K-Medoids Algorithm Analysis in Grouping Students' Level of Understanding of Subjects

Ehrlich F.T Butarbutar[1,*], M. Safii[2]

[1*] STIKOM Tunas Bangsa Pematangsiantar, Sumatera Utara, Indonesia
[2] AMIK Tunas Bangsa Pematangsiantar, Sumatera Utara, Indonesia
[1,2]Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara
[1]andrebutarbutar98@gmail.com *; [2] m.safii@amiktunasbangsa.ac.id

Abstract

Analysis of the teaching and learning process needs to be done as feedback on the understanding of the material for students. One of the obstacles faced by schools is that there is no method of how this feedback can be done so that student achievement is uneven. Student achievement in subjects can be seen from the results of the scores on the report cards obtained by students after taking the final semester exam. Due to the uneven achievement of students, it is necessary to make a method so that feedback analysis can be carried out on the level of student understanding of the subject. Is data mining with clustering techniques using the K-Medoids algorithm. With this algorithm, students' understanding of subjects with high potential can be grouped with high brightness average results.

Keywords: Understanding of the material, *Feedback, Data mining, Clustering, K-Medoids algorithm*

1.   Introduction

Along with the times, computers are increasingly playing a role in people's lives. Almost all areas of life have used computers as tools. The rapid development of computers can be directly felt by the community. The application of information technology in education can also produce abundant data about students and the resulting learning process [1]. One of them is how to find out students who understand the subject. This problem often occurs because it is difficult to determine which students understand and do not understand the subject.

SMA Negeri 5 Pematangsiantar is a public school located on Jalan Medan Pematangsiantar City. As an educational institution, the school is required to provide services and learning processes that are acceptable and can produce outstanding students. The teaching and learning process is one of the activities in a school in the intellectual life of the nation. This role cannot be separated from professional teaching staff and good student understanding. In the teaching and learning process, there is a need for a two-way relationship between students and teaching staff. This is intended so that there is good cooperation during the teaching and learning process. The analysis carried out by the school on the teaching and learning process really needs to be done at the end of the semester and predicting the level of students' understanding of the subject is very important to find out the extent to which students understand the subjects delivered by the educator during teaching and learning activities and to determine the ability of educators to convey subjects. With the K-Medoids method, the author tries to cluster students' understanding of the subject to produce data groups with high potential with large average brightness results.

The K-Medoids algorithm is an algorithm similar to K-Means because both algorithms are partitional which split the dataset into groups[2],[3]. The difference lies in determining the cluster center, where the K-Means algorithm uses the average value of each cluster as the center of the cluster and the K-Medoids algorithm uses data objects as representatives (medoids) as the center of the cluster[4],[5]. Clustering is a method of grouping records, observing, or paying attention and forming classes of objects that have closeness or similarity [6]. Cluster is a collection of several records that have similarities with one another and have dissimilarities with

several records in other clusters. Clustering differs from classification in that there is no target variable in clustering. Clustering does not attempt to classify, estimate, or predict the value of the target variable. However, the clustering algorithm tries to divide all data into several groups that have closeness/similarity (homogeneous), where the similarity of records in one group will be of greater value, while the similarity of records in other groups will be of smaller value [7].

There are several previous studies that became references in this study, among others with the title Application of Clustering and Classification Algorithms at the Interest Level of Learning Systems at the Open University with a background for clustering and classifying data from the questionnaire on the importance of the Open University learning system using RapidMiner 9.0.0.3 software[8],[9]. Further research is the implementation of the K-Medoids Clustering Method to Find Out the Pattern of Selection of New Student Study Programs in 2018 at Kanjuruhan University Malang, the background is with the K-Medoids method so that the pattern of study program selection for new students can be known[10]. Furthermore, using the K-Medoids Algorithm Analysis of the K-Medoids Clustering Algorithm in Grouping the Spread of Covid-19 in Indonesia [11],[12], with a background in order to know the pattern of selection for determining the grouping of the spread of Covid-19 for the community[13],[14]. Furthermore, with the theme of clustering on the handling of diarrhea cases in Indonesia[15]. To find out which areas have the highest and lowest cases of diarrhea in Indonesia.

