International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 1, Number 2, July 2022, *Page. 294 - 301*
Email : editorijhess@gmail.com

Application Of Data Mining Classification Of Student Ability In Learning Using The K-Means Clustering Algorithm Method (Case Study : Sd Negeri 056029 Karya Utama)

Ika Indah Rahayu [1)], Yani Maulita[2)], Husnul Khair[3)]
[1,2,3)]STMIK Kaputama Binjai Sumatera Utara

*Corresponding Author
Email : ikaindah2710@gmail.com

Abstract
The high level of student success and the low level of student failure is a quality of the education world. The world of education is currently required to have the ability to compete by utilizing all resources owned. In addition to facilities, infrastructure and human resources, information systems are one of the resources that can be used to improve competency skills. Data mining is a process of data analysis to find a dataset of data sets. Data mining is able to analyze large amounts of data into information that has meaning for decision supporters. One process of data mining is clustering. Attributes used in the grouping of student achievement are Name, Extracurricular, Value which include UAS Value, . The case study of 20 students with distance calculation using manhattan distance, chbychep distance and euclidian distance yielded 67% accuracy. Keywords: data mining, clustering, k-means, student achievement

Keywords: Data Mining, Clustering, Student Achievement

INTRODUCTION

Learning is a set of actions designed to support the student's learning process, taking into account the extreme events that play a role in the series of internal events that take place experienced by students.

Winkel in Siregar and Nara, (2014:12). The online KBBI (October 30, 2018 12:09 WIB) defines learning as the process, method, act of making people or living things learn. The high and low quality of education is influenced by several factors. Factors that affect education in schools, can come from students, teachers, facilities and infrastructure, and can also come from the environment. According to Eko Prasetyo (2012, p 2) Year (2006) Knowledge discovery and data mining (KDD) is a computer-assisted process to explore and analyze large amounts of data sets and extract useful information and knowledge. There are several different approaches that are classified as information/knowledge seeking techniques in KDD.

Existing conventional decision-making methods cannot handle very large amounts of data. This encourages the emergence of a new branch of science to overcome the problem of extracting important information from data collections, which is called data mining. As the number of student data continues to increase every year, the number of student data continues to increase so that the accumulation of data that has not been processed optimally to explore new information and knowledge through patterns formed from the accumulation of the data.

This increasing amount of data is a number of techniques or methods to process it into information and knowledge that can be used as a consideration for educators in the policy and decision-making process as well as an early warning for certain students based on the results of grouping low achievements potential for non-graduate students. Several studies that have been conducted have conducted modeling regarding rule modeling in predicting student academics, evaluating student academic performance in the 2nd year and classified in the category of students who can graduate on time or not, the concept of clustering with data patterns that have been arranged, Cluster technique makes tropical climate data grouping in the northern Indian Ocean.

RESEARCH METHODS

This research methodology was conducted to find something systematic by using scientific methods and applicable sources. With this process, it can provide good and appropriate research results.

There are several stages of the research methodology used in this study in order to obtain data that can be tested and analyzed for truth.

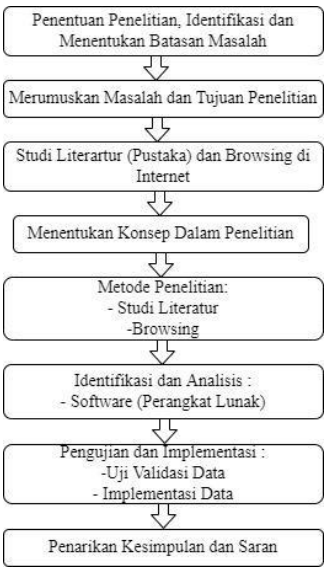

Figure 1. Research Workflow

Based on the picture above, it can be seen that there are several stages in completing research, namely as follows:

a. Problem Formulation
b. Theory
c. of Data Collection Data
d. Analysis
e. Testing and Implementation
f. Evaluation

Data Mining

mining is the process of looking for patterns or information in selected data using certain techniques or methods. Techniques, methods or algorithms in data mining vary widely. The selection of the right method or algorithm is very dependent on the overall goals and processes of KDD

According to eko prasetyo (2012, p 2) Year (2006) Knowledge discovery and data mining (KDD) is a process assisted by computers to explore and analyze large amounts of data sets and extract useful information and knowledge.

