Implementation of Data Mining to Measure Informatic Engineering Graduation Using K-Means Clustering Method

Yogi Yunefri
Faculty of Computer Science
Lancang Kuning University
Pekanbaru, Indonesia
yogiyunefri@unilak.ac.id

Pandu Pratama Putra
Faculty of Computer Science
Lancang Kuning University
Pekanbaru, Indonesia
pandupratamaputra91@gmail.com

Digdaya Arief Wicaksana
Faculty of Computer Science
Lancang Kuning University
Pekanbaru, Indonesia
digdayaarief@gmail.com

***Abstract*—The Faculty of Computer Science at Lancang Kuning University is one of the favorite faculties among the faculties at Lancang Kuning University today. With a good graduation rate every year with the predicate graduation score is very satisfying. Each Study Program is obliged to monitor the progress of studies of its students. Then the study program also has the duty to pay attention to groups of students who have the potential to graduate on time and students who have the potential to experience a setback for the study period and even experience dropping out. To predict it can be done by using data mining techniques with the K-Means Clustering method. In this study, the use of Rapidminer software can be done to build a pattern of grouping the results of student graduation rates using the K-Means Clustering method of data analysis grouping the graduation rate of 2016 academic year informatics students who have conducted lectures up to semester 6 (VI) using data Semester 2 to Semester 5 GPA and total credits taken previously. Prevention of failure is very important for management of study programs. New knowledge gained in this study was used to assist study programs to better understand the situation of their students and to be able to anticipate drop-out students, to improve student achievement, to improve curriculum, improve the process of learning and teaching activities and many other benefits that could be obtained from the results of mining the data.**

Keywords*— *Data Mining, K-Means, Clustering, Graduation Rate, Student

I. INTRODUCTION

Faculty of Computer Science at Lancang Kuning University is one of the favorite faculties among the faculties at Lancang Kuning University today. With a good graduation rate every year with the predicate graduation score is very satisfying. Each Study Program is obliged to monitor the progress of studies of its students. Then the study program also has the duty to pay attention to groups of students who have the potential to graduate on time and students who have the potential to experience a setback for the study period and even experience dropping out. To predict it can be done by using data mining techniques with the K-Means Clustering method.

Understanding information about students who are potentially exposed to drop outs is important to know and understand. Understanding can be done by digging up the data that is owned and then grouping the results of data mining so as to create a pattern or group of students who are potentially exposed to drop outs. Prevention of failure is very important for the management of study programs. This knowledge can be used in helping study programs to find out the situation of their students and can anticipate drop-out students, to improve student achievement, to improve curriculum, improve the process of teaching and learning activities and many other benefits that can be obtained from the results of data mining.

II. METHOD

A. Data Mining

Data mining is a process that uses one or more computer learning techniques (machine learning) to analyze and extract knowledge automatically [3]. The basic concept of data mining is finding hidden information in a database and is part of Knowledge Discovery in Database (KDD) to find useful information and patterns from data. Data mining looks for new, valuable, and useful information in datasets involving computers and humans and is iterative either through automated or manual processes [11].

B. K-Means Clustering

K-Means Clustering is the simplest grouping method that groups data into groups based on the centroid of each group. It's just that the K-Means results are strongly influenced by the k parameters and centroid initialization. In general, K-Means initializes centroids randomly. But the proposed method will modify K-Means in centroid initialization especially in improving performance in document grouping. Following are the steps found in the K-Means algorithm [13]:

1. Determine k as the number of clusters formed.
2. Generate the initial centroid (cluster center point) randomly. Determination of the initial centroid is done randomly from the available objects as many as k clusters, then to calculate the next centroid cluster, the following formula is used :

$$\boxed{v = \frac{\sum_{i=1}^{n} x_i}{n};\ \mathrm{i} = 1,2,3,\ldots\mathrm{n}}$$

