KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
Vol. 19 (2) (2022), 124 –138 p-ISSN: 1693-7554, e-ISSN:2654-3990

Classification of Organizational Data Using the K-Means Algorithm

Abdul Fadlil[1], **Imam Riadi**[2], **Yana Mulyana**[3*]

[1]Program Studi Teknik ElektroUniversitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Program Studi Sistem Informasi, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[3]Program Studi Magister Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Abstract

The University of Muhammadiyah Tasikmalaya (UMTAS) is a private university under the Muhammadiyah association located in Tasikmalaya City, West Java Province. As a form of developing student soft skills, UMTAS carries out many activities, one of which is through student organizations consisting of internal and external organizations. This research was conducted to find out which student organizations are categorized as excellent, good, and not good. The assessment attributes for grouping student organization data are the number of active members, activities in one year, organizational discipline, and achievements. Clustering uses the K-Means algorithm. The results obtained from calculations carried out manually and using the rapidminer application obtained the same results, namely clusters with the "excellent" category totaling 8 data (26.6%), clusters with the "good" category totaling 17 data (56.6%), and clusters with the "not good" category totaling 5 data (16.6%). The results of this study can be used by the head of the bureau of academic administration of student affairs and alumni in providing rewards in the form of priorities in organizational funding, awarding charters and punishments in the form of coaching and revoking student organization decrees.

Keywords: *K-Means; rapidminer; clustering; organizations; classification;*

1. Introduction

Student organizations are one of the attributes of universities that have an important role so it is stated in the Constitution of the Republic of Indonesia Number 12 of 2012 concerning Higher Education in article 77 about student organizations [1] One of the functions of student organizations is as supporting education in student soft skills such as learning to cooperate, solving problems, to express opinions with no fear, thinking critically and being able to be responsible. Students are also referred to as agents of change which have the meaning of being a form of change for society because they are considered capable of overcoming problems that arise in society, by prioritizing clear thinking patterns that are critical, creative, innovative and responsible so that they are expected to be able to provide changes in the form of work that is beneficial to society and the environment [2]. In its implementation, student organization activities must get support from universities, both in the form of activity funds and in the form of coaching so that student organizations can develop, be useful and productive so as to produce innovative work. The provision provided by universities includes competence (hard skills) and the cultivation of character values (soft skills) with the aim

[*]**Corresponding author**. *E-mail address: yana2107048012@webmail.uad.ac.id*
Received: 07 July 2022 , Accepted: 17 July 2022 and available online 30 July 2022
https://doi.org/10.33751/komputasi.v19i2.5676

that students have high sensitivity and are easy to adapt in their environment, both in society and in the industrial world [3].

University of Muhammadiyah Tasikmalaya is a private university located in the city of Tasikmalaya with 3 faculties and 13 study programs, the number of all student organizations at the University of Muhammadiyah Tasikmalaya based on the decree of the college as many as 30 student organizations which are divided into 4 categories, namely the university, faculty, study program and ORTOM (Autonomous Organization) levels. In the implementation of its activities, there are often many student organizations that are not in accordance with the provisions of student guidelines so that they need to be grouped into three categories, namely Good, Not Good and Very Good with the aim of providing information to the Bureau of Academic Administration of Student Affairs and Alumni in providing policies in the form of rewards and punishments. This is very underlying for the author to conduct research on grouping student organizations at the University of Muhammadiyah Tasikmalaya by using the method of applying the K-Means Clustering algorithm. K-Means is a clustering algorithm that can determine an object distance and centroid (central point) by dividing data into a cluster through certain stages [4]. K-Means is a non-hierarchical cluster data method that attempts to divide existing data into one or more clusters/groups. K-Means clustering is a method of data analysis or data mining that carries out an unsupervised modeling process by grouping data into a partition [5], [6]. The application of the K-Means algorithm has been widely carried out by previous researchers including research conducted by sembiring et al, the K-Means method is used to group student achievements so that it can help employees and institutions to improve the quality of education by knowing student learning achievement [7]. The K-means method is also used to analyze factors affecting smokers switching to using alternative tobacco (vape) products [8]. This research also uses the K-Means method in its classification based on the assessment of organizational attributes so as to produce a grouping of organizations with excellent, good and not good categories.

