



## Application of K-Means Clustering on the Performance Evaluation of Lecturers Based on Student Questionnaire

Mutiara Mala Khairunnisa<sup>1</sup>, Agung Triayudi<sup>2</sup>, Endah Tri Esti Handayani<sup>3</sup>

<sup>1</sup>Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Jalan Sawo  
Manila, Jakarta Selatan, 12520, Indonesia

Email: <sup>1</sup>mutimulanis46@gmail.com, <sup>2</sup>agungtriayudi@civitas.unas.ac.id, <sup>3</sup>endahtriesti@civitas.unas.ac.id

### ARTICLE INFO

### ABSTRACT

#### Article history:

Received: 04/04/2020

Revised: 20/04/2020

Accepted: 30/05/2020

#### Keywords:

Data Mining,

K-Means,

Clustering,

Lecturer Assessment

In the world of lectures or universities, conducting performance appraisals needs to be applied to assess the performance of lecturers. The basis used in the assessment uses the basic tri dharma of tertiary institutions. In assessing performance, data is needed related to student satisfaction with the lecturer. The data used in this study are 152 National University student satisfaction data for all lecturers of the Faculty of Communication and Information Technology at the National University, amounting to 28 lecturers. In collecting data using a questionnaire available on the National University Lecture website. There are 10 questions that need to be filled out by all active students in the 2016/2017 school year odd and even semester. Data will be processed by clustering the performance of lecturers in the cluster is good, good enough or less.

Copyright © 2020 Jurnal Mantik.  
All rights reserved.

## 1. Introduction

As one component in evaluating lecturer performance is an assessment given by students to each lecturer who is in charge of the course they are taking. This assessment is given in the form of a questionnaire about the ability of a lecturer to deliver lecture material, both theory in the classroom and practicum in the laboratory. This assessment is given by students before students want to fill KRS in the next semester and can provide an evaluation of lecturers who teach the subjects they teach. Hopefully, with this assessment, input will be obtained regarding the ability of a lecturer to deliver course material as a form of performance appraisal component of a lecturer. Based on the description above, then the problem that arises is how to display the results of lecturer performance based on the assessment of students from learning outcomes using clustering techniques. As for some of the limitations carried out in this study include: a. Active Students in Odd Semester and Even Semester 2016/2017 Academic Year b. Lecturers who take courses in Odd Semester and Even Semester Academic Year 2016/2017 c. The performance appraisal grouping technique used is the clustering technique with the K-Means method d. The assessment weights which initially used abbreviated codes such as SB, B, C, KB, SKB were changed to score 1 to 5. Active Students in Odd Semester and Even Semester 2016/2017 Academic Year b. Lecturers who take courses in Odd Semester and Even Semester Academic Year 2016/2017 c. The performance appraisal grouping technique used is the clustering technique with the K-Means method d. The assessment weights which initially used abbreviated codes such as SB, B, C, KB, SKB were changed to score 1 to 5. Active Students in Odd Semester and Even Semester 2016/2017 Academic Year b. Lecturers who take courses in Odd Semester and Even Semester Academic Year 2016/2017 c. The performance appraisal grouping technique used is the clustering technique with the K-Means method d. The assessment weights which initially used abbreviated codes such as SB, B, C, KB, SKB were changed to score 1 to 5.

The aim to be achieved is to know the extent of the performance of a lecturer who is assessed by students when providing a learning process for the courses they teach during one academic year, which is odd semester and even semester in the 2016/2017 Academic Year. While the benefits obtained are knowing the level of performance evaluation of lecturers in teaching is good, enough or even still lacking, and knowing whether the



assessment tends to increase or decrease in several semesters, so that appropriate evaluation steps can be taken to improve it in the future.

Research related to this research such as:

- a. Sahat Sonang, the research carried out implements the k means algorithm method with the Data Mining technique to get the most accurate test results in processing information from student data as a basis for determining in evaluating lecturer performance [].
- b. Eko Nur Wahyudi, In the research conducted, that is the analysis of the final report to find out the extent of the performance of a lecturer which is assessed by students when giving the learning process to the courses they teach. [].

Basically the use of algorithms in the clustering process depends on the existing data and the conclusions to be achieved. For this reason, the K-Means algorithm is used which makes the following rules:

- 1) Cluster Amount needs to be inputted
- 2) Only has numeric attribute.

K-means is a non-hierarchical (grouped) data grouping method that attempts to partition data into the form of two or more groups that have been predetermined based on the means (means) closest to the centroid [6]. The k-means method is used in several studies because it is simple and able to group large amounts of data with a relatively short computation [7].