Explanation : v : centroid on the cluster
xi : object to-i

Vol 4-Sep2019
ISSN: 2541-3023

n : the number of objects who is a member of the cluster

3. Calculate the distance of each object to each centroid of each cluster. To calculate the distance between objects and centroids, you can use the Euclidian Distance calculation as follows :

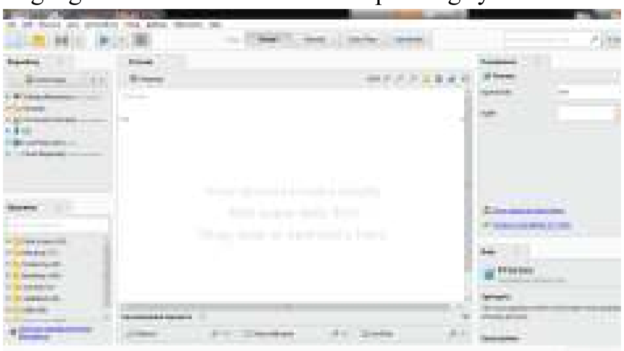
$$d(x,y) = \| x - y \| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \ ; i = 1,2,3,..,n$$

Explanation : xi : object x to-i
yi : centroid y to-i
n : the number of objects

4. Allocate each object to the closest centroid.
5. Do the iteration, then specify the position of the new centroid using equation (2).
6. Repeat step 3 if the position of the new centroid is not the same.

C. *Rapidminer*

Rapidminer is a software that is open (open source). Rapidminer is a solution for analyzing data mining, text mining and prediction analysis. Rapidminer uses a variety of descriptive and predictive techniques to provide insight to users so that they can make the best decisions. Rapidminer has approximately 500 data mining operators, including operators for input, output, data preprocessing and visualization. Rapidminer is a stand-alone software for data analysis and as a data mining machine that can be integrated into its own products. Rapidminer is written by using the java language so that it can work on all operating systems.



Picture 1. *Interface Rapidminer*

III. RESULTS AND DISCUSSION

A. *Data Preprocessing*

In preprocessing this data, it can be done by collecting raw data from the Faculty of Computer Science at Lancang Kuning University in the form of data stored in Excel format (.xls), which will be used as new data consisting of GPA data for semester 2 to semester 5 and total credit taken by Lancang University Informatics Engineering Students in the 2016 academic year which is still active in lectures to date. The process carried out at the data preprocessing stage is data integration, data cleaning, and data transformation.

Preprocessed data consists of 4 attributes to be grouped, namely Informatics Engineering GPA from semester 2 to semester 5 and total credits taken with a total of 22 data samples from a total of 134 data. Students who are active in lectures to date. The following table describes the attributes that will be used for grouping and displaying some of the processed data that has been done.

B. *Clustering Analysis with K-Means Algorithm*

Table 1. Sample Calculation Data

CODE	NIM	Name	GPA - Semester				CREDITS TOTAL
			II	III	IV	V	
M010	1655201017	LUPITA SRI REZEKI	3.74	3.65	3.61	3.44	107
M020	1655201030	TEGUH RIANTO	3.1	2.96	2.96	2.96	62
M030	1655201044	SISILIA TRIANI BR. MANULANG	3	3	3	3	20
M040	1655201058	YUKHRIZUL ALMUFAZRI	2.81	2.81	2.81	2.81	42
M050	1655201071	DEBORA KATARINA SIMAMORA	3.25	3.21	3.26	3.23	128
M060	1655201086	TAUFIK KURRAHMAN	3.13	3.08	3.04	3.06	127
M070	1655201097	ANGGA DANA PRATAMA	1	1	1	1	20
M080	1655201115	MUHAMMAD FUAD FAHRI	1.76	1.76	1.76	1.76	20
M090	1655201131	DENDY NOFRIZAL	2.9	2.88	2.92	2.92	79
M100	1655201147	RYAN BAPTISTA HUTASOIT	2.22	2.22	2.22	2.22	37
M110	1655201158	MAULANA PRAYOGO PANGESTU	3.03	2.98	2.88	2.89	91
M120	1655201172	DHEA ANISYA	2.59	2.82	2.92	2.95	122
M130	1655201187	NUZWARDANA FIKRI	1.5	1.5	1.5	1.5	32
M011	1655201018	ADE SAPUTRA	3.73	3.73	3.79	3.77	130
M022	1655201032	RONI SAFRIYADI	3.68	3.48	3.47	3.57	130
M033	1655201047	RAMADHAN SAPUTRA	2.25	2.25	2.25	2.25	35
M044	1655201063	SYAHRUL RAMADHAN	3.52	3.49	3.47	3.48	129
M055	1655201079	RIRI KUSHENDAR	2.91	2.8	2.89	2.94	127
M066	1655201092	REYNAL SATRIO	1.92	1.92	1.92	1.92	20
M077	1655201106	NOPEBRIN DUY PUTRA MANIK	2.59	2.54	2.4	2.4	66
M088	1655201128	SYAHRUN NUR	3.09	2.92	3.01	3.01	126
M099	1655201146	SISKA FERONIKA SIRINGO-RINGO	3.34	3.27	3.32	3.32	130