2. Methods

2.1 Stages of Research

This method provides important images in making the first step of the research, the initial steps of the research carried out by the author are as follows:

(a) **Study of literature**

Literature research is a scientific activity carried out to solve a problem, the purpose is to make a theoretical or practical contribution to the development of relevant disciplines [9]. The literature in this study includes the processing of student organization data at the University of Muhammadiyah Tasikmalaya and the implementation of the K-Means algorithm. Literature study is one of the important elements in research to examine problems that will be discussed theoretically through the study of various journals and other references.

(b) **Data collection**

The data collection method is to obtain the information to achieve research objectives [7]. At this stage, the data collection process is carried out by interviewing with student managers, namely the head of the student academic bureau and alumni of the University of Muhammadiyah Tasikmalaya. For the data collection procedure can be seen in figure 1:

a The author conducts interviews with the heads of student academic bureaus and alumni.

b The author asks for data on student organizations at the University of Muhammadiyah Tasikmalaya.

c The author collects data from the requested student organization.

d The author groups the data from those that have been collected based on the name of the student organization.

e The author makes an analysis of the data that has been grouped.

f From the results of data analysis, an expected result was obtained.

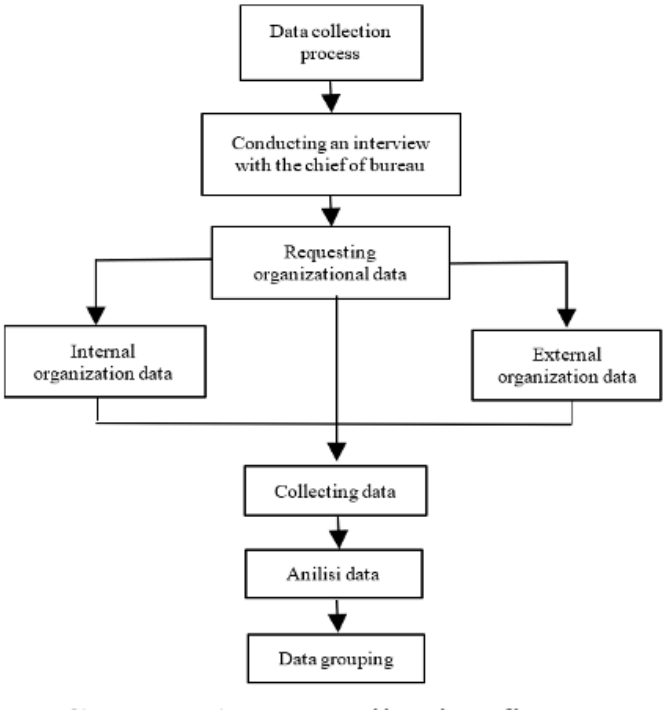

Figure 1. Data collection flow

(c) **Data analysis**
Analysis is an effort to understand a problem so that it can provide an assessment with the aim of achieving a more specific understanding of the problem being investigated [10]. Data is a set of information obtained from search results or investigations from a particular source of information that is considered competitive. In this study the authors used qualitative data analysis.

(d) **Prepare the dataset**
In this study involving all student organizations of the University of Muhammadiyah Tasikmalaya, the data includes the number of active members, the number of activities in 1 year, organizational discipline and achievements. All data is prepared in .xslx format. can be seen in Table 1.

Table 1. Student organization data

No	Organization name	Active member	Activities for 1 year	Organizational Discipline	Achievements
1	DPM-U	20	10	Complete	None
2	DPM – FIKes	15	9	Incomplete	None
3	DPM-FKIP	18	11	Incomplete	None
4	DPM-FT	10	8	Incomplete	None
5	BEM-U	25	15	Incomplete	Local
6	BEM-FIKes	17	9	Incomplete	Local
7	BEM-FKIP	19	10	Incomplete	Local
8	BEM-FT	10	9	Incomplete	None
...	...	...	...	...	...
30	PK IMM FT	15	7	Incomplete	None

(e) **Preprocess the data**
Data pre-processing is an important and necessary step in the data mining process

because it is closely related to the preparation and creation of the original data set. Data quality is a determinant in the data mining process, therefore data must not be empty or there are missing data values, so it must be ensured in advance that the data entered is correct and there is no data that is anomalous because it can affect the clustering results [8], [11]. Easier way to cluster data, data initialization must be carried out first, namely converting nominal data into numerical data. In this study, what is referred to as numerical data is organizational discipline and achievement.

a For the disciplinary attributes of the organization, changes are made from nominal to numerical, the assessment parameters in this attribute are work programs, financial reports of activities, performance evaluations and organizational accountability reports. If all parameters exist then given a full description, if the parameter is less than < 4 is given an incomplete description and if there is no parameter at all given a description does not exist. The change in nominal to numeric in this attribute can be seen in Table 2. In table 2. Initialization of organizational discipline data

