Proceedings of the
1st International Conference on Environmental Health, Socioeconomic and Technology 2022
Lamongan, Indonesia, August 10-11 2022
Vol. 1, No. 1 (https://conference.unisla.ac.id/index.php/icehst/)

AN IMPLEMENTATION OF K-MEANS CLUSTERING ON E-LEARNING DATA

M. Ghofar Rohman[1], Purnomo Hadi Susilo[2], Riky Yoga Perkasa[3], Auliya'ur Rohman [4]
[1,2,3,4] Informatics Engineering of Universitas Islam Lamongan, Jl. Veteran No. 53 A, Lamongan, East Java, Indonesia Country

[1]m.ghofarrohman@unisla.ac.id (+62 856-451-90-498) ;
[2]purnomo@unisla.ac.id (+62 857-332-90-212);
[3]reizzzazz@gmail.com (+62 838-301-75-697);
[4]akunsayapertama01@gmail.com (+62 881-027-413-971);

ABSTRACT

With the COVID-19 pandemic, requiring learning to be done online, the assessment of learning in online learning is certainly different from learning outcomes in class. This study aims to implement the clustering kmeans method on online learning assessment data in order to determine clusters on the learning outcomes data. This study uses the k-means method for the clustering process in order to get the results of classifying learning assessments in online learning activities. The results of this study indicate that in 5 clusters with the following details: in cluster 1 there are 113 students, in cluster 2 there are 11 students, in cluster 3 there are 4 students, in cluster 4 there are 12 students, and in cluster 5 there are 5 students.

KEYWORDS: *K-means, Clustering, e-learning, Learning assessment*

1 INTRODUCTION

As the COVID-19 epidemic continues to spread, one effort or method the government is using to enhance educational opportunities is distance learning. A circular about the prevention of the Corona Virus Disease (COVID-19) in universities was released by the Ministry of Education and Culture (Kemendikbud) Directorate of Higher Education Number 1 of 2020 as part of this government initiative. Learn from their homes, or take remote education courses, to stop and reduce the development of COVID-19 (Hulukati et al., 2021).

Online learning is learning that is accessible, connected, flexible, and allows for a variety of interactive learning activities to take place through an e-learning platform at any time and from any location (Hulukati et al., 2021) (Chu et al., 2021). The government made the proper decision to implement the educational process through online learning in the midst of the COVID-19 pandemic.

By incorporating online learning, the use of digital technology has significantly benefited educational institutions, particularly those at the university level, during the

COVID-19 epidemic. In order to achieve successful learning outcomes, teachers require students to be able to follow and comprehend learning materials. The implementation of university rules for online learning is equally challenging since it necessitates changes to both the technology and human resources, particularly lecturers who deliver the lecture materials. As a result, measurements are required to identify the group of students identified from the outcomes of the completed online learning.

Using techniques from data mining processing, specially a clustering algorithm that enables the discovery of clusters from online learning outcomes (Xie et al., 2019). Although in practice perfect separation is typically not possible, a decent clustering algorithm will typically create clusters with various non-overlapping borders (Werdiningsih et al., 2019) (Delgado et al., 2021). (Oding et al., 2021) A non-hierarchical data clustering technique called K-Means aims to divide existing data into one or more clusters or groupings (Bisilisin et al., 2014). As a result of this method's data clustering (Xu et al., 2020) (Oding et al., 2021), data that have similar qualities are clustered together. The K-Means method is critical to the success of this study.

The centroid model is described in the K-Means Algorithm. The model used in generating the cluster is the notion of the centroid model. Where the center point of a cluster can be referred to as the centroid, the initial centroid can be determined randomly (Xu et al., 2020). The shape of the centroid is in the form of a value. Which is the center point used to calculate the distance between a data object and the centroid (Mahmud et al., 2012). A data object can be said to be a cluster if it has the shortest distance from the centroid of a cluster. The K-Means algorithm explains how an algorithm can solve a problem in grouped data with the ultimate goal of being able to reduce failures or defects repeatedly (Ariasa et al., 2020).

According to a study by (Yustanti et al., 2021), In this study, 91 data from students who completed database courses using an unsupervised learning approach for the process clustering were employed along with 22 variables, 15 indicators for internal factors, and 7 indications for external ones. and achieve a k=4 ideal number of clusters. Furthermore, in research (Yamasari et al., 2021). in this study DBSCAN-based grouping can be applied to student behavior in the e-Learning system to detect students who perform unusual behavior. This method can work optimally if the parameter value is determined with the appropriate value. Additionally, according to (Abdo et al., 2021) research, this study's clustering results reveal that pupils with higher involvement tend to perform better than groups with.