The k-means algorithm is the simplest and most common clustering method. This is because K-means has the ability to group large amounts of data with relatively fast and efficient computing time.

## 2. Research methods

In this study there are several work sequences that must be followed, this work order is the steps that must be passed and carried out in accordance with the main problems so that they do not deviate from the specified problem boundaries and this research goes well. The steps that must be done starting from the problem found to get a solution of the problem. Following is the framework of the stages of research used in conducting this research:

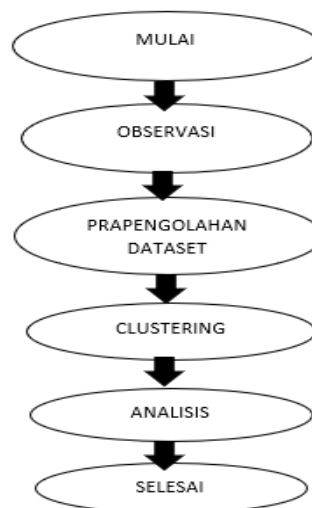


Fig 1. Research Outline Framework

### A. Observation

At this stage, data collection was carried out on each lecturer who was actively teaching, who would be grouped through filling out questionnaires by students. After getting the data, the results of the data are inputted to the application.

### B. Pre-processing Dataset

At this stage the dataset to be processed is 10 indicator indicators which are divided into 3 aspects, to determine the performance of lecturers. The 10 indicators are indicators selected from the Academic Administration Handbook for Lecturers because they are in accordance with the material that will be used as a case. In this research, what will be used as a case is the performance evaluation of lecturers who are actively

taking courses at the Faculty of Communication and Information Technology from the National University. Before indicator data is processed into clusters, it is necessary to do data cleaning and data transformation first. Indicator data is changed in numerical or numeric form, so that we can get the means (means) on each indicator.

TABLE 2  
EXAMPLE OF DATASETS THAT ARE READY TO BE PROCESSED INTO CLUSTERS

Name of Lecturer Initials	Preparation	Learning Process	Ability
IF	10.56	17.61	7.04
UD	10.53	17.55	7.02
SA	10.62	17.7	7.08
F	10.69	17.81	7.12
IW	10.48	17.47	6.99

The following is an overview of flowcharts based on the stages of the tables above. Shown in Figure 2 below.

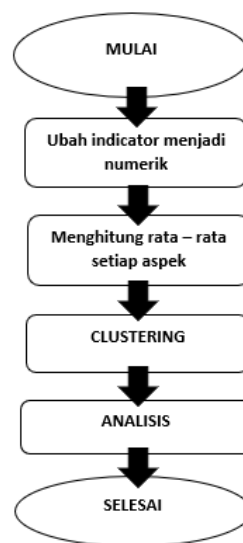


Fig 2. Clustering flowchart

C. Clustering

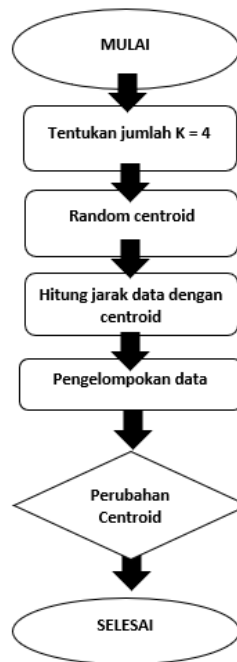
Clustering has been using the k-means method a lot, as explained in the introduction above. K-means grouping data based on the calculation of the closest distance between data and centroid. There are several ways that can be used in calculating the distance between data and centroids, namely Euclidean Distance, Manhattan Distance, and Minkowski Distance. In this research, the author will use the Euclidean Distance calculation because it provides the best calculation results [5]. The following is the formula for calculating Euclidean Distance:

$$Dist(xy) = \sqrt{\sum_{k=1}^m (X_{ik} - X_{jk})^2} \dots\dots\dots (1)$$

Note: Dist (xy) = Distance between x and y  
 $X_{ik} - X_{jk}$  = Difference between data and centroid

Clustering using the k-means method is generally illustrated in Figure 3 with the basic algorithm as follows:





**Fig 3.** Clustering Algorithm Flowchart

### 3. Results and Discussion

#### A. The calculation results

In this study the authors will display some test data results that will be used to compare calculations with the RapidMiner tools with manual calculations. The following table is the result of test data.