Table 2. Initialization of organizational discipline

Discipline	Initial Organizational
None	1
Incomplete	2
Complete	3

is carried out, namely in the description of data there is no initialization 1, in the description of incomplete data is given initialization 2 and in the description of the complete data is initialized 3.

b For the attributes of achievement achievements, changes are made from nominal to numerical, the parameters in this achievement are international, national, regional, local/university and there is no achievement. The nominal to numerical change can be seen in Table 3

Table 3. Initialize achievements

Achievements	Initial
There isn't any	1
Local/University	2
Region	3
National	4
International	5

In Table 3, initialization of achievement data is carried out, in this data for student organizations that do not have achievements, initialization is given 1, student organizations that have achievements at the local / University level are initialized 2, student organizations that have achievements at the regional level are given initialization 3, student organizations that have achievements at the national level are initialized 4 and student organizations that have achievements at the international level are given initialization 5.

2.2 Data Mining

Data mining is a method to extract prediction information hidden in a database, it is a technology that has great potential to improve data warehouses. Data mining has a different meaning knowledge discovery or pattern recognition used to extract hidden knowledge from very large data sets [12]. "The main purpose of data mining is to find, dig, or mine knowledge from the data or information we have"[13]. In a board outline, data mining divided to two main categories:

a Descriptive mining : the process of finding important characteristics of the data in the database. Data mining techniques, including descriptive mining, are grouping, association, and sequential mining.

b Predictive : This is the process of finding patterns in data by using some other variable in the future. One of the techniques involved in predictive mining is classification simply put data mining and usually referred to the process by which knowledge is filtered or "mined" from a large amount of data another term for data mining is the Knowledge Discovery Database (KDD) [14]. For KDD steps can be seen in Figure 2.

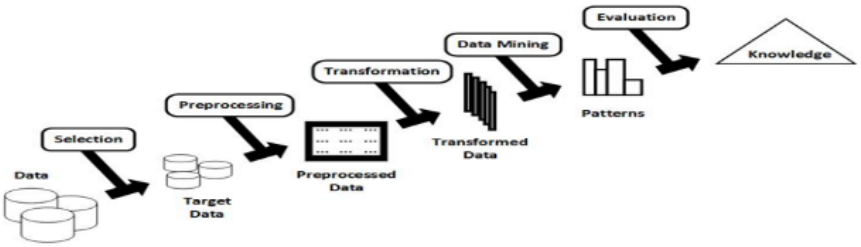

Figure 2. KDD steps

2.3 RapidMiner

Rapidminer is Software developed by Dr. Markus Hofmann of the Blanchardstown Institute of Technology and Ralf Klinkenberg of rapid-i.com. Rapidminer features a graphically based user interface so it is easier to use, this open source based software is written using Java programs under the GNU Public License, rapidminer can run on any operating system. Rapidminer has about 500 data mining operators such as input, output, data preprocessing, visualization operators and others [15],[16]. In this study, the author will compare the results of manual data processing and data processing using the RapidMiner application.

2.4 K-means Clustering

K-Means is sometimes also referred to data analysis technique or data mining technique that runs the data modeling process without supervision, this is one of the techniques for grouping data in groups or partition systems that are adapted from characteristics of each piece of data. The term clustering in k-means is used to describe an algorithm that assigns each element to a cluster with the nearest centroid (mean) starting with splitting or separating objects into the initial cluster k, then assigning objects to the entire nearest centroid cluster usually calculated based on Euclidean distances [17]. K-Means clustering technique is a clustering technique that is known to be fast and easy. K-means clustering is a non-hierarchical data clustering technique that groups data into one or more clusters. Data with the same characteristics are grouped into clusters, and data with different characteristics are grouped into other clusters, so there is little variation of data within the cluster . K-Means algorithm performs can be seen in Figure 3.