1. Analysis and Process of K-Means Clustering
   a. Determine the Initial Cluster Center
   Determining the initial centroid is randomly determined from the available data / objects in the number of clusters k. Where the initial number of centroids is determined as 3 initial centroids, the value for C1 is taken from the data line M020, the value of C2 is taken from row M010, the value of C3 is taken from the data line M070. Following this is the initial centroid value in the study, C is a cluster :
   C1 = (3.1 ; 2.96 ; 2.96 ; 2.93 ; 62)
   C2 = (1.5 ; 1.5 ; 1.5 ; 1.5 ; 107)
   C3 = (1.92 ; 1.92 ; 1.92 ; 1.92 ; 20)
   b. Calculating Distance with the Cluster Center
   The following is the calculation of distance with Euclidean Distance for iteration 1 with centroid 1:

$$M010 = \sqrt{(3.74-3.1)^2 + (3.65-2.96)^2 + (3.61-2.96)^2 + (3.44-2.96)^2 + (107-62)^2} = 45.02$$

$$M020 = \sqrt{(3.1-3.1)^2 + (2.96-2.96)^2 + (2.96-2.96)^2 + (2.96-2.96)^2 + (62-62)^2} = 0.00$$

$$M030 = \sqrt{(3-3.1)^2 + (3-2.96)^2 + (3-2.96)^2 + (3-2.96)^2 + (20-62)^2} = 42.00$$

$$M040 = \sqrt{(2.81-3.1)^2 + (2.81-2.96)^2 + (2.81-2.96)^2 + (2.81-2.96)^2 + (42-62)^2} = 20.00$$

$$M050 = \sqrt{(3.25-3.1)^2 + (3.21-2.96)^2 + (3.26-2.96)^2 + (3.23-2.96)^2 + (128-62)^2} = 66.00$$

....

$$M099 = \sqrt{(3.34-3.1)^2 + (3.27-2.96)^2 + (3.32-2.96)^2 + (3.32-2.96)^2 + (107-62)^2} = 68.00$$

After calculating the distance with Euclidean Distance for iteration 1 with centroid 1 is

Vol 4-Sep2019
ISSN: 2541-3023

done, then the next step is to calculate with centroid 2, and 3 with the same formula with the calculation above.

c. Data Grouping
Allocate each data to the closest centroid. In reallocating data into each cluster based on a comparison of the distance between the data with the centroid of each cluster, data is explicitly allocated into the cluster that has the distance to the nearest centroid to that data. The following is the result of a comparison of the distance between the data with the centroid of each existing cluster based on the calculation of distance with Euclidean Distance for iterations 1, C (cluster) and M (data).