Clustering

According to R. Muliono (Zulfikar Sembiring, 2019, Vol. 4) *Clustering* on a data is a stage to classify data sets whose class attributes have not been described, the concept of clustering is to maximize and minimize intra-class similarities. For example, there is a set of objects, the first process can be clustered into several sets of classes, then it becomes a regular set so that it can be derived based on certain classification groups. Cluster can also be interpreted as a group. Then the clustering analysis will basically produce a number of clusters (groups).

Equality

According to the journal Technology and Open Source by M. Hasyim Siregar, S.Kom., M.Kom (2018) 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 computation time. group (cluster) and each object of observation is owned by a group with the closest mean. In determining the value of the centroid for the start of the iteration, the initial value of the centroid is done randomly.

- Determine the number of clusters
- Allocate objects into clusters randomly
- Calculate the sample centroids in each cluster To calculate the distance between objects and centroids can use Euclidian Distance.

$$Euclidean\ (X,Y,Z)\ \sqrt{\sum(X1-X2)^2+\overline{(Y1-Y2)^2+\ (Z1-Z2)^2}}$$

- Allocate each data to the nearest centroid.
- Return to step 3 if there is still data moving clusters or there is still a change in the centroid value, some are above the specified threshold value or if the change in the value of the objective function used is above the specified threshold value.

System Analysis

System analysis is very important in research. This is done with the aim of digging up data and processing it as information for making decisions. System analysis is a very critical and very important stage, because errors in this stage will cause errors in the next stage. The main task of system analysis at this stage is to find the weaknesses of the running system so that improvements can be proposed.

RESULTS AND DISCUSSION

Clustering Calculations The

Grouping in this study was carried out by collecting data on the classification of students' abilities in learning. The following data will be sampled.

TABLE I. DATA TO BE PROCESSED

No	*Name*	*UAS*	*Skill*	*Ability*
1	*AHF*	*82.76*	*85.67*	*High*
2	*AA*	*90.55*	*83.67*	*High*
3	*AZ*	*73.25*	*70.97*	*Low*
4	*DF*	*87.75*	*86.33*	*High*
5	*EF*	*90.33*	*78.88*	*Medium*
6	*FAF*	*87.52*	*83.33*	*Medium*
7	*HRR*	*74.57*	*75.45*	*Low*

8	***HI***	***78.88***	***83.25***	***Medium***
9	***CI***	***80.11***	***81.54***	***Medium***
10	***KN***	***76.76***	***74, 23***	***Low***
11	***MDP***	***79.52***	***78.89***	***Medium***
12	***MAP***	***82.67***	***90.78***	***High***
13	***MF***	***80.01***	***79.12***	***Medium***
14	***MAS***	***76.05***	***80.67***	***Medium***
15	***NAR***	***79.11***	***75.12***	***Medium***
16	***RABH***	***76 ,06***	***80.05***	***Medium***
17	***RFDG***	***85.76***	***90.01***	***High***
18	***TAA***	***77.62***	***76.67***	***High***
19	***XNP***	***76.75***	***84.33***	***Medium***
20	***ZA***	***79.67***	***75.76***	***Medium***

Then initialize each variable in the clustering process,

Initialization

From the existing data, data initialization can be carried out according to the needs of the variables as follows:

a. Initialization of UAS Criteria The

following is a table of initialization of UAS criteria which can be seen in the table below:

TABLE II. INITIALIZATION OF AREA CRITERIA

Code	***Ability***
1	***65-75***
2	***76-85***
3	***86-95***

b. Initialization of Skills Criteria The

following is a table of initialization of Skills criteria which can be seen in the table below:

TABLE III. INITIALIZATION OF SKILL CRITERIA

Code	***Ability***
1	***65-75***
2	***76-85***
3	***86-95***

c. Initialization of Ability Criteria The
following is a table of initialization of ability criteria which can be seen in the table below:

TABLE IV. INITIALIZATION OF ABILITY CRITERIA

Code	*Ability*
1	*Low*
2	*Medium*
3	*High*

Then perform data transformation for each UAS variable (X), Skills (Y), Ability (Z).

Cluter into 3 groups (K=3) and determine the center point of the centroid. The clustering calculation process is as follows:

Iteration I

Centroid 1 = (2,2,3) is taken randomly from data 1
*Centroid*2 = (3,2,3) is taken randomly from data 2
Centroid 3 = (2, 3,3) taken randomly from the data 3
Note:
The *centroid* is done randomly.

TABLE V. RESULTS OF GROUP 1 DETERMINATION

No	*Name*	*UAS*	*Skill*	*Ability*	*Distance From C1*	*Distance From C2*	*Distance From C3*	*Group*
		X	*Y*	*Z*				
1	*AHF*	*2*	*2*	*3*	*0*	*1*	*2,45*	*1*
2	*AA*	*3*	*2*	*3*	*1*	*0*	*3*	*2*
3	*AZ*	*1*	*1*	*1*	*2.45*	*3*	*0*	*3*
4	*DF*	*3*	*3*	*3*	*1.41*	*1*	*3.46*	*2*
5	*EF*	*3*	*3*	*3*	*1.41*	*1*	*2.45*	*2*
6	*FAF*	*3*	*2*	*2*	*1.41*	*1*	*2.45*	*2*
7	*HRR*	*1*	*1*	*1*	*2.45*	*3*	*0*	*3*
8	*HI*	*2*	*2*	*2*	*1.41*	*1.41*	*1.73*	*1*
9	*IK*	*2*	*2*	*2*	*1.41*	*1.41*	*1.73*	*1*
10	*KN*	*2*	*1*	*1*	*1.73*	*2.45*	*1*	*3*
11	*MDP*	*2*	*2*	*2*	*1.41*	*1.41*	*1.73*	*1*
12	*MAP*	*1*	*1*	*1*	*2.45*	*3*	*0*	*3*
13	*MF*	*2*	*2*	*2*	*1.41*	*1.41*	*1.73*	*1*
14	*MAS*	*2*	*2*	*2*	*1.41*	*1, 41*	*1.73*	*1*
15	*NAR*	*2*	*1*	*2*	*1.41*	*1.73*	*1.41*	*1*
16	*RABH*	*2*	*2*	*2*	*1.41*	*1.41*	*1.73*	*1*
17	*RFDG*	*3*	*3*	*3*	*1.41*	*1*	*2.45*	*2*
18	*TAA*	*2*	*2*	*3*	*0*	*1*	*2.45*	*1*
19	*XNP*	*2*	*2*	*2*	*1.41*	*1.41*	*1.73*	*1*
20	*ZA*	*2*	*2*	*2*	*1.41*	*1.41*	*1.73*	*1*

Group based on the minimum distance to the nearest centroid, namely:
If the distance the shortest distance is in C1 then the data is entered in group 1
If the shortest distance is in C2 then the data is entered in group 2
If the shortest distance is in C3 then the data is entered in group 3
Old Group = {0,0,0,0,0 ,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
New Group = {1,2,3,2,2,2,3 ,1,1,3,1,3,1,1,1,2,1,1,1}
There is a change in the group, then proceed to the next iteration.
For group 1 there are 11 data, namely:
C1 = (2+2+2+2+2+2+2+2+2+2+2) / 11 = 2
C2 = (2+2+2+2+2+ 2+1+2+2+2+2) / 11 = 1.90

