https://jurnal.seaninstitute.or.id/index.php/jukomi
Jurnal Komputer Indonesia (JU-KOMI)
Volume 3 No 01, 2024

JU-KOMI
E-ISSN: 2963-0460

K-Means Clustering of Student Mid-Term and Final Exam Data

Nella Ane Br Sitepu[1], Agnesia Rointan Sijabat[2], Cindy Rounali Limbong[3], Lenny Evalina Pasaribu[4], Einson O.B Nainggolan[5], Michael Manulang[6]
Universitas Katolik Santo Thomas Medan Fakultas Ilmu Komputer

ARTICLE INFO

Keywords:
k-means clustering, UAS scores, UTS scores

Email :
anebrusitepu@gmail.com

ABSTRACT

This study examines the use of the k-means clustering method in grouping students based on UAS and UTS scores to identify patterns of academic achievement. Clustering is an effective data mining technique for grouping data based on similar characteristics. By applying the k-means algorithm, this study aims to make it easier for lecturers to identify student abilities, so that they can provide appropriate support to those who need help. Data were taken from UTS and UAS scores of students at a university in Indonesia, and the results of the analysis showed that k-means clustering can group students according to their level of achievement. These findings are expected to help in the development of more effective teaching strategies and interventions, improving the quality of education and overall academic performance of students.

INTRODUCTION

Clustering is a method in data mining that aims to group data because of the similarity of characteristics between the data created. This method is often used to recognize patterns in a large and varied set of data. The clustering algorithm is a method that is often used, this method is quite appropriate for grouping data into several clusters that have been created in advance, so that the data in a cluster has a fairly similar similarity.

This aims to use k-means clustering to create groups of data on students' UAS and UTS scores. By implementing this method, it can help determine groups of smart, less smart, and fairly smart students. This method can make it easier for lecturers to identify their students' abilities, so that lecturers can provide assistance to students who are lacking in their learning abilities.

The advancement of the current era has created an impact that needs to be studied, including in the field of Education. In this digital era, all campuses or universities around the world are required to improve the quality of Education. One thing that needs to be observed in improving the quality of Education is careful observation of the progress of student grades. This observation can be related to the UAS value, as well as the student's UTS value.

The output of this research is expected to provide insight for those who regulate Education in mapping student work results by determining student groups based on learning outcomes or achievement levels. Teaching and learning techniques can address the needs of a group of students. In addition, this clustering can also provide motivation for student groups to be more diligent so that learning outcomes are more satisfying.

In Indonesia, the problem related to monitoring the progress of student grades is one of the important aspects. In some cases, the observation process is often done manually, which is troublesome and wastes a lot of time. Such problems can be related to the large number of classes, which can make it difficult for lecturers to direct adequate attention to each student.

There are various studies that have been conducted to facilitate observation in the field of Education. Some studies are conducted on the use of technology to shorten the time for assessment, while others refer to data observation methods to gain insight into student work results.

Although the effort to simplify the observation process, there are some gaps in this observation. One of them is using the clustering method to observe student performance results related to UTS, UAS scores. This method provides a good understanding of student work results and makes it easier for campuses to create the right techniques to improve the learning process.

This observation aims to use the k-means clustering method on students' UTS and UAS scores to observe how students work differently. So, the expected output is the level of understanding of aspects that can influence student performance results that increase well and can improve student performance to be better than before. This is studied to fill the gap, namely by using k-means clustering to understand how students work based on students' UTS and UAS scores, where the output of this observation can provide development of higher education.