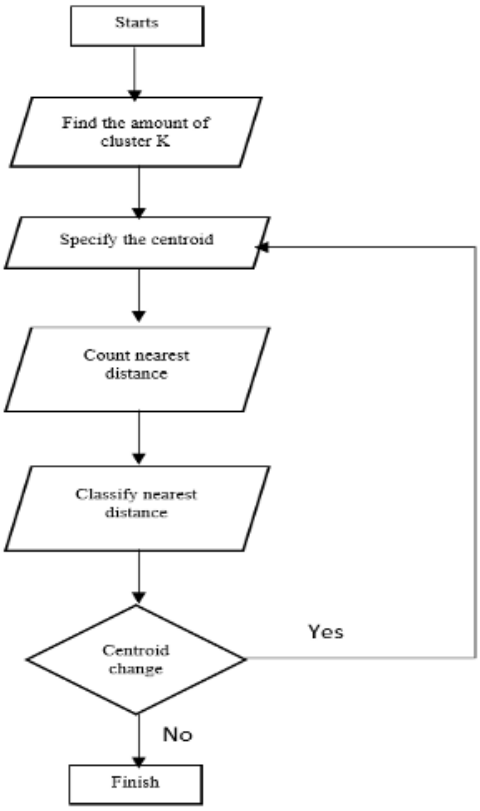

Figure 3. Flowchart algorithm of the K-Means method

Figure 3 shows the steps in determining clustering using the K-means algorithm method. The clustering stage uses the K-means algorithm as follows:

a. Enter data
b. Specify amount of cluster
c. The clustering stage uses the K-means algorithm as follows
d. Calculate the distance between the data and the cluster center, using the Eq. 1

$$D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + ... + (x_{ki} - x_{kj})^2} \quad (1)$$

Where
(i,j) : data distance i to cluster center j
x_{ki} : data to i on the attribute to j
x_{kj} : center point to j on the attribute to i

e. Recalculate the cluster center with the new cluster member, To determine the center of the cluster (centroid) from the data present in each cluster obtained from the following Eq.2

$$CKJ = \frac{x_{1j} + x_{2j} + ... + x_{nj}}{n} \quad (2)$$

Explanation
C_{kj} : K-th cluster center on the j-th variable, where $j = 1, 2, ..., p$.
n : Lots of data on the k-th cluster

f. If the cluster center has not changed then the clustering process has been completed, if not then repeat the step to d until the cluster center has not changed again.

3. Result and Discussion

3.1 Specifying the value of K as the number of clusters to be created

Data processing on student organization information is calculated manually, this process uses 3 clusters "Excellent (C1), Good (C2), Not Good (C3) and data attributes such as the number of active members, the number of activities for 1 year, organizational discipline and achievements. for data on the attributes of student organization institutions can be seen in Table 4

Table 4. Student organization attribute data

N0	Organization Name	Active Members	Activities for 1 year	Organizational Discipline	Achievements
1	DPM-U	20	10	3	1
2	DPM – FIKes	15	9	2	1
3	DPM-FKIP	18	11	2	1
4	DPM-FT	10	8	2	1
5	BEM-U	25	15	2	2
6	BEM-FIKes	17	9	2	2
7	BEM-FKIP	19	10	3	2
8	BEM-FT	10	9	2	1
9	HIMA S1 Keperawatan	25	12	3	3
10	HIMA D3 Kebidanan	20	8	2	1
11	HIMA D3 keperawatan	25	11	2	1
12	HIMA PTI	20	5	2	1
13	HIMA BK	30	13	3	1
14	HIMA PGPAUD	15	9	3	1
15	HIMA PGSD	18	10	2	1
16	HIMA SENDRATASIK	20	8	2	4
17	HIMA Teknik Mesin	18	14	2	1
18	HIMA Teknik Lingkungan	15	9	2	1
19	HIMA Teknik Pertambangan	18	7	2	1
20	HIMA Teknik Elektro	17	8	2	1
21	UKM Voly	30	17	2	3
22	UKM Futsal	20	7	1	1
23	UKM Badminton	25	8	1	1
24	UKM Tapak suci	24	12	2	4
25	UKM Mapala	33	15	3	3
26	Hizbuwaton	22	7	3	3
27	IMM Korkom	11	5	2	1
28	IMM PK FIKes	8	4	1	1
29	IMM PK FKIP	12	5	1	1
30	PK IMM FT	15	7	2	1

Furthermore, what needs to be considered is to choose the number of clusters you want then determine the centroid randomly. Random determination of centroids can be seen in Table 5

Table 5. Sample centroid

Cluster Amount Cluster	3 Cluster Center Random Organization Name	Members	Activities	Discipline	Achievement
C1	HIMA D3 KEP	25	11	2	1
C2	HIMA TL	15	9	2	1
C3	IMM PK FIKES	8	4	1	1