The aims of this study is to apply the k-means clustering algorithm in determining clusters on online learning outcomes. By using the Matlab tool on the data of online learning outcomes in the online learning system of the Islamic University of Lamongan.

the following sections of this paper are structured as follows. section 2 describes the method used, namely K-mean clustering, k-mean flowchart, and the data used. in section 3 is the result and discussion, explaining the results of the research and discussion of the kmeans clustering algorithm, and section 4 is the conclusion of the paper. and lastly are there Acknowledgements

2. MATERIAL AND METHOD

A. K-Means Algorithm

K-Means is a method that is often applied to group data objects using the Euclidean distance metric (Zhai et al., 2021). how the K-Means grouping algorithm works is clearly shown in the flowchart, namely Figure 3.1

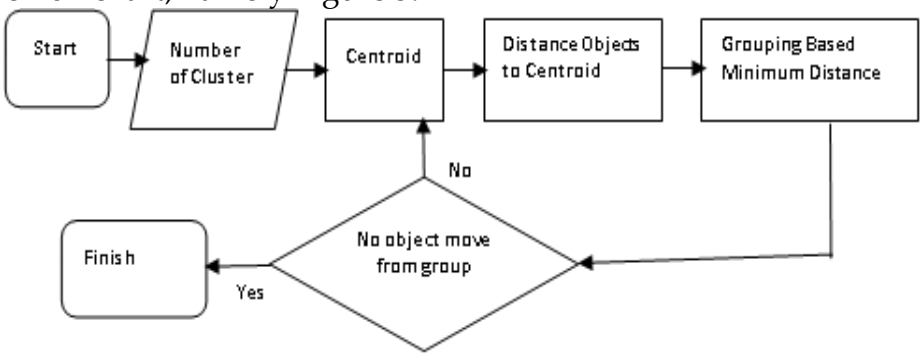

Figure 2.1. Flowchart on K-means Clustering

Here are the steps in the K-Means algorithm:

1. Initialization of the number of clusters (K)
2. centroids are selected randomly.
3. performs the calculation of the distance between the data points and the centroid of the cluster. .

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

4. Similar data points that are close to the centroid point will then move to the closest cluster.
5. The new cluster center point is formed.
6. 6. Repeat steps 3 to 5 until the centroid cluster does not change or has reached the optimum number of iterations.

B. Data Collection Stage

The data used is online learning assessment data on operating system courses taken by 145 students of the Informatics Engineering University of Lamongan Islamic University, starting from Presence (activity), Quiz 1, Midterm Exam, Quiz 2, Assignments, UAS, and Final Score (Chu et al., 2021) (see table 2.1). The data will be processed by clustering online learning outcomes using 5 clusters.

C. Data Processing Stage

The data obtained in the data collection stage, data from each student will be added up for each criterion, so at the data processing stage a value calculation will be obtained which will later be operated in the next or clustering stage.

D. Clustering Stage

The process of grouping data without maintenance and separating a set of data from the set according to the criteria into several classes is a clustering process. Equations and

stages regarding the distance algorithm, or commonly called Euclidean Distance can be applied in the process of grouping data. The method implemented to classify the data set into several parts according to the same category which has been previously defined is the cluster study (Wu et al., 2019) (Oding et al., 2021).

To get the cluster in accordance with the data already owned, we need a flow chart to assist in getting a series of calculations as a way to find out the final result of implementing the cluster on the data to be processed. Below is a flowchart or flowchart to find out the cluster with K-Means (see Figure 3.1).

Table 2.1. Online Learning Outcome

No	Name	Presence	Quiz 1	Mid Exam	Quiz 2	Assignment	Final Exam	Final Score
1	Abdillah Efendi	0	0	0	0	0	0	0
2	Dienanda Rahmani Royyan	0	0	88	81	42	0	0
3	Muhammad Harjo	86	86	89	83	86	93	87
4	Oktova Bara Adji Putra	87	85	89	83	86	87	83
5	Moch. Fikri Haikal	87	85	89	82	86	93	87
6	Brilian Angga Putra	83	88	92	81	86	80	85
7	Dwi Oktaviani	88	89	88	80	86	93	87
8	Fakhri Ahmad Fahim	88	89	88	82	87	100	89
9	Laelatul Qomaria	87	87	90	80	86	87	86
10	Rendra Rusma Dharmawan	87	85	88	83	86	100	89
11	Ahmad Faisol Akbar	77	78	0	83	60	80	62
12	Ahmad Sabil Adani	79	89	94	83	86	93	88
13	Novandy Fahrizal Fanani	87	88	90	84	87	80	83
14	Bagus Dwi Santoso	85	87	90	81	86	93	87
15	Didik Eko Prasetyo	88	87	90	80	86	93	87
……	……	….	….	….	….	….	….	….
141	Adinda Safitri	86	86	90	82	85	100	89
142	Achmad Shofiyuddin Hasan	0	87	90	82	40	64	64
143	M. Zidan Sa'id	0	85	88	82	61	100	75
144	Fatma Sa'diyah	87	87	90	92	86	100	91
145	M. Akmal Syarifuddin	86	85	88	86	84	100	89