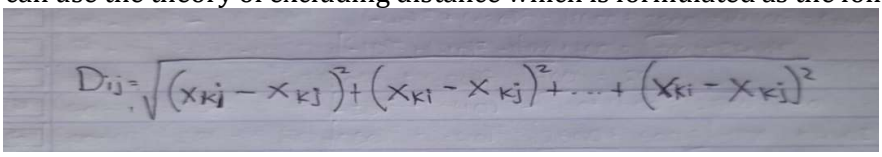
METHOD

The method that will be used in this observation is the k-means clustering method, because the k-means clustering method can group data based on similarities between data. This method is also appropriate for use in this study because it can observe patterns of student UTS and UAS value data. This k-means clustering method uses excel

1. Clustering
   Clustering is one of the many data mining methods that has an unspervised nature (without direction) namely the characteristics of each Cluster are not predetermined, but the Cluster, which is formed refers to similar attributes of a Cluster. Therefore, from a set of data and similar attributes will be grouped according to their characteristics and represented as points in a multidimensional space.
2. K-Means Algorithm
   K-Means is grouped using a cluster analysis method that is directed at partitioning. N observation objects into K groups or referred to as Clusters, where each observation object has a group with an average or mean. The K-Means algorithm is included in the application of Clustering data mining which can be used to group data into several groups. The groups formed have predetermined criteria, then the data that matches the group is collected into one cluster.

The center point or Centroid is something that every cluster has. The stages of the K-Means algorithm are as follows:

1. Define k as the number of clusters you want to form.
2. Initialize k initial centroids (cluster center points) randomly.
3. Allocate each data or object to the nearest cluster. The distance between objects and the distance between objects with a particular cluster is determined by the distance between the data and the cluster center. To calculate the distance of all data to each cluster center, you can use the theory of excluding distance which is formulated as the following equation.



Where:
D(i,j) = distance of data i to cluster center j,
Xki = i data in k data attribute,
Xkj = center point of cluster j on attribute k.

The distance of the cluster center is recalculated with the current cluster membership. The cluster center is the average of all data or objects in a particular cluster, if desired the median value of the cluster can also be used.

4. Repeat the third step until the result of the iteration is the same as the previous iteration.

Sample case:

NAME	UTS	UAS
BUDI	40	30

https://jurnal.seaninstitute.or.id/index.php/jukomi
Jurnal Komputer Indonesia (JU-KOMI)
Volume 3 No 01, 2024
JU-KOMI
E-ISSN: 2963-0460

NAME	UTS	UAS
YU	70	90
SISKA	80	60
YUDI	50	40
BAYU	30	40
ARDI	90	90

The problem
Group the data with the following provisions:

1. the number of data groups is 2
2. 2. The number of attributes is 2, namely the UTS and UAS scores.
3. the amount of data is 6