3.1 **Calculating the distance in each data against each centroid**

Furthermore, calculating the distance from each data to the cluster center with calculations using the Euclidean Distance equation, then a matrix of calculating the distance from the first data to the 30th data will be obtained, to get the C1 result can be calculated as follows:

$$\begin{aligned} d &= \sqrt{(25-20)^2+(11-10)^2+(2-3)^2+(1-1)^2} = 5.20 \\ d &= \sqrt{(25-15)^2+(11-9)^2+(2-2)^2+(1-1)^2} = 10.20 \\ d &= \sqrt{(25-18)^2+(11-11)^2+(2-2)^2+(1-1)^2} = 7 \\ d &= \sqrt{(25-10)^2+(11-8)^2+(2-2)^2+(1-1)^2} = 15.30 \end{aligned}$$

Continued until the 30th data...

To get the C2 result, calculate the distance from the first data to the 30th data as follows:

$$\begin{aligned} d &= \sqrt{(25-20)^2+(9-10)^2+(2-3)^2+(1-1)^2} = 5.20 \\ d &= \sqrt{(25-15)^2+(9-9)^2+(2-2)^2+(1-1)^2} = 0 \\ d &= \sqrt{(25-18)^2+(9-11)^2+(2-2)^2+(1-1)^2} = 3.61 \\ d &= \sqrt{(25-10)^2+(9-8)^2+(2-2)^2+(1-1)^2} = 5.10 \end{aligned}$$

Continued until the 30th data...

To get the $C3$ result, calculate the distance from the first data to the 30th data as follows:

$$\begin{aligned} d &= \sqrt{(8-20)^2+(4-10)^2+(1-3)^2+(1-1)^2} = 13.56 \\ d &= \sqrt{(8-15)^2+(4-9)^2+(1-2)^2+(1-1)^2} = 8.66 \\ d &= \sqrt{(8-18)^2+(4-11)^2+(1-2)^2+(1-1)^2} = 12.25 \\ d &= \sqrt{(8-10)^2+(4-8)^2+(1-2)^2+(1-1)^2} = 4.58 \end{aligned}$$

The results of the calculation of the first iteration of the entire data can be seen in Table 6

Table 6. The result of the calculation of distance and grouping of the first iteration data

Data ke	C1	C2	C3	Nearest distance	Group of data
1	5,20	5,20	13,56	5,20	C1
2	10,20	0,00	8,66	0,00	C2
3	7,00	3,61	12,25	3,61	C2
4	15,30	5,10	4,58	4,58	C3
5	4,12	11,70	20,30	17	C1
6	8,31	2,24	10,39	2,24	C2
7	6,24	4,36	12,73	4,36	C2
8	15,13	5,00	5,48	5,00	C2
9	2,45	10,68	19,00	2,45	C1
10	5,83	5,10	12,69	5,10	C2
11	0,00	10,20	18,41	0,00	C1
12	7,81	6,40	12,08	6,40	C2
13	5,48	15,56	23,85	5,48	C1
14	10,25	1,00	8,83	1,00	C2
15	7,07	3,16	11,70	3,16	C2
16	6,56	5,92	13,04	5,92	C2
17	7,62	5,83	14,18	5,83	C2
18	10,20	0,00	8,66	0,00	C2
19	8,06	3,61	10,49	3,61	C2
20	8,54	2,24	9,90	2,24	C2
21	8,06	17,12	25,65	8,06	C1
22	6,48	5,48	12,37	5,48	C2
23	3,16	10,10	17,46	3,16	C1
24	3,32	9,95	18,17	3,32	C1
25	9,22	19,10	27,46	9,22	C1
26	5,48	7,62	14,59	5,48	C1
27	15,23	5,66	3,32	3,32	C3
28	18,41	8,66	0,00	0,00	C3
29	14,35	5,10	4,12	4,12	C3
30	10,77	2,00	7,68	2,00	C2

Based on the matric results obtained in table 6, student organizations can be grouped into:
C1 : number of data 1,5,9,11,13,21,23,24,25,26
C2 : number of data 2,3,6,7,8,10,12,14,15,16,17,18,19,20,22,30
C3 : number of data 4,27,28,29
The total number of C1 is 10 data, C2 is 16 data and C3 is 4 data. Furthermore, recalculation is carried out to obtain a new centroid obtained from the results of the first iteration, in the process of obtaining a new centroid value, the calculation of the average number of each member attribute is carried out divided by the number of data on each cluster, the calculation process is as follows:

$$\begin{aligned} C1 &= \frac{20+25+25+25+30+30+25+24+33+22}{10}; \\ &= \frac{0+15+12+11+13+17+8+12+15+7}{10}; \\ &= \frac{(3+2+3+2+3+2+1+2+3+3)/11}{10}; \\ &= \frac{1+2+3+1+1+3+1+4+3+3}{10} \end{aligned}$$

