M., & Nataliani, Y. (n.d.). *Analysis of the Effect of Assessor Assessment on Subject Teacher Performance with k-Means Clustering*.

Indriyani, F., & Irfiani, E. (2019). *Clustering Data Penjualan* (Vol. 7, Issue 2).

Madhulatha, T.S., 2012. An Overview On Clustering Methods. IOSR Journal of Engineering, II(4), pp.719-725

Congratulations Siregar, Vanreider Sinaga, &; Elisa Citra Manihuruk. (n.d.). *THE EFFECT OF THE COMPENSATION SYSTEM ON EMPLOYEE WORK PRODUCTIVITY ON CV. ZEGES MEDAN.*

Tanjung, R., Supriani, Y., Mayasari, A., Arifudin, O., Rakeyan, S., &; Karawang, S. (n.d.). *Quality Management in the Provision of Education*. https://doi.org/10.32529/glasser.v6i1.1481

Y. Agusta, *"K-Means Application, Related Problems and Methods," Journal of Systems and Informatics, vol. 3,* pp. 47-60, February 2007.