Literacy 1 >>

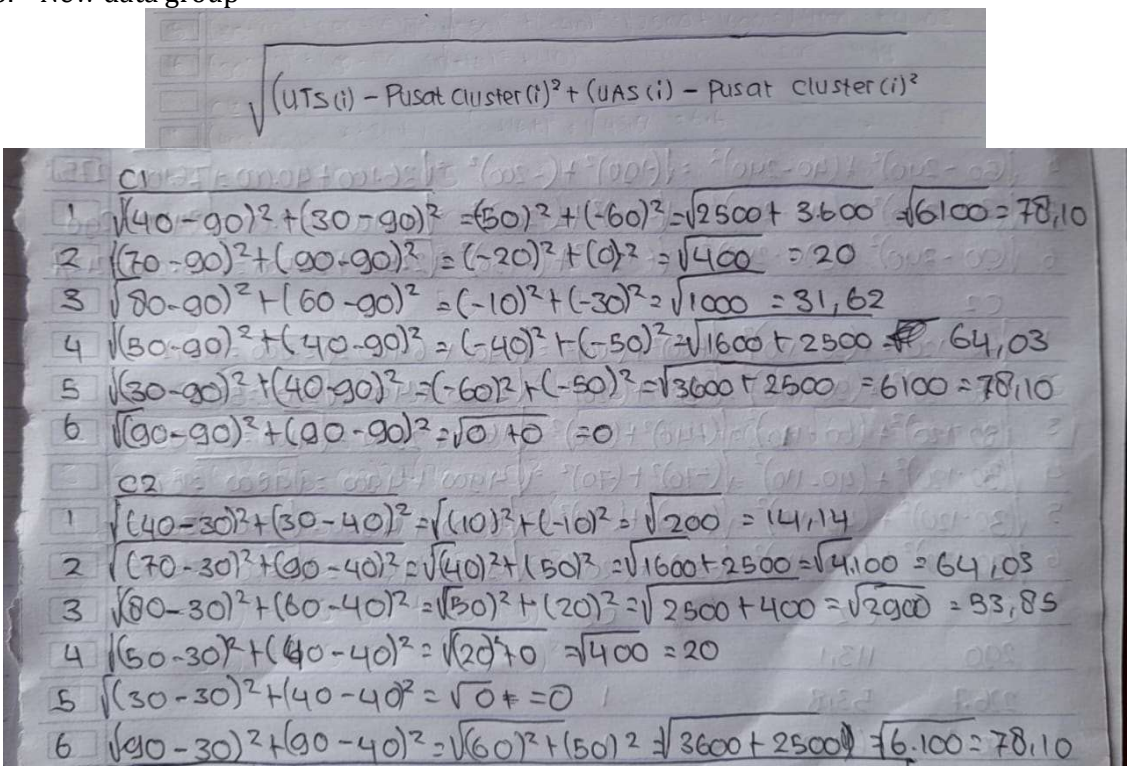
1. center point determination
   The 6th data is taken as the center of the 1st cluster.

c1	90	90

   The 5th data is taken as the center of the 2nd cluster.

c2	30	40

2. Determining the distance of the center of each cluster
3. New data group



The shortest distance is the ratio of the smallest values of c1 and c2.

2nd

Student name	UTS	UAS	C1	C2	Shortest distance	C1	C2
Budi	40	30	78.10	14.14	14.14		1
Beautiful	70	90	20	64.03	20	1	
Siska	80	60	31.62	53.85	31.62	1	
Yudi	50	40	64.03	20	20		1
Bayu	30	40	78.10	0	0		1
ardi	90	90	0	78.10	0	1	

Clustering>>

https://jurnal.seaninstitute.or.id/index.php/jukomi
Jurnal Komputer Indonesia (JU-KOMI)
Volume 3 No 01, 2024
JU-KOMI
E-ISSN: 2963-0460

The 1st cluster center ((70+80+90)/3=80 and (90+40+40)/3=80)

c1	80	80

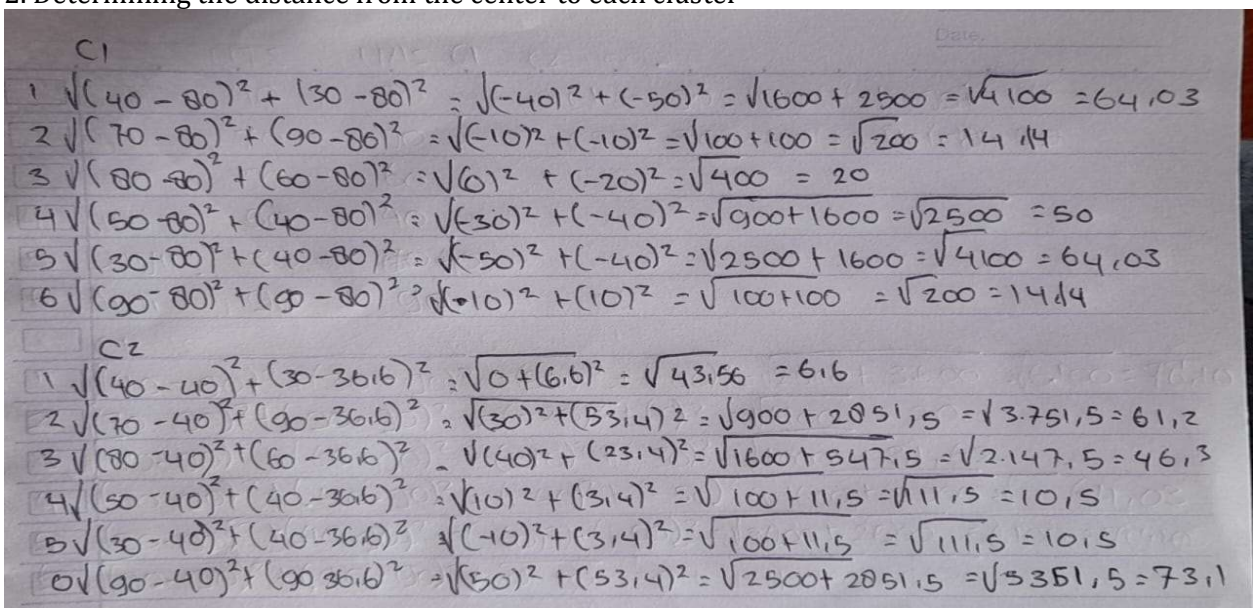
2nd cluster center ((40+50+30)/3=40 and (30+40+40)/3=36.6)