III. RESULTS AND DISCUSSION

This study will describe how the k-means approach is used to cluster student value data in e-learning. However, the MATLAB tool will be used to illustrate this debate. There are tools available in MATLAB for clustering data. Before clustering data using

MATLAB, you can prepare data in the form of files to cluster using the k-means algorithm (can be .xls, .csv, or .dat). Next, decide how many clusters will be used. Then, enter the code for the k-means method function in MATLAB as follows:

```
clear all;
clc;
load dat.csv
[idx,C] = kmeans(dat,5);

figure;
plot(dat(idx==1,1),dat(idx==1,2),'r.','MarkerSize',
12)
hold on
plot(dat(idx==2,1),dat(idx==2,2),'b.','MarkerSize',
12)
hold on
plot(dat(idx==3,1),dat(idx==3,2),'g.','MarkerSize',
12)
hold on
plot(dat(idx==4,1),dat(idx==4,2),'y.','MarkerSize',
12)
hold on
plot(dat(idx==5,1),dat(idx==5,2),'r.','MarkerSize',
12)
plot(C(:,1),C(:,2),'kx',...
     'MarkerSize',15,'LineWidth',2)
legend('Cluster 1','Cluster 2','Cluster 3','Cluster
4','Cluster 5','Centroids',...'Location','NW')
title 'Cluster Assignments and Centroids'
hold off
```

When this program has been saved, then when it is run it will generate groups of data. These data groups can be represented using graphs/plots in MATLAB, so that the distribution of data in each cluster is obtained based on its proximity to the cluster center, the output of the coding can be seen in Figure 3.1 below.

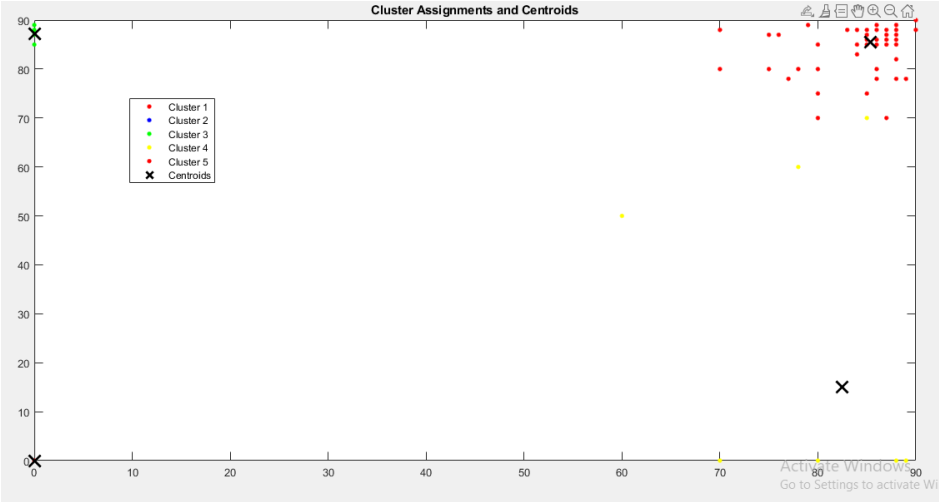

Figure 3.1. Clustering value data Using KMeans

The number of clusters to be utilized is decided after the data is entered since the centroid of the cluster in each feature is randomly determined. In this case, the sign "c" is used for clusters with a value of = 5, which denotes that this program will produce 5 clusters. In order to group the data in each cluster using the reference to the closest distance to the data point in each cluster, the system will then continue to find the center point (centroid) of each cluster in each data feature. using iteration continuously till no data is moving cluster

Table 3.1 Centroid of each cluster on each feature

No	Presence	Quiz 1	Mid Exam	Quiz 2	Assignment	Final Exam	Final Score
C1	85.3805	85.5841	88.6018	84.5929	80.9602	93.2280	86.1605
C2	0	0	89.0000	76.4545	58.8864	64.9351	54.0143
C3	0	87.2500	89.0000	83.0000	69.2500	86.0714	74.1893
C4	82.5000	15.0000	89.5833	62.4167	46.5625	68.7302	63.0277
C5	0	0	0	0	0	17.1429	3.4286

In this case, the results of clustering with k-means can be seen in table 3, in the class column. In table 3.2, it contains numbers and so on, up to class, where the class here contains the numbers 1, 2, 3, 4, and 5 which states that the data in that row is included in the respective cluster according to the entries in the class column.