$C1 = 25.90; 12.00; 2.40; 2.20$

$$\begin{aligned} C2 &= \frac{5+18+17+19+10+20+20+15+18+20+18+15+18+17+20+15}{16}; \\ &= \frac{9+11+9+10+9+8+5+9+10+8+14+9+7+8+7+7}{16}; \\ &= \frac{(2+2+2+3+2+2+2+3+2+2+2+2+2+2+1+2}{16}; \\ &= \frac{1+1+2+2+1+1+1+1+1+4+1+1+1+1+3+1}{16} \end{aligned}$$

$C2 = 17,19; 8,75; 2,06; 1,31$

$$\begin{aligned} C2 &= \frac{(10+11+8+12)/4;}{4}; \\ &= \frac{8+5+4+5}{11}; \\ &= \frac{2+2+1+1}{4}; \\ &= \frac{1+1+1+1}{4} \end{aligned}$$

$C3 = 10,25; 5,50; 1,50; 1,00.$

From the results of the calculation of the determination of the cluster center of the first iteration can be seen in Table 7

Table 7. New centroid of the first iteration data

Cluster	members	activities	disciplines	achievements
C1	25,90	12,00	2,40	2,20
C2	17,19	8,75	2,06	1,31
C3	10,25	5,50	1,50	1,00

Furthermore, recalculation is carried out as calculated in the first iteration by using a new cluster / centroid center to produce a second iteration, from the results of the calculation of the second iteration, a new centroid center is obtained which can be seen in Table 8

Table 8. New centroid of the second iteration data

Cluster	members	activities	disciplines	achievements
C1	27,13	12,88	2,25	2,25
C2	18,06	8,71	2,18	1,41
C3	10,20	6,20	1,60	1,00

Based on the results of the 2nd iteration matrix with a centroid center in table 8, student organizations can be grouped as follows:
C1 : number of data 5,9,11,13,21,23,24,25
C2 : number of data 1,2,3,6,7,10,12,14,15,16,17,18,19,20,22,26,30
C3 : number of data 4,8,27,28,29
The total number of C1 is 8 data, C2 is 17 data and C3 is 5 data

The results of the calculation of the second iteration obtained a new centroid center which can be seen in Table 9, Then the calculations are carried out to produce the third iteration,

Based on the results of the 3rd iteration matrix, student organizations can be grouped as follows:
C1 : number of data 5,9,11,13,21,23,24,25

Table 9. New centroid center of third iteration data

Cluster	members	activities	disciplines	achievements
C1	27,13	12,88	2,25	2,25
C2	18,06	8,71	2,18	1,41
C3	10,20	6,20	1,60	1,00

C2 : number of data 1,2,3,6,7,10,12,14,15,16,17,18,19,20,22,26,30
C3 : number of data 4,8,27,28,29
The total number of C1 is 8 data, C2 is 17 data and C3 is 5 data. In this calculation, the author only iterated 3 times because the results obtained were convergent, meaning that no data had moved from one cluster to another but if the calculation results were not yet at the convergence point, the iteration calculation must be continued until convergence. For more details can be seen in Table 10

Table 10. Results of grouping data of the first, second, and third iterations

First Iteration				Second Iteration				Third Iteration			
Data	C1	C2	C3	Data	C1	C2	C3	Data	C1	C2	C3
1	1			1		1		1		1	
2		1		2		1		2		1	
3		1		3		1		3		1	
4			1	4			1	4			1
5	1			5	1			5	1		
6		1		6		1		6		1	
7		1		7		1		7		1	
8		1		8			1	8			1
9	1			9	1			9	1		
10		1		10		1		10		1	
11	1			11	1			11	1		
12		1		12		1		12		1	
13	1			13	1			13	1		
14		1		14		1		14		1	
14		1		14		1		14		1	
14		1		14		1		14		1	
15		1		15		1		15		1	
16		1		16		1		16		1	
17		1		17		1		17		1	
18		1		18		1		18		1	
19		1		19		1		19		1	
20		1		20		1		20		1	
21	1			21	1			21	1		
22		1		22		1		22		1	
23	1			23	1			23	1		
24	1			24	1			24	1		
25	1			25	1			25	1		
26	1			26		1		26		1	
27			1	27			1	27			1
28			1	28			1	28			1
29			1	29			1	29			1
30		1		30		1		30		1	

From the data of table 10, it can be seen the grouping of student organizations at UMTAS,

organizations categorized as "Excellent" as many as 8 organizations, organizations categorized as "Good" as many as 17 organizations and organizations categorized as "Not Good" as many as 5 organizations.