c2	40	36.67

1. Determining the center point

Student name	UTS	UAS	C1	C2
Budi	40	30		1
Beautiful	70	90	1	
Siska	80	60	1	
Yudi	50	40		1
Bayu	30	40		1
ardi	90	90	1	

2. Determining the distance from the center to each cluster



3.new data group

Student name	UTS	UAS	C1	C2	Shortest distance	C1	C2
Budi	40	30	64.03	6.6	6.6		1
Beautiful	70	90	14.14	61.2	14.14	1	
Siska	80	60	20	46.3	20	1	
Yudi	50	40	50	10.5	10.5		1
Bayu	30	40	64.03	10.5	10.5		1
ardi	90	90	14.14	73.1	14.14	1	

RESULTS AND DISCUSSION

After doing literacy 1 to literacy 2, there will be no more changes if repeated until literacy 3, the results will remain the same as literacy 2. So here are the results of the grouping of UTS and UAS value data, students and comparison of literacy 1 and 2

Literacy 1

NAME	UTS	UAS	c1	c2	c1	c2
BUDI	40	30	78.1025	14.14214		1
YU	70	90	20	64.03124	1	

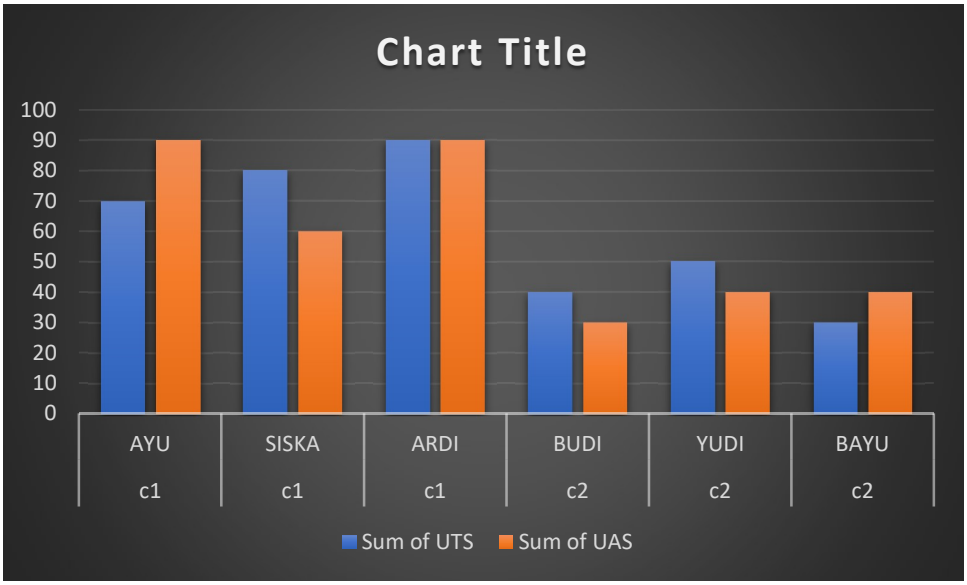
NAME	UTS	UAS	c1	c2	c1	c2
SISKA	80	60	31.62278	53.85165	1	
YUDI	50	40	64.03124	20		1
BAYU	30	40	78.1025	0		1
ARDI	90	90	0	78.1025	1	