To be able to access the centroid in table 3.1 above, you can use the "C" code in the command window in the matlab application, to display the class you can use the "idx" code, while to display the data row with the cluster you can use the "[dat idx]" code because the data will be displayed is a matrix.

Table 3.2. Clustering Results using K-means Agorithm in Matlab

No	Name	Presence	Quiz 1	Mid Exam	Quiz 2	Assignment	Final Exam	Final Score	Class
1	Abdillah Efendi	0	0	0	0	0	0	0	5
2	Dienanda Rahmani Royyan	0	0	88	81	42	0	0	2
3	Muhammad Harjo	86	86	89	83	86	93	87	1
4	Oktova Bara Adji Putra	87	85	89	83	86	87	83	1
5	Moch. Fikri Haikal	87	85	89	82	86	93	87	1
6	Brilian Angga Putra	83	88	92	81	86	80	85	1
7	Dwi Oktaviani	88	89	88	80	86	93	87	1
8	Fakhri Ahmad Fahim	88	89	88	82	87	100	89	1
9	Laelatul Qomaria	87	87	90	80	86	87	86	1
10	Rendra Rusma Dharmawan	87	85	88	83	86	100	89	1
11	Ahmad Faisol Akbar	77	78	0	83	60	80	62	1
12	Ahmad Sabil Adani	79	89	94	83	86	93	88	1
13	Novandy Fahrizal Fanani	87	88	90	84	87	80	83	1
14	Bagus Dwi Santoso	85	87	90	81	86	93	87	1
15	Didik Eko Prasetyo	88	87	90	80	86	93	87	1
…..	……	….	…	…	….	….	...	….	…..
141	Adinda Safitri	86	86	90	82	85	100	89	1
142	Achmad Shofiyuddin Hasan	0	87	90	82	40	64	64	3
143	M. Zidan Sa'id	0	85	88	82	61	100	75	3
144	Fatma Sa'diyah	87	87	90	92	86	100	91	1
145	M. Akmal Syarifuddin	86	85	88	86	84	100	89	1

From the results of clustering using the K-means method using the Matlab tools above, it can be concluded that:

1. the group of students in cluster 1 of 113 students with online learning outcomes cluster are on cluster 1,

2. The group of students in cluster 2 is 11 students with the cluster online learning outcomes are on cluster 2,
3. The group of students in cluster 3 of 4 students with the cluster of online learning outcomes is on cluster 3,
4. The group of students in cluster 4 of 12 students with clusters of online learning outcomes is on cluster 4, and
5. The group of students in cluster 5 of 5 students with clusters of online learning outcomes is on cluster 5.

4. CONCLUSION

The conclusions are as follows:

1. Data mining with the K-Means clustering algorithm can be applied to online learning outcomes.
2. The calculation process is carried out using the k-means function on the matlab tool with the output being the amount of data in each cluster.
3. the group of students in cluster 1 is 113 students with the online learning outcomes cluster are on cluster 1,
4. The group of students in cluster 2 is 11 students with the cluster online learning outcomes are on cluster 2,
5. The group of students in cluster 3 of 4 students with the cluster of online learning outcomes are on cluster 3,
6. The group of students in cluster 4 of 12 students with the cluster of online learning outcomes are on cluster 4, and
7. The group of students in cluster 5 of 5 students with clusters of online learning outcomes are on cluster 5.

5. ACKNOWLEDGEMENTS

Thank you to the Litbang Pemas of the Islamic University of Lamongan, for the support in the form of a research fund program for the internal grant program of the Islamic University of Lamongan in 2022.

REFERENCES

Abdo, A. M., Maisara, N., Rasid, A., Amira, N., Mohd, H., Nur, S., Sulaiman, A., Wani, S., & Zainol, Z. (2021). *Student ' s Performance Based on E-Learning Platform Behaviour using K-means Clustering*. *7*(1), 72–78.