3.3 Rapid Miner Studio Implementasi

After obtaining the results of the analysis of ormawa data grouping manually, the author analyzed the oramawa data using the RapidMiner Studio tool. Data processing using rapidminer tools can be seen in Figure 4.

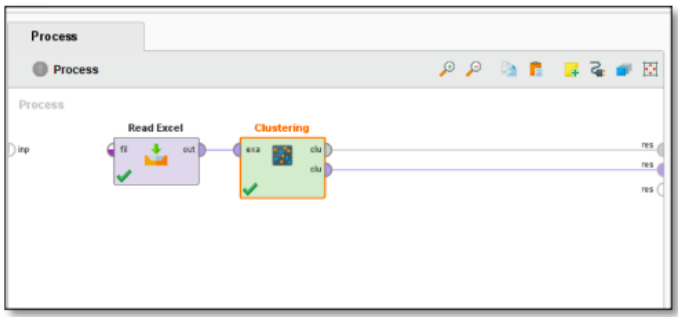

Figure 4. K-Means clustering data processing results

K-means modeling using rapidminer in figure 4 with the setting K = 3, max run 10, measure types: Bregman Divergencer, Divergence: SquaredEuclideanDistance, the modeling results are obtained as shown in Figure 5.

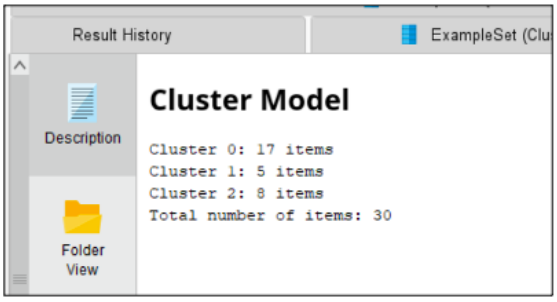

Figure 5. K-Means clustering data processing results

From the results of data processing, the following data groupings were obtained, for cluster 0 or the "good" category totaled 17 data, for cluster 1 or the "not good" category totaled 5 data and cluster 2 or the "very good" category as many as 8 data. To see which data is in each cluster, you can see Figure 6.

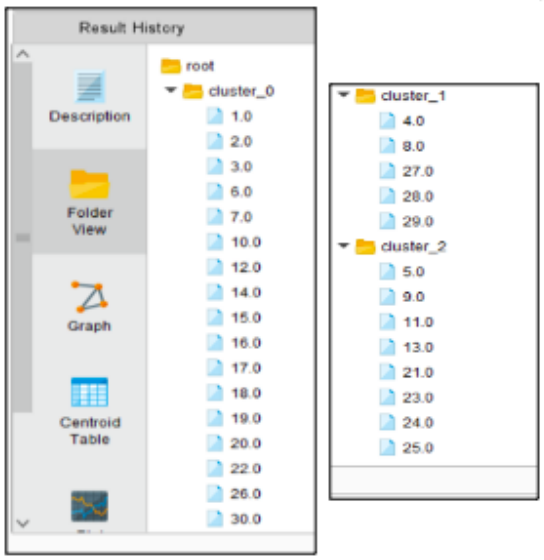

Figure 6. Display data in each cluster

From Figure 6, it can be seen that each data in each cluster is different according to the proximity of the distance between the data and the cluster center, the rapidminer clustering data in each cluster is in accordance with the grouping of data which is done manually using Ms.Excel, namely in each cluster there are 17, 5 and 8 data. The cluster deployment pattern can be seen in Figure 7. From the results of data processing, the following data groupings were obtained, for cluster 0 or the "good" category totaled 17 data, for cluster 1 or the "not good" category totaled 5 data and cluster 2 or the "very good" category as many as 8 data. To see which data is in each cluster, you can see Figure 7.