**TABLE 3**  
 DATA RESULTS

Name of Lecturer Initials	Preparation	Learning Process	Ability
IF	10.56	17.61	7.04
UD	10.53	17.55	7.02
SA	10.62	17.7	7.08
F	10.69	17.81	7.12
IW	10.48	17.47	6.99

Based on table 3 of the test data results, the highest value of the preparation aspects is 10.9 with the name of the Lecturer Initials namely IDS, B, and AR, for the lowest value of 10.3 with the Initial Lecturers namely N, RTK, W, RS, and EM.

For the aspect of the learning process the highest value of 18.2 with the initials of the lecturer name is B, for the lowest value of 1 with 4 initials of lecturers namely N, RTK, RS, and EM.

For the aspect of ability the highest score was 7.2, with the initials of IDS, B, and AR lecturers, for the lowest value of 6.8 with 2 initial lecturer names namely RTK and EM.

Then determine the initial centroid randomly, table 4 below designates the initial centroid.

**TABLE 4**  
 INITIAL CENTROIDS DATA

Centroids	x	y	z
1	10.67	17.79	7.11
2	10.53	17.55	7.02
3	10.47	17.46	6.98
4	10.62	17.7	7.08

Based on table 3 of the test data results, the highest value of the preparation aspects is 10.9 with the name of the Lecturer Initials namely IDS, B, and AR, for the lowest value of 10.3 with the Initial Lecturers namely N, RTK, W, RS, and EM.

For the aspect of the learning process the highest value of 18.2 with the initials of the lecturer name is B, for the lowest value of 1 with 4 initials of lecturers namely N, RTK, RS, and EM.

For the aspect of ability the highest score was 7.2, with the initials of IDS, B, and AR lecturers, for the lowest value of 6.8 with 2 initial lecturer names namely RTK and EM.

Then determine the initial centroid randomly, table 4 below designates the initial centroid.

**TABLE 5**  
THE RESULTS OF THE FIRST ITERATION MANUAL CALCULATION BETWEEN THE DISTANCE OF DATA AND CENTROID

Name of Lecturer Initials	Preparation	Learning Process	Ability	Distance 1	Distance 2	Distance 3	Distance 4	c
IF	10.56	17.61	7.04	0.22	0.07	0.38	0.12	2
UD	10.53	17.55	7.02	0.29	0.00	0.32	0.18	2
SA	10.62	17.7	7.08	0.11	0.18	0.30	0.00	4
F	10.69	17.81	7.12	0.03	0.32	0.44	0.14	1
IW	10.48	17.47	6.99	0.39	0.10	0.02	0.28	3

Based on the results of the first iteration using the application of the author of lecturers in cluster 1 numbered 9, lecturers in cluster 2 numbered 6, lecturers in cluster 3 totaled 8, and lecturers in cluster 4 totaled 5.

In calculations using this author's application the convergent centroid is found in the 7th iteration with the final cluster and centroid results shown in table 6.

**TABLE 6**  
THE RESULTS OF CALCULATING ITERATION 7 MANUALLY

Name of Lecturer Initials	Preparation	Learning Process	Ability	Distance 1	Distance 2	Distance 3	Distance 4	c
IF	10.56	17.61	7.04	0.70	0.05	0.45	0.21	2
UD	10.53	17.55	7.02	0.77	0.02	0.38	0.28	2
SA	10.62	17.7	7.08	0.59	0.16	0.56	0.10	4
F	10.69	17.81	7.12	0.45	0.30	0.70	0.04	4
IW	10.48	17.47	6.99	0.87	0.12	0.28	0.38	2

Based on the results of iteration 7 manually, there are 3 lecturers in cluster 1, lecturers in cluster 2 are 12, lecturers in cluster 3 are 5 and lecturers in cluster 4 are 8. The results of calculations using the rapidminer application are shown in table 7.

**TABLE 7**  
THE RESULTS OF CALCULATIONS USING RAPIDMINER VERSION 9.0.3

id	the cluster	16311 ...	16311 ...	16311 ...	16311 ...	16311 ...
1	Cluster_0	34	40	33	45	35
2	Cluster_0	33	45	35	45	44
3	Cluster_3	36	34	34	38	33
4	Cluster_2	32	31	34	35	44
5	Cluster_0	33	30	27	29	25

The results of rapidminer calculation version 9.0.3 were determined initially by the centroid of the application, namely, cluster\_0 = cluster 2, cluster\_1 = cluster 3, cluster\_2 = cluster 1, cluster\_3 = cluster 4. To see the results of the calculations using the rapidminer application, in figure 4 shows visualization of that data.