Ariasa, K., Gunadi, I. G. A., & Candiasa, I. M. (2020). Optimasi Algoritma Klaster Dinamis pada K-Means dalam Pengelompokkan Kinerja Akademik Mahasiswa (Studi Kasus: Universitas Pendidikan Ganesha). *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, *9*(2), 181–193. https://ejournal.undiksha.ac.id/index.php/janapati/article/view/23491

Bisilisin, F. Y., Herdiyeni, Y., & Silalahi, B. I. B. P. (2014). Optimasi K-Means Clustering

Menggunakan Particle Swarm Optimization pada Sistem Identifikasi Tumbuhan Obat Berbasis Citra K-Means Clustering Optimization Using Particle Swarm Optimization on Image Based Medicinal Plant Identification System. *Ilmu Komputer Agri-Informatika*, *3*(2002), 38–47.

Chu, Y. W., Tenorio, E., Cruz, L., Douglas, K., Lan, A. S., & Brinton, C. G. (2021). Click-Based Student Performance Prediction: A Clustering Guided Meta-Learning Approach. *Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021*, 1389–1398. https://doi.org/10.1109/BigData52589.2021.9671729

Delgado, S., Moran, F., Jose, J. C. S., & Burgos, D. (2021). Analysis of Students' Behavior through User Clustering in Online Learning Settings, Based on Self Organizing Maps Neural Networks. *IEEE Access*, *9*, 132592–132608. https://doi.org/10.1109/ACCESS.2021.3115024

Hulukati, E., Achmad, N., & Bau, M. A. (2021). Deskripsi Penggunaan Media E-Learning dalam Pembelajaran Matematika di Masa Pandemi Covid-19. *Jambura Journal of Mathematics Education*, *2*(1), 21–27. https://doi.org/10.34312/jmathedu.v2i1.10061

Mahmud, M. S., Rahman, M. M., & Akhtar, M. N. (2012). Improvement of K-means clustering algorithm with better initial centroids based on weighted average. *2012 7th International Conference on Electrical and Computer Engineering, ICECE 2012*, 647–650. https://doi.org/10.1109/ICECE.2012.6471633

Oding, H., Shanti, M., & Firdaus, E. A. (2021). STRATEGI PEMASARAN PRODUK INDUSTRI KREATIF MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING BERBASIS PARTICLE SWARM OPTIMIZATION. *Jurnal Nuansa Informatika*, *15*, 1–13.

Werdiningsih, D., Rahmawati, D., Cholissodin, I., & Santoso, N. (2019). Optimasi K-Means untuk Pengelompokan Data Kinerja Akademik Dosen menggunakan Particle Swarm Optimization. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(4), 4102–4110.

Wu, M., Li, X., Liu, C., Liu, M., Zhao, N., Wang, J., Wan, X., Rao, Z., & Zhu, L. (2019). Robust global motion estimation for video security based on improved k-means clustering. *Journal of Ambient Intelligence and Humanized Computing*, *10*(2), 439–448. https://doi.org/10.1007/s12652-017-0660-8

Xie, H., Zhang, L., Lim, C. P., Yu, Y., Liu, C., Liu, H., & Walters, J. (2019). Improving K-means clustering with enhanced Firefly Algorithms. *Applied Soft Computing Journal*, *84*, 105763. https://doi.org/10.1016/j.asoc.2019.105763

Xu, H., Yao, S., Li, Q., & Ye, Z. (2020). An Improved K-means Clustering Algorithm. *IDAACS-SWS 2020 - 5th IEEE International Symposium on Smart and Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems, Proceedings*, *i*, 0–4. https://doi.org/10.1109/IDAACS-SWS50031.2020.9297060

Yamasari, Y., Qoiriah, A., Rochmawati, N., Yustanti, W., Wintarti, A., & Ahmad, T. (2021). Clustering the Students' Behavior on the e-Learning using the Density-Based Algorithm. *Proceedings - 2021 International Seminar on Application for Technology of Information and Communication: IT Opportunities and Creativities for Digital Innovation*

and Communication within Global Pandemic, ISemantic 2021, 22–27. https://doi.org/10.1109/iSemantic52711.2021.9573234

Yustanti, W., Rochmawati, N., Yamasari, Y., & Qoiriah, A. (2021). Clustering Students in the Online Learning Climate based on Internal and External Factors. *Proceedings - 4th International Conference on Vocational Education and Electrical Engineering: Strengthening Engagement with Communities through Artificial Intelligence Application in Education, Electrical Engineering and Information Technology, ICVEE 2021*, 0–5. https://doi.org/10.1109/ICVEE54186.2021.9649723

Zhai, T., Qu, J., & Ling, Y. (2021). Analysis of Procrastination State of E-learning Based on Clustering and Social Network. *Proceedings - 2021 2nd International Conference on Artificial Intelligence and Education, ICAIE 2021*, 357–360. https://doi.org/10.1109/ICAIE53562.2021.00081