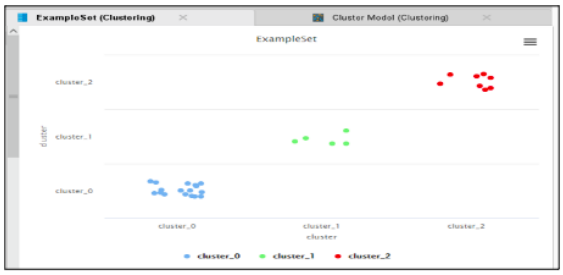

Figure 7. Display the distribution of each cluster on rapidminer

4. Conclusion

From the results of the discussion, it can be seen that the grouping of data based on each value of each attribute is grouped into 3 clusters, namely "very good", "good" and "not good". The results of the calculations carried out manually using excel and the rapidminer application obtained the same results, namely clusters with the "excellent" category totaling 8 data (26.6%), clusters with the "good" category totaling 17 data (56.6%), and clusters with the "not good" category totaling 5 data (16.6%). The results of this study can be used by the Chief of the Student Affairs, Alumni and Academic Administration Bureau as a consideration in providing rewards in the form of priorities in organizational funding, awarding charters and punishments in the form of coaching and revoking student organization decrees.

References

[1] Kementrian Sekretaris Negara, "UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 12 TAHUN 2012 TENTANG PENDIDIKAN TINGGI," Sekretaris Negara, vol. 5, no. 1. Jakarta, pp. 1–8, 2012.

[2] K. Kosasih, "Peranan Organisasi Kemahasiswaan Dalam Pengembangan Civic Skills Mahasiswa," *J. Pendidik. Ilmu Sos.*, vol. 25, no. 2, p. 188, 2017, doi: 10.17509/jpis.v25i2.6196.

[3] B. Basri and N. R. Dwiningrum, "Peran Ormawa dalam Membentuk Nilai-nilai Karakter di Dunia Industri (Studi Organisasi Kemahasiswaan di Politeknik Negeri Balikpapan)," Al-Adabiya J. Kebud. dan Keagamaan, vol. 15, no. 01, pp. 139–160, 2020, doi: 10.37680/adabiya.v15i01.273.

[4] N. Agustina and P. Prihandoko, "Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 3, pp. 621–626, 2018, doi: 10.29207/resti.v2i3.492.

[5] D. Sunia, Kurniabudi, and P. Alam Jusia, "Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means," *J. Ilm. Mhs. Tek. Inform.*, vol. 1, no. 2, pp. 121–134, 2019.

[6] H. Sy, Rismayani, and A. Syam, "Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare di Kota Makassar," *SISITI Semin. Ilm. Sist. Inf. dan Teknol. Inf.*, vol. 8, no. 1, pp. 73–82, 2019.

[7] S. N. B. S. H. W. S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," Pap. Knowl. . Towar. a Media Hist. Doc., vol. 1, pp. 82–86, 2014, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsi/article/download/4784/320.

[8] A. M. Fadhilah, M. I. Wahyuddin, and D. Hidayatullah, "Analisis Faktor yang Mempengaruhi Perokok Beralih ke Produk Alternatif Tembakau (VAPE) menggunakan Metode K-Means Clustering," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 5, no. 2, p. 219, 2020, doi: 10.35870/jtik.v5i2.182.

[9] A. K. Wardhani, "K-Means Algorithm Implementation for Clustering of Patients Disease in Kajen Clinic of Pekalongan," *J. Transform.*, vol. 14, no. 1, p. 30, 2016, doi: 10.26623/transformatika.v14i1.387.

[10] Asahar Johar, "IMPLEMENTASI ALGORITME HARD K- MEANS CLUSTERING DALAM PENENTUAN MASA PENSIUN," vol. 7, no. 2, pp. 179–186, 2019.

[11] R.Setiawan, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik Lp3i Jakarta), " J. Lentera Ict, vol. 3, no. 1, pp. 76–92, 2016.

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

[13] Susanto, Sani, and D. Suryadi, Pengantar Data mining menggali pengetahuan dari bongkahan data. Yogyakarta: Andi, 2010.

[14] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.

[15] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.

[16] M. L. Sibuea and A. Safta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustring," *Jurteksi*, vol. 4, no. 1, pp. 85–92, 2017, doi: 10.33330/jurteksi.v4i1.28.

[17] H. Annuril Chusna and A. Tuti Rumiati, "Penerapan Metode K-Means dan Fuzzy C-Means untuk Pengelompokan Sekolah Menengah Pertama (SMP) di Indonesia Berdasarkan Standar Nasional Pendidikan (SNP)," vol. 9, no. 2, pp. 216–223, 2020.