2.   Research Methodology

The research method consists of two methods, namely qualitative research methods and quantitative research methods. Qualitative research method is a method based on post-positivism philosophy, as well as an artistic method because the research process is more artistic in nature less patterned. While quantitative research methods are one type of research whose specifications are systematic, planned, and clearly structured from the beginning to manufacture research design.

2.1. Literature Review
Related researches in this study are previous studies that have been carried out with various kinds of problems solved using the K-Medoids algorithm, such as:

Research conducted by [8] with the title Application of Clustering and Classification Algorithms at the Interest Level of Learning Systems at the Open University Providing analysis results using the K-Medoids algorithm by dividing clusters into 2 clusters is very helpful in grouping data on the importance of learning systems at UT. The results of the analysis using several classification algorithms obtain the highest accuracy value shown in the Naïve Bayes algorithm of 72.70% with an AUC value of 0.499, but the accuracy and AUC values are still categorized as low.

Research conducted by [16] with the title Clustering Technique With K-Medoids Algorithm to Handle Promotional Strategies at Tedc Polytechnic Bandung, Stating that the K-Medoids Algorithm for the data clustering method can be applied to produce information that can be proposed to the Director of the Bandung TEDC Polytechnic in the future so that it can support new student admission promotion activities more efficiently and effectively. namely by collaborating with SMKs in West Java Province.

2.2. Data Analisys
The data preprocessing stage is the data selection stage which aims to obtain clean and ready data for use in research. Conduct selection by paying attention to data consistency, missing value and redundant data. Then the data is processed using RapidMiner using Performance which functions as data validation and reliability to find data accuracy.

The data used will be data on grades XII-PMIA 1-5 at SMA 5 Pematangsiantar with a total of 176 data and the variables used are 6 subjects, namely Indonesian, English, Mathematics, Chemistry, Biology and Physics subjects.

Table 1: Data Value of Class XII-PMIA1-5 SMA 5 Pematangsiantar Students

NO	Student's Name	Score					
		Indonesian	English	Mathematics	Physics	Chenistry	Biology
1	ABDUL HAMID	73	80	70	75	70	75
2	AHMAD FIRMANSYAH	75	70	75	80	70	75
3	ANDI	80	75	80	75	70	80
4	ANDI RAHMAD DAULAY	76	80	70	70	80	70

5	ANDINI SAPITRI	74	75	75	75	70	70
6	BOY SERGIO LUMBAN	70	70	70	80	70	73
7	BUDI PRAYOGI	78	75	75	70	73	75
8	BUNGA SALEHA LUBIS	75	80	80	75	75	80
9	CECILIA PRAMUDITHA	70	70	70	70	80	76
10	CLAUDYA AZHARI LUBIS	75	75	75	75	76	74
11	DAHWIYAH SYAH PUTRI	80	70	80	80	74	70
12	DERMAWAN RIZKY	70	75	75	75	70	78
13	ELCA ARIO ANJATRI	75	80	70	70	78	75
14	GABRIEL FAGA	70	75	75	75	75	70
15	GILBERT WISUDAWAN	75	70	80	80	70	75
16	GLADIS DWI TRINITY	80	75	75	70	75	80
17	ISRO AYU MARBUN	70	80	80	75	78	75
18	KIKI ALFANI SINURAT	75	70	70	80	75	80
19	MHD FATHUR RIDWAN	80	75	75	70	70	70
20	MIA KESUMA	75	80	70	75	75	70
21	MUHAMMAD FADHLY	75	75	75	70	80	73
22	NAIA RENCANI SUCI	75	75	80	75	76	75
23	NATHASYA CRISTIN	70	75	75	80	74	80
24	NISA IMA GUSTIANI	75	70	70	75	70	76
25	NUNING MAYANG SARI	80	75	75	70	70	74
176	....	.....	....	.....	....	....	....