Table 2. Data Grouping in Iteration-1

CODE	C1	C2	C3	Clustering
M010	45.02	0.00	87.16	2
M020	0.00	45.02	42.19	1
M030	42.00	87.01	4.00	3
M040	20.00	65.02	22.30	1
M050	66.00	21.01	108.09	2
M060	65.00	20.03	107.08	2
M070	42.19	87.16	0.00	3
M080	42.07	87.08	1.52	3
M090	17.00	28.04	59.12	1
M100	25.05	70.06	17.17	3
M110	29.00	16.06	71.11	2
M120	60.00	15.09	102.07	2
M130	30.15	75.12	12.04	3
M011	68.02	23.00	110.14	2
M022	68.01	23.00	110.12	2
M033	27.04	72.05	15.21	3
M044	67.01	22.00	109.11	2
M055	65.00	20.05	107.07	2
M066	42.06	87.07	1.84	3
M077	4.13	41.06	46.10	1
M088	64.00	19.04	106.08	2
M099	68.00	23.01	110.10	2

d. Determination of New Cluster Centers
Determine the position of the new centroid by calculating the average value of the data in the same centroid.

$$C1 = \begin{pmatrix} \frac{3.1+2.81+2.9+2.59}{4}; \frac{2.96+2.81+2.88+2.54}{4}; \frac{2.96+2.81+2.92+2.4}{4}; \\ \frac{2.96+2.81+2.92+2.4}{4}; \frac{62+42+79+66}{4} \end{pmatrix} = (2.85;2.80;2.77;2.77;62.25)$$

$$C2 = \begin{pmatrix} \frac{3.74+3.25+3.13+3.03+2.59+3.73+3.68+3.52+2.91+3.09+3.34}{11}; \\ \frac{3.65+3.21+3.08+2.98+2.82+3.73+3.48+3.49+2.8+2.92+3.27}{11}; \\ \frac{3.61+3.26+3.04+2.88+2.92+3.79+3.47+3.47+2.89+3.01+3.32}{11}; \\ \frac{3.44+3.23+3.06+2.89+2.95+3.77+3.57+3.48+2.94+3.01+3.32}{11}; \\ \frac{107+128+127+91+122+130+130+129+127+126+130}{11} \end{pmatrix} = (3.27;3.22;3.24;3.24;122.45)$$

$$C3 = \begin{pmatrix} \frac{3+1+1.76+2.22+1.5+2.25+1.92}{7}; \frac{3+1+1.76+2.22+1.5+2.25+1.92}{7}; \\ \frac{3+1+1.76+2.22+1.5+2.25+1.92}{7}; \frac{3+1+1.76+2.22+1.5+2.25+1.92}{7}; \\ \frac{20+20+20+37+32+35+20}{7} \end{pmatrix} = (13.65;13.65;13.65;13.65;184)$$

Repeat the 4 steps above until the cluster position does not change again (Iteration-n). Then do a comparison of the results of grouping data in each iteration.

e. Comparison of Iteration Grouping Results
Determine the position of the new centroid by calculating the average value of existing data.

Table 3. Comparison of Data Grouping Each Iteration

CODE	Iteration-1	Iteration-2	Iteration-3
M010	2	2	2
M020	1	1	1
M030	3	3	3
M040	1	3	3
M050	2	2	2
M060	2	2	2
M070	3	3	3
M080	3	3	3
M090	1	1	1
M100	3	3	3
M110	2	1	1
M120	2	2	2
M130	3	3	3
M011	2	2	2
M022	2	2	2
M033	3	3	3
M044	2	2	2
M055	2	2	2
M066	3	3	3
M077	1	1	1
M088	2	2	2
M099	2	2	2

In iteration-1 and iteration-2 there is still a cluster position that is still changing, it is necessary to recalculate the iteration-3. Then do a comparison between iterations-2 and iterations-3. Because the 3rd iteration of the cluster position has not changed / the same as the position of the cluster in the 2nd iteration then the next iteration process does not need to be done, it can be concluded that the iteration process can be stopped at the 3rd iteration with the results :

Cluster 1 member (C1) : {M020, M090, M110, M077} = 4 data

Cluster 2 member (C2) : {M010, M050, M060, M120, M011, M022, M044, M055, M088, M099} = 10 data

