

The Level of Student Satisfaction with the Online Learning Process During a Pandemic Using the K-means Algorithm

Talitha Syahla Janiar Arifin^{1*}, Nakia Natassa², Dinda Khoirunnisa³, Retno Hendrowati⁴

^{1,2,3,4}*Informatics Engineering Program, Universitas Paramadina, Indonesia*

^{1,2,3}[talitha.arifin*, nakia.natassa, dinda.khoirunnisa]@students.paramadina.ac.id

⁴retno.hendrowati@paramadina.ac.id

Received: 2021-06-26; Accepted: 2021-07-26; Published: 2021-07-31

Abstract— The number of cases of Covid-19 in this pandemic era is increasing and getting out of control every day. This triggers the Indonesian government to set policies on schools with online learning methods. Of course, online learning cannot ensure that it runs smoothly in all circles because several factors hinder the learning process. The difficulty of the internet network, limited quotas, unfamiliarity with the use of learning media, and an unsupportive environment for conducting online learning are the obstacles to ineffective online learning. The purpose of this study was to determine the level of satisfaction with online learning during the pandemic. This study uses quantitative research methods with a descriptive approach. Quantitative research methods will be processed into data mining using the K-Means Clustering Algorithm. The clustering process is carried out to get the results of clustering the level of student satisfaction. The dataset was obtained from the results of the questionnaire by submitting statements of satisfaction and dissatisfaction. The cluster type is based on high, medium, and low class. The test results obtained a value with the final iteration, namely the level of satisfied statements is categorized as high with a value of 11.79 compared to the dissatisfied statement, which is categorized as moderate with a value of 7.46. In contrast, for the low category level, there is no value of 0.00 cluster results state that the category is satisfied with online learning with a value of 9.33.

Keywords— Coronavirus, Data Mining, K-Means Clustering, Quantitative Methods

I. INTRODUCTION

Coronavirus (COVID 19) is a virus that is spreading around the world. This virus first appeared in Wuhan, China, at the end of 2019 [1]. This virus is growing so fast that it has spread worldwide with increasing and uncontrolled cases [2]. On March 2, 2020, the Indonesian government announced two positive COVID-19 patients; Indonesia was designated as one of the countries exposed to COVID-19 [3]. The government has also established a Social Distancing policy to break the chain of transmission of the Covid-19 virus by maintaining distance and limiting human interaction. The implementation of this policy is carried out by encouraging all people to stay at home, do all activities at home, and have educational activities carried out online or online [4].

Since March 16, 2020, Paramadina University has implemented social distancing and work from home (WFH) policies for students, lecturers, and staff. With this policy, the face-to-face learning process is replaced with online learning methods that can connect students and teachers who are far apart but can communicate with each other or interact with technology and internet support. Online learning uses digital platforms such as zoom, google meet, e-learning, WhatsApp, and others to assist in implementing online learning and a means of communication between teachers and students [5]. Of course, this situation is an obstacle experienced by teachers and students, such as the slow internet network, causing problems in submitting assignments and disconnection while lectures are in progress [6].

During the covid-19 pandemic, students conduct online or online lectures that have various positive and negative impacts on their application. The positive impact of online lectures is growing independence in learning. Students are required to be more active in exploring the material that the lecturer has given to make it easier for them to understand the material [7]. Online lectures also have time flexibility because virtual classes can be accessed anytime and anywhere [8]. The negative impact of online lectures is that excessive use of gadgets and computers can lead to serious health problems [7]. In addition, many students complain of increasing assignments during online lectures, resulting in fatigue and easy stress [9]. The problem discussed in this paper's discussion is to determine the level of student satisfaction with online learning using the K-Means algorithm. This algorithm will make class groupings based on their similarity [10]. K-Means was chosen because it is a very reliable method [11].

Previous research that uses data mining to create clusters in dataset analysis, including the Fuzzy C-Means Algorithm [12]–[14], Multifactor Evaluation Process (MFEP) [15], Comparison of Single Linkage, Complete Linkage and Average Linkage Clustering Methods [16] [14], and K-Means Algorithm [14] [17] by Using Particle Swarm [18]. However, no one has discussed the paper to make groupings know the level of student satisfaction with online learning. Policymakers can use the results of this study to evaluate online learning during this pandemic.