2.3. Research Workflow
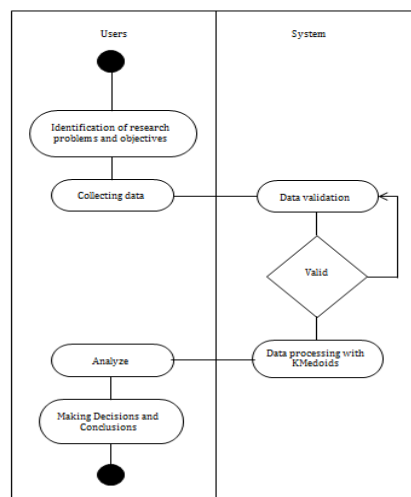The workflow carried out in this study is presented in the following figure:

Figure 1: Research Flow

Figure 1 is the flow of activities carried out by research on the system used. in the figure describes the user to identify the problem and the purpose of the research carried out, then validate the data using the performance on the RapidMiner application. If the data is valid, then the data is processed in RapidMiner, the results of the information provided by the RapidMiner application, the user creates rules and analyzes that have been carried out. The author or user makes decisions from the results of the research and draws conclusions on the research conducted.

3.   Result and Discussion
3.1. K-Medoids . Algorithm Solution
The following are the steps in data processing using the K-Medoids Algorithm[17],[18],[19]:

a. Determine the data to be processed

   The data on grades XII-PMIA 1-5 students at SMA 5 Pematangsiantar with a total data of 176 data and variables used in 6 subjects, namely Indonesian, English, Mathematics, Chemistry, Biology and Physics subjects which have been attached to the research data table.

b. Determine the Number of Clusters

   The number of clusters or k used is 3, namely High Clusters, Medium Clusters and Low Clusters.

c. Determine the Medoids Center

   Determining the Medoids Center is done randomly in each Cluster. The Medoid Centers used are:

Table 2: Determine the Medoids Center

Student's Name	Ind	Eng	Math	Chem	Bio	Phy
ZHAFFRAN RULINSANI	73	70	75	75	70	70
TRI NOVIANTI	80	76	75	70	70	75
YULVA INDAH RIZKI	75	74	70	75	80	80

d. Calculating Euclidian Value

   To calculate the distance between the Medoid point and the point of each object using Euclidian Distance. The formula to calculate the Euclidian Value is:

$$D_{(i,f)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2}$$

Then the calculation for the distance from the Medoid is as follows:

$$D_{x3,c1} = \sqrt{\begin{array}{l}(73-73)^2 + (80-70)^2 + (70-75)^2 \\ +(75-75)^2 + (70-70)^2)^2 + (75-70)^2\end{array}} = 12{,}2474487$$

And so on until D_{x176c1}. Furthermore, the calculation for the distance from the 2nd Medoid is as follows:

$$D_{x1,c1} = \sqrt{\begin{array}{l}(73-80)^2 + (80-76)^2 + (70-75)^2 \\ +(75-70)^2 + (70-75)^2)^2 + (75-70)^2\end{array}} = 10{,}72380529$$

And so on until D_{x176c2}. Furthermore, the calculation for the distance from the 3rd Medoid is as follows:

$$D_{x2,c1} = \sqrt{\begin{array}{l}(73-75)^2 + (80-74)^2 + (70-70)^2 \\ +(75-75)^2 + (70-80)^2)^2 + (75-80)^2\end{array}} = 12{,}84523258$$

And so on until D_{x176c3}. So that the distance table from Medoid is obtained and find the minimum value of the three Medoids.