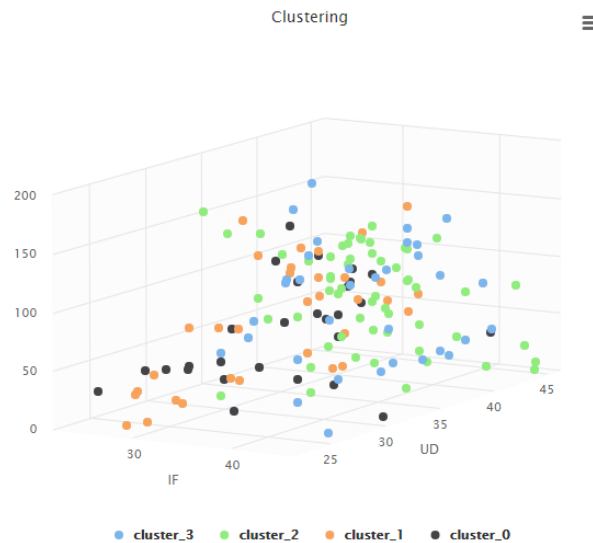


Fig 4. Visualization of the results of calculations using rapid miners.

Information :

X-axis = aspect of preparation

Y-axis = learning process

Z axis = ability

#### 4. Conclusion

From the results of an experiment comparing 2 calculations into the k-means clustering method to map aspects of lecturer performance indicators, it can be concluded that, manual calculation has clustering results that are close to the results of calculations using the rapid miner application. The calculation using rapidminer version 9.0.3 has different clustering results because the centroid value in the rapidminer calculation is not determined by the author but is determined randomly by the rapidminer application. So that the results of clustering calculated manually by the author can be used as a reference for conducting regular evaluations to improve the performance of lecturers.

#### 5. Reference

- [1] Gama, S. (2014). Clustering Portal Jurnal Internasional Untuk Rekomendasi Publikasi Berdasarkan Kualitas Cluster Menggunakan Kernel K-Means (Doctoral dissertation, Universitas Brawijaya).
- [2] Hambali, H. A., Abdullah, S., Lailee, S., Jamil, N., & Harun, H. (2016). Intelligent segmentation of fruit images using an integrated thresholding and adaptive K-means method (TsNKM). *Jurnal Teknologi*, 78(6-5), 13-20.
- [3] Hidayatullah, K. H. (2014). Analisis Klaster Untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 2(1).
- [4] Nurzahputra, A., Muslim, M. A., & Khusniati, M. (2017). Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa. *Techno. Com*, 16(1), 17-24.
- [5] Purba, W., Tamba, S., & Saragih, J. (2018, April). The effect of mining data k-means clustering toward students profile model drop out potential. In *Journal of Physics: Conference Series* (Vol. 1007, No. 1, p. 012049). IOP Publishing.
- [6] Ramdhani, F., Hoyyi, A., & Mukid, M. A. (2015). Pengelompokan Provinsi di Indonesia berdasarkan Karakteristik Kesejahteraan Rakyat menggunakan Metode K-Means Cluster. *Jurnal Gaussian*, 4(4), 875-884.
- [7] Santoso, B., Cholissodin, I., & Setiawan, B. D. (2017). Optimasi K-Means untuk Clustering Kinerja Akademik Dosen Menggunakan Algoritme Genetika. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- [8] Sonang, S. (2018). IMPLEMENTASI K-MEANS DALAM EVALUASI KINERJA DOSEN POLITEKNIK BISNIS INDONESIA. *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, 1(1), 32-40.
- [9] Triayudi, A., & Nathasia, N. D. (2020). Analysis of Factors Affecting Student Graduation Using the K-Means clustering Method. *Jurnal Mantik*, 3(4, Feb), 135-143.
- [10] Tüzüner, V. L., & Yüksel, C. A. (2009). Segmenting potential employees according to firm's employer attractiveness dimensions in the employer branding concept. *Journal of Academic Research in Economics*, 1(1), 46-61.

- [11] Wahyudi, E. N., & Jananto, A. (2013). Final Report Penilaian Kinerja Dosen oleh Mahasiswa pada Satu Periode Tahun Akademik menggunakan Teknik Klustering (Studi Kasus: Universitas Stikubank Semarang). *Dinamik*, 18(2).