II. RESEARCH METHODOLOGY

This study uses quantitative research methods with a descriptive approach. Quantitative research methods will be processed into a form of data mining. Data mining itself is a series of processes obtained from data sets manually. One of the techniques contained in data mining is the K-Means Clustering Algorithm. A descriptive approach is a research approach that solves a problem by obtaining decisions based on data [19].

A. Online Learning

The online learning model can also be referred to as a blended learning (BL) model. Blended learning is a learning approach that combines the advantages of offline learning and e-learning [20]. Online learning is a learning method that is tried by using various digital technology features such as smartphones, computers, applications, or websites based on internet networks. Digital platforms that are often used for online learning are WhatsApp, google meet, zoom, e-learning, google classroom, video conferencing, live chat, and various other digital platforms [5]. The application of online learning immediately requires utilizing technology and using the internet network [21].

This online learning certainly cannot ensure that it runs smoothly in all circles because not all students have supporting facilities such as laptops, adequate network connections, economic conditions in the absence of quotas, and an unsupportive environment for online learning. These constraints make online learning ineffective [22]. In addition, online learning spends more time in front of a smartphone or computer screen. The use of these devices for a very long duration can cause physical problems, especially health. In the results of Mustakim's research through a survey procedure, while online learning, students face many physical complaints such as headaches, eye fatigue, difficulty resting, often drowsy, and other physical complaints [23].

During online lectures, students spend a lot of time in front of a gadget or computer screen. This can lead to health problems such as headaches, eye fatigue, and other ailments [23]. Less effective learning occurs because of the slow network, lack of facilities, inadequate environment, lack of understanding in the use of technology, difficulty in understanding the material because it is easier for students to understand the material face to face than online. This can lead to ineffective learning while online [5]. Various academic demands that students must complete cause academic stress [24]. Academic stress is the pressure that arises on students due to competition and academic demands. Academic stress is caused by academic stress. Academic stressors start from the learning process, such as pressure to get good grades, increased assignments, length of study, low grades/performance, and fear of taking exams [25]. The inability of students to adapt to these conditions causes students to experience stress [23].

B. Datasets

The subject of this study is a dataset of students from Paramadina University's Faculty of Engineering batches of

2019 and 2020, with a total of about 200 students. However, only 10% of the entire random sample was used in this study. Questionnaires, observation, and study theories were utilized to collect data in this study. The author's data-collecting period runs from April through July 2021. The variables in this study are declarations of extreme satisfaction and dissatisfaction. As a result, for numerous categories of questions on the questionnaire, it becomes a measure of satisfaction.

In this study, the data collected from the outcomes of the questionnaire distribution by 20 random sample respondents will be processed first, followed by clustering. At the clustering step, the processed data will be processed.

The information will be analyzed by grouping student satisfaction with the degree of the online learning process. High satisfaction level cluster, medium satisfaction level cluster, and low satisfaction level cluster are the three clusters. In this study, data is processed using a spreadsheet/excel that is automatically linked to the questionnaire data.

C. K-Means Clustering Algorithm

Figure 1 depicts the step of categorizing and solving a collection of data from groups into numerous classes based on criteria. During this step of data grouping, the Euclidean Distance theory is applied. The clustering flowchart can be shown in Figure 1.1 to aid in getting a series of computations to figure out the ultimate result [26].

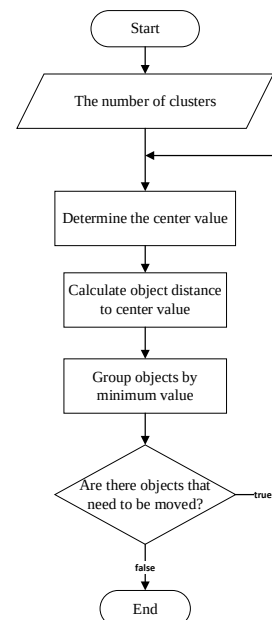


Figure 1. Clustering stages using the K-Means method

Data mining decision-making is a collection of procedures that are manually obtained from data sets. The K-Means Clustering Algorithm is one of the data mining strategies. This approach employs the Clustering principle, which involves categorizing objects into class groups based on their resemblance [10]. Overall, the K-Means Clustering Algorithm is a non-hierarchical clustering approach that divides sample

data into one or more groups based on similar features [27]. The K-Means algorithm classifies data by detecting the significant categories of each cluster and searches for group members within each cluster [28].