Table 3: 1 Iteration Medoid Distance Results

C1	C2	C3	Range
12,24744871	10,72380529	12,8452326	10,72380529
7,348469228	12,68857754	13,820275	7,348469228
14,10673598	8,717797887	15,0332964	8,717797887
16,09347694	13,49073756	12,7279221	12,72792206
5,099019514	9,327379053	15,0665192	5,099019514
8,246211251	16,2788206	14,6628783	8,246211251
10,44030651	3,741657387	11,5758369	3,741657387
15,93737745	11,87434209	12,6885775	11,87434209
13,96424004	16,18641406	9,05538514	9,055385138
9	9,38083152	8,83176087	8,831760866

10,72380529	14,2126704	17,3781472	10,72380529
9,899494937	11,61895004	12,4498996	9,899494937
15,58845727	11,40175425	9,48683298	9,486832981
12,76714533	11,40175425	10	10
10,72380529	10,58300524	11,9163753	10,58300524
…………….	…………….	…………….	…………….
0	11,61895004	15,6524758	0
	Total Range		**1542,881135**

From table 3, it can be seen the results of the Medoid distance of iteration 1 Grouping the Level of Understanding of High School 5 Students on the Subjects that will be used to determine the Cost value. This Near Distance value will be used as the location of the cluster or group based on the smallest value per row.

e. Calculating Cost
The Cost value is obtained from the total sum of the Medoid short distance values obtained, namely:
***Cost* = 1542,881135**

f. Repeat Steps 4 and 5 with the new Medoid Center value.
Repeat the process of calculating the Medoid close range value with a new random Medoid center value, namely:

Table 4: Determine the Medoids Center

Student's Name	Ind	Eng	Math	Chem	Bio	Phy
NUNING MAYANG SARI	80	75	75	70	70	74
YABES ARMOND	75	75	80	70	75	75
MHD. ALWI YUDHA	73	75	74	74	70	76

From processing using the new Medoid center, the results are as follows: Table 5:

Table 5: 2 Iteration Medoid Distance Results

C1	C2	C3	Range
11,18034	13,37909	6,557439	6,557438524
13,2665	13,22876	8,185353	8,185352772
9,848858	10	10,0995	9,848857802
71,11259	13,2665	14,21267	13,26649916
75,51159	10,04988	6,244998	6,244997998
71,59609	16,70329	9,746794	9,746794345
75,09328	6,164414	7,211103	6,164414003
80,84553	8,660254	10,34408	8,660254038
71,61704	13,2665	12,8841	12,88409873
75,57116	7,211103	6,78233	6,782329983
80,9753	13,2665	14,07125	13,26649916
75,93418	10,44031	3,872983	3,872983346
70,81666	11,57584	11,22497	11,22497216
76,09862	10	8,485281	8,485281374
80,93825	12,24745	10,0995	10,09950494
………….	………….	………….	………….
75,76279	11,35782	7,937254	7,937253933
	Total Range		**1510,4528**

From the table … it can be seen that the results of the 2nd iteration Medoid distance will be used to determine the Cost value. Cost Value =1510.4528

g. Finding the value of S

The S value is obtained by reducing the Cost value in the new iteration to the initial iteration. If the value of S < 0 then the processing is continued by using the new medoid center value. If the S value > 0 or the new iteration Cost value is greater than the old iteration Cost value, the process is terminated. So that the value of S is obtained:

S = New Cost Value – Old Cost Value = 1542.881135-1510.4528 = 32.428335

Because New Cost Value > Old Cost Value, the iteration is stopped and the last iteration is in Iteration 1.

h. Looking for Clusters or Grouping

To determine the Cluster, it is obtained by determining the closest Cluster 1, 2 and 3 values to the Distance Value so that the last iteration Cluster is obtained as follows:

Table 6: Last Iteration Cluster

NO	Students Name	Cluster
1	ABDUL HAMID MUKHTAMRIN	2
2	AHMAD FIRMANSYAH HARAHAP	1
3	ANDI	2
4	ANDI RAHMAD DAULAY	3
5	ANDINI SAPITRI	1
6	BOY SERGIO LUMBAN TOBING	1
7	BUDI PRAYOGI	2
8	BUNGA SALEHA LUBIS	2
9	CECILIA PRAMUDITHA	3
10	CLAUDYA AZHARI LUBIS	3
11	DAHWIYAH SYAH PUTRI SIMANJUNTAK	1
12	DERMAWAN RIZKY SIDABUTAR	1
13	ELCA ARIO ANJATRI SIMBOLON	3
14	GABRIEL FAGA SIHOMBING	1
15	GILBERT WISUDAWAN BATUBARA	1
...	....	...
176	ZHAFFRAN RULINSANI	1