Cluster 3 member (C3) : {M030, M040, M070, M080, M100, M130, M033, M066} = 8 data

Table 4. Cluster Data Processing Results

CLUSTER 1 - NUMBER OF MEMBERS = 4 STUDENTS. CONSISTS OF :						
CODE	STUDENTS NAME	GPA SMT 2	GPA SMT 3	GPA SMT 4	GPA SMT 5	CREDIT S TOTAL
M020	TEGUH RIANTO	3.1	2.96	2.96	2.96	62
M090	DENDY NOFRIZAL	2.9	2.88	2.92	2.92	79
M110	MAULANA PRAYOGO PANGESTU	3.03	2.98	2.88	2.89	91
M077	NOPEBRIN DUY PUTRA MANIK	2.59	2.54	2.4	2.4	66
NILAI RATA-RATA		2.91	2.84	2.79	2.79	74.50
NILAI MINIMUM		2.59	2.54	2.40	2.40	62.00
NILAI MAXIMUM		3.10	2.98	2.96	2.96	91.00
CLUSTER 2 - NUMBER OF MEMBERS = 10 STUDENTS. CONSISTS OF :						
CODE	STUDENTS NAME	GPA SMT 2	GPA SMT 3	GPA SMT 4	GPA SMT 5	CREDIT S TOTAL
M010	LUPITA SRI REZEKI	3.74	3.65	3.61	3.44	107
M050	DEBORA KATARINA SIMAMORA	3.25	3.21	3.26	3.23	128

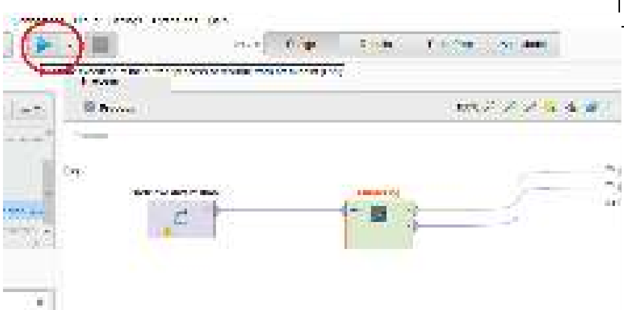
Vol 4-Sep2019
ISSN: 2541-3023

M060	TAUFIK KURRAHMAN	3.13	3.08	3.04	3.06	127
M120	DHEA ANISYA	2.59	2.82	2.92	2.95	122
M011	ADE SAPUTRA	3.73	3.73	3.79	3.77	130
M022	RONI SAFRIYADI	3.68	3.48	3.47	3.57	130
M044	SYAHRUL RAMADHAN	3.52	3.49	3.47	3.48	129
M055	RIRI KUSHENDAR	2.91	2.8	2.89	2.94	127
M088	SYAHRUN NUR	3.09	2.92	3.01	3.01	126
M099	SISKA FERONIKA SIRINGO-RINGO	3.34	3.27	3.32	3.32	130
NILAI RATA-RATA		3.30	3.25	3.28	3.28	125.6
NILAI MINIMUM		2.59	2.80	2.89	2.94	107.0
NILAI MAXIMUM		3.74	3.73	3.79	3.77	130.0

CLUSTER 3 - NUMBER OF MEMBERS = 8 STUDENTS. CONSISTS OF :						
CODE	STUDENTS NAME	GPA SMT 2	GPA SMT 3	GPA SMT 4	GPA SMT 5	CREDIT S TOTAL
M030	SISILIA TRIANI BR. MANULANG	3	3	3	3	20
M040	YUKHRIZUL ALMUFAZRI	2.81	2.81	2.81	2.81	42
M070	ANGGA DANA PRATAMA	1	1	1	1	20
M080	MUHAMMAD FUAD FAHRI	1.76	1.76	1.76	1.76	20
M100	RYAN BAPTISTA HUTASOIT	2.22	2.22	2.22	2.22	37
M130	NUZWARDANA FIKRI	1.5	1.5	1.5	1.5	32
M033	RAMADHAN SAPUTRA	2.25	2.25	2.25	2.25	35
M066	REYNAL SATRIO	1.92	1.92	1.92	1.92	20
NILAI RATA-RATA		2.06	2.06	2.06	2.06	28.25
NILAI MINIMUM		1.00	1.00	1.00	1.00	20.00
NILAI MAXIMUM		3.00	3.00	3.00	3.00	42.00