The K-Means Algorithm is calculated by first identifying the data to be grouped. The k variable, where n is the number of data group members and m is the number of variables. Because the K-Means Algorithm is directionless in theory, each cluster will be selected independently at first. As a result, the distance between each data and each primary category of the cluster is established in order to become a fixed data. The Euclidean Distance concept is used to determine the distance between the data and the cluster's center. The concept is derived from the data's value by comparing the data's similarity to other data using the Euclidean Distance equation, which is based on Equation (1).

$$de = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (1)$$

x, y variable is object coordinate, s, t variable is centroid coordinate, and i variable is the number of objects. After determining the data distance, the next step is to determine the center of the cluster by finding the average of each cluster member using Equation (2).

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

where the variable is the average centroid in the i variable is the cluster for the j variable, is the number of the i variable is Cluster members, the i, k variable is the index of the cluster, the j variable is the variable index, and the X_{kj} variable is the k data value and j variable in the cluster.

The cluster grouping process will be completed if the cluster members do not move clusters. But if the cluster members are still moving, it must be recalculated from the first method, which calculates the data distance. Then calculate the cluster average to determine the cluster center. This method is still carried out until the cluster members do not experience displacement [17].

III. RESULT AND DISCUSSION

The trial's findings were based on a dataset created from a questionnaire filled out by students at Paramadina University's Faculty of Engineering. The number of data samples is 20, the number of clusters is 3, and the number of statements or characteristics is 2 with the categories Satisfied and Dissatisfied. There are 20 respondents who become the sample data to establish the degree of student satisfaction.

The starting centroid is determined at random in the first iteration till the n th data. The data utilized is from the datasets in Table I with $C1 = (10,0)$, $C2 = (5,5)$, and $C3 = (10,0)$. (1,9). Based on the results of the Equation computation, the nearest centroid to the nearest centroid was calculated sample data is (9,1). (3).

$$C1 = \sqrt{(9 - 10)^2 + (1 - 0)^2} = 1$$

$$C2 = \sqrt{(9 - 5)^2 + (1 - 5)^2} = 5,656854 \quad (3)$$

$$C3 = \sqrt{(9 - 1)^2 + (1 - 9)^2} = 11,31371$$

TABLE I
ITERATION DISTANCE CALCULATION RESULTS

Sample Data	Statement		C1	C2	C3	Nearest Distance
	Satisfied	Not satisfied				
I	4	6	7,54	1,03	00	1,03
II	6	4	4,71	1,80	00	1,80
III	1	9	11,79	5,27	00	5,27
IV	8	2	1,89	4,63	00	1,89
V	10	0	0,94	7,46	00	0,94
VI	9	1	0,47	6,04	00	0,47
VII	10	0	0,94	7,46	00	0,94
VIII	10	0	0,94	7,46	00	0,94
IX	6	4	4,71	1,80	00	1,80
X	5	5	6,13	0,39	00	0,39
XI	5	5	6,13	0,39	00	0,39
XII	7	3	3,30	3,21	00	3,21
XIII	8	2	1,89	4,63	00	1,89
XIV	7	3	3,30	3,21	00	3,21
XV	9	1	0,47	6,04	00	0,47
XVI	4	6	7,54	1,03	00	1,03
XVII	10	0	0,94	7,46	00	0,94
XVIII	4	6	7,54	1,03	00	1,03
XIX	10	0	0,94	7,46	00	0,94
XX	3	7	8,96	2,44	00	2,44

The data has been categorized with the findings in Table II based on Table I above.

TABEL II
ITERATION DATA GROUPING

Cluster	Sample Data
C1	IV, V, VI, VII, VIII, XIII, XV, XVII, XIX
C2	I, II, III, IX, X, XI, XII, XIV, XVI, XVIII, XX

The value of the first iteration did not change, the value of the third iteration was the same. The iteration process then comes to a halt until the final result has three clusters. During the covid-19 epidemic, studies on student satisfaction with online learning produced the graph in Figure 2.