3.2. Implementation of the K-Medoids Algorithm with RapidMiner

The results obtained from processing the K-Medoids Algorithm on RapidMiner are as follows:
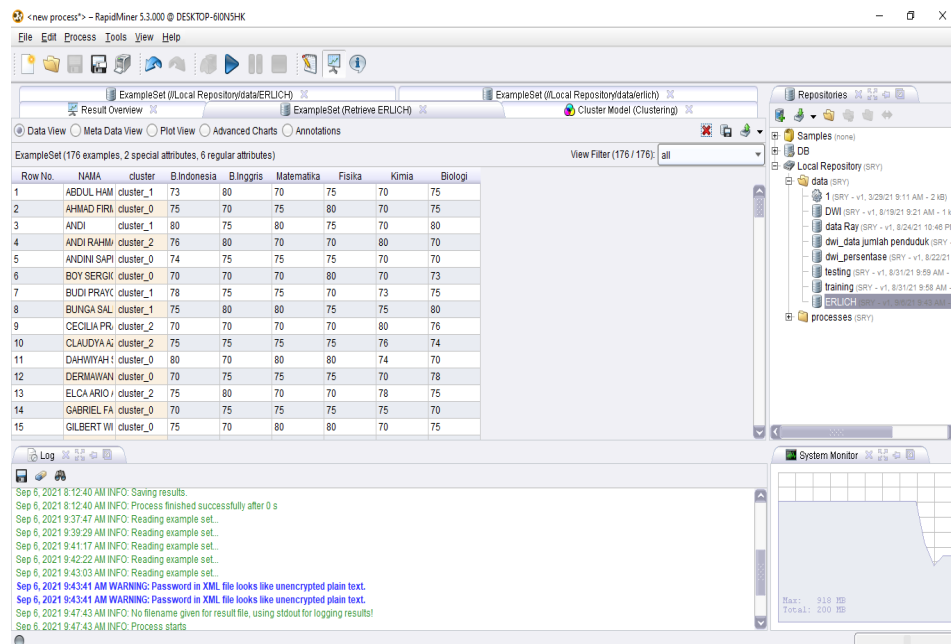

Figure 2: RapidMiner Processing Results

Figure 2 describes the results of RapidMiner processing by displaying Cluster Model with Cluster 0 as Low Cluster with 7 items, Cluster 1 as Medium Cluster with 1 items, and Cluster 2 as High Cluster with 4 items.

4. Conclusion

The results obtained in the calculation of K-Medoids Grouping the Level of Understanding of High School 5 Students on Subjects Cluster 0: Very Understanding with a total of 80 students, Cluster 1: Understanding with a total of 56 students, and Cluster 2: Not understanding with a total of 40 students. The process is carried out in 2 iterations to get good results.

References

[1]   . Jo Shan Fu and J. S. Fu, "ICT in Education : A Critical Literature Review and Its Implications," *Int. J. Educ. Dev. Using Inf. Commun. Technol.*, vol. 9, no. 1, pp. 112–125, 2013.

[2]   S. Defiyanti, M. Jajuli, and N. Rohmawati, "K-Medoid Algorithm in Clustering Student Scholarship Applicants," *Sci. J. Informatics*, vol. 4, no. 1, pp. 27–33, 2017, doi: 10.15294/sji.v4i1.8212.

[3]   Nurhayati, N. S. Sinatrya, L. K. Wardhani, and Busman, "Analysis of K-Means and K-Medoids's Performance Using Big Data Technology," *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1–5, 2019, doi: 10.1109/CITSM.2018.8674251.