C3 = (3+2+2+2+2+2+2+2+2+3+2) / 11 = 2.18

For group 2 contains 5 data, namely:

C1 = (3+3+3+3+3) / 5 = 3

C2 = (2+3+3+2+3) / 5 = 2.6

C3 = (3+3+ 3+2+3) / 5 = 2.8

For group 3 there are 4 data, namely:

C1 = (1+1+2+1) / 4 = 1.25

C2 = (1+1+1+1) / 4 = 1

C3 = (1+1+1+1) / 4 = 1

Iteration II

Centroid 1 = (2, 1,90, 2,18)

Centroid 2 = (3 , 2.6 ,2.8)

Centroid 3 = (1,5, 1 , 1)

TABLE VI. RESULTS OF DETERMINATION OF GROUP 2

No	*Name*	*UAS*	*Skill*	*Ability*	*Distance From C1*	*Distance From C2*	*Distance From C3*	*Group*
		X	*Y*	*Z*				
1	*AHF*	*2*	*2*	*3*	*0.82*	*1.18*	*2.29*	*1*
2	*AA*	*3*	*2*	*3*	*1.29*	*0.63*	*2.69*	*2*
3	*AZ*	*1*	*1*	*1*	*1.78*	*3.13*	*0.5*	*3*
4	*DF*	*3*	*3*	*3*	*1.69*	*0.44*	*3.20*	*2*
5	*EF*	*3*	*3*	*3*	*1.69*	*0.44*	*3.20*	*2*
6	*FAF*	*3*	*2*	*2*	*1, 02*	*1*	*2, 06*	*2*
7	*HRR*	*1*	*1*	*1*	*1.78*	*3.13*	*0.5*	*3*
8	*HI*	*2*	*2*	*2*	*1, 20*	*1.18*	*1.50*	*1*
9	*IK*	*2*	*2*	*2*	*1, 20*	*1.18*	*1 .50*	*1*
10	*KN*	*2*	*1*	*1*	*1.48*	*2.60*	*0.5*	*3*
11	*MDP*	*2*	*2*	*2*	*1.41*	*1.41*	*1.73*	*1*
12	*MAP*	*1*	*1*	*1*	*2.44*	*3*	*0*	*3*
13	*MF*	*2*	*2*	*2*	*1 , 20*	*1.18*	*1.50*	*1*
14	*MAS*	*2*	*2*	*2*	*1, 20*	*1.18*	*1.50*	*1*
15	*NAR*	*2*	*1*	*2*	*0.91*	*1.89*	*1.11*	*1*
16	*RABH*	*2*	*2*	*2*	*1, 20*	*1.18*	*1.50*	*1*
17	*RFDG*	*3*	*3*	*3*	*1.69*	*0.44*	*3.20*	*2*
18	*TAA*	*2*	*2*	*3*	*0.82*	*1.18*	*2.29*	*1*
19	*XNP*	*2*	*2*	*2*	*1.41*	*1.41*	*1.73*	*1*
20	*ZA*	*2*	*2*	*2*	*1, 20*	*1,18*	*1,50*	*1*

Group based on the minimum distance to *centroid* , namely:

Old Group = {1,2,3,2,2,2,3,1,1,3,1,3 ,1,1,1,1,2,1,1,1}

New Group = {1,2,3,2,2,2,3,1,1,1,1,1,1,1,1 ,1,2,1,1,1}

Because in the 1st and 2nd iterations the *cluster* does not change or there are similarities, the iteration calculation is stopped.