Literacy 2

NAME	UTS	UAS	c1	c2	c1	c2
BUDI	40	30	64.03	6.666667		1
YU	70	90	14.14214	61.19187	1	
SISKA	80	60	20	46.30815	1	
YUDI	50	40	50	10.54093		1
BAYU	30	40	64.03124	10.54093		1
ARDI	90	90	14.14214	73.10571	1	

So from the table above it has the same clustering
Here is the graph:

CLASTER MEMBERSHIP	STUDENT NAME	Sum of UTS	Sum of UAS
c1	YU	70	90
c1	SISKA	80	60
c1	ARDI	90	90
c2	BUDI	40	30
c2	YUDI	50	40
c2	BAYU	30	40



So from this table we can determine which students are grouped into students with the highest and lowest scores.

CONCLUSION

For the final result after using the k-means clustering method on the UAS and UTS student value data through this excel application, it has been obtained to group students based on the similarity of their achievement results. In this process, we first determine the data, then determine the central ceteroid. Then after that we will calculate the closest distance to the closest distance, after that we will do literacy until the literacy has a value that does not change between

the final value and the previous literacy. So this k-means clustering helps identify hidden patterns in the achievement level data and allows decision making for teachers to be able to find solutions to improve student academic achievement.

REFERENCE

Adistya, A. P., Luthfyani, N., Tara, P., Adriyan, R., & Rosyani, P. (2023). Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Kelulusan Mata Kuliah Mahasiswa. OKTAL: Jurnal Ilmu Komputer dan Sains, 2(08), 2301-2306.

Poerwanto, B., & Fa'rifah, R. Y. (2016). Analisis cluster k-means dalam pengelompokan kemampuan mahasiswa. Jurnal Scientific Pinisi, 2(2), 92-96.

Maulana, I., & Rosalina, U. (2020). Clustering Data Nilai Ujian Akhir Semester Menggunakan Algoritma Data Mining K-Means. PERISKOP: Jurnal Sains Dan Ilmu Pendidikan, 1(2), 76-85.

Pradnyana, G. A., & Permana, A. A. J. (2018). Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode K-Means Dan K-Nearest Neighbors Untuk Meningkatkan Kualitas Pembelajaran. Jurnal Ilmiah Teknologi Informasi, 16(1), 59-68.

Talakua, A. C. (2023). Penerapan Data Mining untuk memprediksi Nilai Akhir Matakuliah Menggunakan Metode K-Means Clustering. JURNAL TEKNIK INFORMATIKA INOVATIF WIRA WACANA, 2(3), 178-186.

Ariawan, M. P. A., Peling, I. B. A., & Subiksa, G. B. (2023). Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus: Matakuliah Pemrograman Dasar). J. Nas. Teknol. dan Sist. Inf, 9(2), 122-131.

Poerwanto, B., & Fa'rifah, R. Y. (2016). Analisis Cluster Menggunakan Algoritma K-Means. Jurnal Ilmiah Teknologi Informasi, 6(2).

Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustring. JURTEKSI (Jurnal Teknologi dan Sistem Informasi), 4(1), 85-92.

Wiyono, D. E. (2020). Pengelompokan Mahasiswa Berdasarkan Pencapaian Prestasi Belajar dari Mata Kuliah yang Ditempuh Berbasis Web dengan K-Means Clustering. Journal of Telecommunication Electronics and Control Engineering (JTECE), 2(2), 69-77.