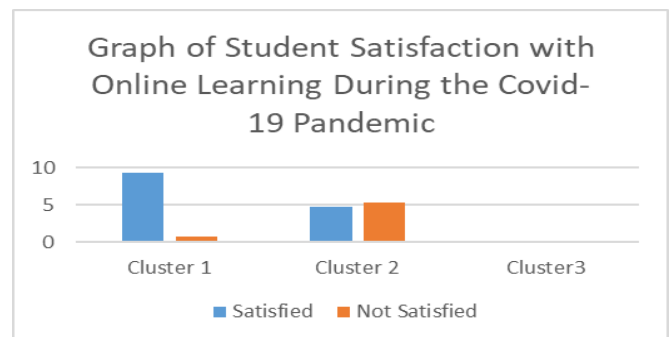


Figure 2. graph of K-means method results for clustering

IV. CONCLUSION

Based on pleased and unsatisfied with online learning, the K-Means algorithm may be used to generate clusters in evaluating the level of satisfaction with the online learning

process. There are three clusters of data handled in the online learning process. The satisfied category with online learning was categorized as "high" with a value of 9.33, while the unsatisfied category was categorized as "low" with a value of 0.67 in the first cluster. The category of being happy with online learning is classed as "low" with a value of 4.73, while the category of being dissatisfied is rated as "high" with a value of 5.27 in the second cluster. The happy or unsatisfied category has no value, according to the third cluster. These findings suggest that online learning is satisfactory at Paramadina University's Faculty of Engineering.