From the table above it can be concluded that based on the average value of each attribute in each cluster can be seen a comparison of the average number, then the group graduation rate of students who have the potential to graduate with a decline in study period is in Cluster 1 with an average grade point average Semester 2 to Semester 5 and total credits = 2.91; 2.84; 2.79; 2.79; 74.50 and consists of 4 data. For groups the graduation rate of students who have the potential to graduate on time is in Cluster 2 with the average grade of GPA of Semester 2 to Semester 5 and total credits = 3.30; 3.25; 3.28; 3.28; 125.6 and consists of 10 members. For groups of student graduation rates that have the potential to drop out can be found in Cluster 3 with the average scores of Semester 2 to Semester 5 and credits total = 2.06; 2.06; 2.06; 2.06; 28.25 and consists of 8 members.

2. Testing with the Rapidminer Tool
The testing process is a very important process to determine the extent to which the design of Data Mining can be tested using a software.



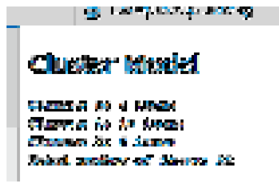
Picture 2. Testing with the Rapidminer Tool

After entering the data and operators needed to run the testing process, it will display some forms of display produced by Rapidminer on the K-Means process that has been carried out previously. The results are as follows :

a. *ExampleSet*

Picture 3. Display of Cluster Results (Data View)

b. *Cluster Model* (Clustering)



Picture 4. Display of Cluster Results (Description)

3. Comparison of Calculation Results
Based on the results of testing sample data using Rapidminer Tools, it can be concluded that the results of manual calculations and calculations using Rapidminer Tools, the results are the same. Members of each cluster in manual calculations are the same as cluster members in the test results.

Table 5. Comparison of Calculations Results

Manual Calculations		
Cluster	Members	Total
1	M020, M090, M110, M077	4
2	M010, M050, M060, M120, M011, M022, M044, M055, M088, M099	10
3	M030, M040, M070, M080, M100, M130, M033, M066	8

RapidminerTools		
Cluster	Members	Total
0	M020, M090, M110, M077	4
1	M010, M050, M060, M120, M011, M022, M044, M055, M088, M099	10
2	M030, M040, M070, M080, M100, M130, M033, M066	8

From the tests conducted above, it can be concluded that there are similarities between the results of manual processing and testing using Rapidminer Tools. The only difference lies in the number of clusters produced and the number of cluster members. Based on testing the entire data, cluster 0 is the graduation rate of students who have the potential to graduate with a decrease in study period, cluster 1 is the graduation rate of students who have the potential to graduate on time and cluster 2 is a group of students who have the potential to break up. This can be read on the centroid of each cluster in the attribute of the average value.

IV. Conclusion

In this study, the grouping of student graduation results using the K-Means Clustering method from the data analysis process grouping student graduation rates up to semester 6 (VI) using data on GPA and total credit values taken previously, succeeded in producing 3 clusters namely, for groups of students who had the potential to graduate on time has the potential to pass with a decrease in study period and potentially drop out. The use of Rapidminer tools can be done to build a pattern of grouping the results of student graduation rates with data from the Faculty of Computer Science at Lancang Kuning University and as a comparison with manual calculations. In this study, this can also help study programs to recognize the situation of their students and be able to anticipate students who drop out of lecture, to improve student achievement, to improve the curriculum, improve the process of teaching and learning activities and many others benefits that can be obtained from data mining.