[4]   D. F. Pramesti, Lahan, M. Tanzil Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas(Hotspot)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 723–732, 2017.

[5]   A. D. Andini and T. Arifin, "Implementasi Algoritma K-Medoids Untuk Klasterisasi Data Penyakit Pasien Di Rsud Kota Bandung," *J. RESPONSIF Ris. Sains …*, vol. 2, no. 2, pp. 128–138, 2020.

[6]   A. Saxena *et al.*, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, 2017, doi: 10.1016/j.neucom.2017.06.053.

[7]   T. Bimantoro and A. K. Wardhani, "Implementasi Algoritma Partitioning Around Medoids Dalam Pengelompokan Restoran," *Indones. J. Technol. Informatics Sci.*, vol. 2, no. 1, pp. 33–36, 2020, doi: 10.24176/ijtis.v2i1.5651.

[8]   N. Hidayati, A. I. Rizmayanti, C. B. S. Dewi, R. Fatmasari, and W. Gata, "Penerapan Algoritma Klasterisasi dan Klasifikasi pada Tingkat Kepentingan Sistem Pembelajaran di Universitas Terbuka," *Swabumi*, vol. 8, no. 2, pp. 134–142, 2020, doi: 10.31294/swabumi.v8i2.8385.

[9]   S. Liu and M. D'Aquin, "Unsupervised learning for understanding student achievement in a distance learning setting," *IEEE Glob. Eng. Educ. Conf. EDUCON*, pp. 1373–1377, 2017, doi: 10.1109/EDUCON.2017.7943026.

[10]   B. Wira, A. E. Budianto, and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiwa Baru Tahun 2018 Di Universitas Kanjuruhan Malang," *RAINSTEK J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 53–68, 2019, doi: 10.21067/jtst.v1i3.3046.

[11]   A. M. H. Pardede *et al.*, "Implementation of Data Mining to Classify the Consumer's Complaints of Electricity Usage Based on Consumer's Locations Using Clustering Method," in *Journal of Physics: Conference Series*, 2019, vol. 1363, no. 1, doi: 10.1088/1742-6596/1363/1/012079.

[12]   W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia," *Ilk. J. Ilm.*, vol. 13, no. 1, pp. 31–35, 2021, doi: 10.33096/ilkom.v13i1.763.31-35.

[13]   S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.

[14]   S. Samudi, S. Widodo, and H. Brawijaya, "The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic," *SinkrOn*, vol. 5, no. 1, p. 116, 2020, doi: 10.33395/sinkron.v5i1.10649.

[15]   D. P. Sari and B. Budyanra, "The Risk Factor that Affect Children Diarrhea in The Island of Java 2013 (Riskesdas 2013 Data Analysis)," *J. Educ. Heal. Community Psychol.*, vol. 6, no. 1, p. 1, 2017, doi: 10.12928/jehcp.v6i1.6615.

[16]   Syahrul dkk, "Teknologi informasi dan pendidikan," *Al-Manar (Edisi 1)*, vol. 12, no. 2, pp. 1–7, 2004.

[17]   E. H. S. Atmaja, "Implementation of k-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta," *Int. J. Appl. Sci. Smart Technol.*, vol. 1, no. 1, pp. 33–44, 2019, doi: 10.24071/ijasst.v1i1.1859.

[18]   A. Hermawati, S. Jumini, M. Astuti, F. Ismail, and R. Rahim, "Unsupervised Data Mining with K-Medoids Method in Mapping Areas of Student and Teacher Ratio in Indonesia," *TEM J.*, vol. 9, no. 4, pp. 1614–1618, 2020, doi: 10.18421/TEM94-37.

[19]   A. Bhatti, J. Kim, and R. Li, "Implementation of Image Quality and Design Time for Block-based Lossy VQ Image Compression using K-Means and K-Medoids Algorithm in Spatial Domain," *Int. J. New Technol. Res.*, vol. 3, no. 12, p. 263166, 2017.