REFERENCE

- [1] A. Susilo *et al.*, "Coronavirus Disease 2019: Tinjauan Literatur Terkini," *J. Penyakit Dalam Indones.*, vol. 7, no. 1, p. 45, 2020.
- [2] R. N. Putri, "Indonesia dalam Menghadapi Pandemi Covid-19," *J. Ilm. Univ. Batanghari Jambi*, vol. 20, no. 2, p. 705, 2020.
- [3] A. Harapani, "Pengaruh Kuliah Daring Saat Pandemi Covid-19 Terhadap Kemampuan Mahasiswa," 2020.
- [4] R. D. Buana, "Analisis Perilaku Masyarakat Indonesia dalam Menghadapi Pandemi Covid-19 dan Kiat Menjaga Kesejahteraan Jiwa," *Sos. dan Budaya, Fak. Syariah dan Huk. Univ. Islam Negeri Syarif Hidayatullah Jakarta*, vol. 53, no. 9, pp. 1689–1699, 2017.
- [5] Y. Fitriyani, I. Fauzi, and M. Z. Sari, "Motivasi Belajar Mahasiswa Pada Pembelajaran Daring Selama Pandemi Covid-19," *Profesi Pendidik. Dasar*, vol. 7, no. 1, pp. 121–132, 2020.
- [6] A. Hutauruk and R. Sidabutar, "Kendala pembelajaran daring selama masa pandemi di kalangan mahasiswa pendidikan matematika: Kajian kualitatif deskriptif," *J. Math. Educ. Appl.*, vol. 02, no. 01, pp. 45–51, 2020.
- [7] D. A. Kusuma, "Dampak Penerapan Pembelajaran Daring Terhadap Kemandirian Belajar (Self-Regulated Learning) Mahasiswa Pada Mata Kuliah Geometri Selama Pembelajaran Jarak Jauh Di Masa Pandemi Covid-19," *Teorema Teor. dan Ris. Mat.*, vol. 5, no. 2, p. 169, 2020.
- [8] A. Sadikin and A. Hamidah, "Pembelajaran Daring di Tengah Wabah Covid-19," *Biodik*, vol. 6, no. 2, pp. 109–119, 2020.
- [9] M. Barseli and I. Idfil, "Konsep Stres Akademik Siswa," *J. Konseling dan Pendidik.*, vol. 5, no. 3, p. 143, 2017.
- [10] R. Novita, "Teknik Data Mining : Algoritma C4.5," *Ilmu Komputer.com*, pp. 1–12, 2016.
- [11] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, no. November, pp. 295–316, 2020.
- [12] A. F. Wardatul Maghfiroh, Nurissaidah Ulinnuha, "Penerapan Fuzzy C-Means dalam Mengelompokkan Kabupaten/Kota Berdasarkan Fasilitas Pelayanan Kesehatan Di Jawa Timur," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 4, no. 1, pp. 8–14, 2019.
- [13] W. Maghfiroh and N. Ulinnuha, "Application of Fuzzy C-Means in Grouping Districts/Cities Based on Health Service Facilities in East Java," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 3, no. 2, pp. 62–68, 2018.
- [14] M. Kurniawan, R. R. Muhima, and S. Agustini, "Comparison of Clustering K-Means, Fuzzy C-Means, and Linkage for Nasa Active Fire Dataset," *Int. J. Artif. Intell. Robot.*, vol. 2, no. 2, p. 34, 2020.
- [15] M. Khozin, D. Cahyono, and T. A. Wijaya, "Sistem Pendukung Keputusan Penerimaan dan Pengelompokan Kelas Peserta Didik Baru di SMP Islam Parlaungan Waru Sidoarjo," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 1, no. 2, pp. 98–107, 2016.
- [16] S. Sikri and N. Ulinnuha, "Perbandingan Metode Single Linkage , Complete Linkage Dan Average Linkage dalam Pengelompokan Kecamatan Berdasarkan Variabel Jenis Ternak Kabupaten Sidoarjo," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 4, no. 2, 2019.
- [17] R. R. Muhima, M. Kurniawan, and O. T. Pambudi, "A LOF K-Means Clustering on Hotspot Data," *Int. J. Artif. Intell. Robot.*, vol. 2, no. 1, p. 29, 2020.
- [18] R. Hariyanto and M. Z. Sarwani, "Optimizing K-Means Algorithm by Using Particle Swarm Optimization in Clustering for Students Learning Process," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 6, no. 1, pp. 65–68, 2021.
- [19] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *J. Ultim.*, vol. 6, no. 1, pp. 15–20, 2014.
- [20] Mapilindo and Anim, "Persepsi Mahasiswa FKIP tentang Pembelajaran Daring Selama Pandemi Covid-19," *Pros. Semin. Nas. Multi Disiplin Ilmu Univ. Asahan*, no. September, pp. 1–8, 2020.
- [21] M. Juliya and Y. T. Herlambang, "Analisis Problematika Pembelajaran Daring dan Pengaruhnya Terhadap Motivasi Belajar Siswa," *Genta Mulia*, vol. XII, no. 1, pp. 281–294, 2021.
- [22] F. AGUS SUSILO, "Peningkatan Efektivitas Pada Proses Pembelajaran," *MATHEdunesa*, vol. 2, no. 1, 2013.
- [23] M. Mustakim, "Efektivitas Pembelajaran Daring Menggunakan Media Online Selama Pandemi Covid-19 Pada Mata Pelajaran Matematika," *Al asma J. Islam. Educ.*, vol. 2, no. 1, p. 1, 2020.
- [24] A. C. P. Harahap, D. P. Harahap, and S. R. Harahap, "Analisis Tingkat Stres Akademik Pada Mahasiswa Selama Pembelajaran Jarak Jauh Dimasa Covid-19," *Biblio Couns. J. Kaji. Konseling dan Pendidik.*, vol. 3, no. 1, pp. 10–14, 2020.
- [25] E. Fachrosi, "74. Perbedaan Stres Akademik Antara Kelompok Siswa Minoritas dengan Mayoritas di SMP WR.Supratman 2 Medan," *Fak. Psikol. Univ. Sumatera Utara*, pp. 17–20, 2012.
- [26] K. D. R. Sianipar, S. W. Siahaan, M. Siregar, and P. P. P. A. N. W. F. I. R. H. Zer, "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Online," *Infomatek*, vol. 22, no. 1, pp. 23–30, 2020.
- [27] N. Narwati, "PENGELOMPOKAN MAHASISWA MENGGUNAKAN ALGORITMA K-MEANS," *J. Din. Inform.*, vol. 2, no. 2, 2010.
- [28] A. R. Rinaldi, L. Surlanto, D. Sudrajat, and D. A. Kurnia, "AnalisaTingkat Kepuasan Mahasiswa Terhadap Layanan Pembelajaran Menggunakan K-Means dan Algoritma Genetika," *J. ICT Inf. Commun. Technol.*, vol. 18, no. 1, pp. 60–64, 2019.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