References

[1] AKBAR, R. (2015). *Penerapan Data Mining dengan Menggunakan Metode Clustering K-Mean Untuk Mengukur Tingkat Ketepatan Kelulusan Mahasiswa Program Teknik Informatika S1 Fakultas Ilmu Komputer Universitas Dian Nuswantoro*. *Dokumen Karya Ilmiah*. Retrieved from http://dinus.ac.id/

[2] Fan, Z., & Sun, Y. (2017). *Clustering of College Students Based on Improved K-Means Algorithm*. *Proceedings - 2016 International Computer Symposium, ICS 2016*, *28*(6), 676–679. https://doi.org/10.1109/ICS.2016.0139

[3] Hermawati, F.A. (2013). *Data Mining*. Yogyakarta: Penerbit Andi.

[4] Asril, Elvira, Fana Wiza, and Yogi Yunefri. "Analisis Data Lulusan dengan Data Mining untuk Mendukung Strategi Promosi Universitas Lancang Kuning." Digital Zone: Jurnal Teknologi Informasi dan Komunikasi 6.2 (2015): 24-32..

[5] Jannah, A. R., Arifianto, D., & Kom, M. (2015). *Penerapan Metode Clustering dengan Algoritma K-Means untuk Prediksi Kelulusan Mahasiswa Jurusan Teknik Informatika di Universitas Muhammadiyah Jember*. (1210651237).

[6] Kartini, D. (2016). *Rancang Bangun Aplikasi K-Means untuk Klasifikasi Kelulusan Siswa Sekolah Kepolisian Negara Daerah Kalimantan Selatan*. *ProTekInfo*, *3*(1), 14–21.

[7] Metisen, B. M., & Sari, H. L. (2015). *Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila*. *11*(2), 110–118.

[8] Muslihudin, M., & Arumita, A. W. (2016). *Pembuatan Model Penilaian Proses Belajar Mengajar Perguruan Tinggi Menggunakan Fuzzy Simple Additive Weighting (SAW) (Sudi : STMIK Pringsewu)*. 6–7.

[9] Noercholis, A., & Hakim, M. L. (2016). *Sistem Pendukung Keputusan Penilaian Kinerja Guru PNS Di SMKN Sukorejo 1 Dengan Menggunakan Metode Fuzzy AHP*. *10*(2), 65–72.

[10] Novita Sari, V., Yupianti, Y., & Maharani, D. (2019). *Penerapan Metode K-Means Clustering Dalam Menentukan Predikat Kelulusan Mahasiswa Untuk Menganalisa Kualitas Lulusan*. *Jurteksi*, *4*(2), 133–140. https://doi.org/10.33330/jurteksi.v4i2.53

[11] Reza, A., Syukur, A., & Soeleman, M. A. (2017). *Penentuan Jurusan Siswa Sekolah Menengah Atas Disesuaikan Dengan Minat Siswa Menggunakan Algoritma*, *13*, 57–68.

[12] Rosmini, R., Fadlil, A., & Sunardi, S. (2018). *Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah. It Journal Research and Development*, *3*(1), 22. https://doi.org/10.25299/itjrd.2018.vol3(1).1773

[13] Siska, S. T. (2016). *Analisa Dan Penerapan Data Mining Untuk Menentukan Kubikasi Air Terjual Berdasarkan Pengelompokan Pelanggan Menggunakan Algoritma K-Means Clustering*. *Jurnal Teknologi Informasi & Pendidikan*, *9*(1), 48–55.

[14] Syarif, M. (2011). *Aplikasi Data Mining Untuk Mengukur Tingkat Kelulusan Mahasiswa Dengan Metode Apriori Dan K-Mean Clustering*. *2*(2), 283–292.

[15] Widaningrum, I. (2015). *Sistem Pendukung Keputusan Evaluasi Kinerja Dosen Menggunakan Analitycal Hierarchy Process (Ahp) Dan Fuzzy Analitycal Hierarchy Process (Fuzzy-Ahp)*. 6–8.