Graph Results

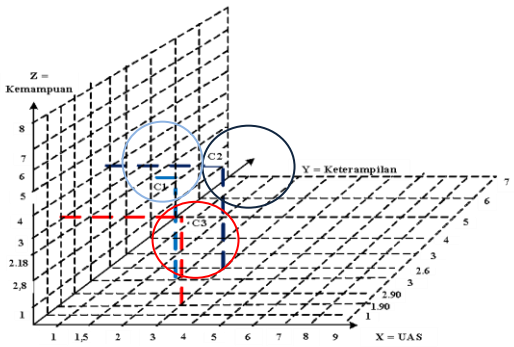

Figure 1. Clustering Graph of Students' Ability in Learning

Description:

	X	Y	Z
Centroid 1 :	(2	1.90	2.18)
Centroid 2 :	(3	2.6	2.8)
Centroid 3:	(1.25	1	1)

CONCLUSION

In the final result, it can be seen that the grouping in SD Negeri 056029 Karya Utama with 442 data can be divided into 3 groups, namely for *centroid 1* (1.06 1.23 1.31) The data group classification of students' abilities in learning in *group* UASis 65-75 with Skill 65-75 and Ability is Low. And for *Centroid 2* (2.98 2.89 2.92) The data group for the classification of students' abilities in learning in the UAS group is 86-95 with 86-95 Skills and High Ability. And for *Centroid 3* (2.21 2.11 2.00) the data group classification of students' abilities in learning in the UAS group is 76-85 with 76-85 Skills and Medium Ability..

REFERENCES

Angraini, G. (2014). Analysis of Science Literacy Ability of Class X High School Students in Solok City. *Proceedings of the 2014 Mathematics and Sciences Forum*, 161–170.

Asroni, A., Fitri, H., & Prasetyo, E. (2018). Application of the Clustering Method with the K-Means Algorithm in Grouping Prospective New Student Data at the University of Muhammadiyah Yogyakarta (Case Study: Faculty of Medicine and Health Sciences, and Faculty of Social and Political Sciences). *Engineering Universe*, *21*(1), 60–64. https://doi.org/10.18196/st.211211

Darmawan, EW (2018). Analysis of Learning Media Development Needs According to Guided Discovery. *Proceedings of the National Seminar on Ethnomatnesia*, 222–224. http://jurnal.ustjogja.ac.id/index.php/etnomatnesia/article/view/2318

Elizawati, N., & Lesmana, LS (2017). Analysis of Grade X Student Report Cards in the Multimedia Department on Interest in Productive Lessons in Grade XII to Determine Student Competence with the K-Means Algorithm Clustering Method (Case Study at SMKN 4 Padang). *Journal of Applied Computers*, *3*(2), 133–148. http://jurnal.pcr.ac.id

Gais, Z., & Afriansyah, EA (2017). Analysis of Students' Ability in Solving High Problems. *Mosharafa*, *6*, 255–266.

Informatics, PST, & Majapahit, UI (2019). Utilization of Knowledge Data Discovery (KDD) in Badminton Athletes' Game Patterns. *Explore IT : Journal of Science and Informatics Engineering Applications*, *11*(1), 1–6. https://doi.org/10.35891/explorit.v11i1.1467

Mardi, Y. (2017). Data Mining: Classification Using the C4.5 Algorithm. *Edik Informatics*, *2*(2), 213–219. https://doi.org/10.22202/ei.2016.v2i2.1465

Mirnawati, M., & Firman, F. (2019). Application of Clustering Techniques in Developing Essay Writing Skills for Class IV Students of MI Islamic Boarding School Datuk Sulaiman Palopo. *Journal of Teacher Studies and Learning*, *2*(2), 165–177. https://doi.org/10.30605/jsgp.2.2.2019.1373

Muliono, R., & Sembiring, Z. (2019). Data Mining Clustering Using the K-Means Algorithm for Clustering the Level of Lecturer Teaching Tridharma. *CESS (Journal of Computer Engineering, Systems and Science)*, *4*(2), 2502–2714.

Poerwanto, B., & Fa'rifah, RY (2016). K-Means Cluster Analysis in Student Ability Grouping. *Scientific Pinisi Journal*, *2*(2), 92–96.

Ridho, S., Ruwiyatun, R., Subali, B., & Marwoto, P. (2020). Analysis of Students' Critical Thinking Skills Subject Classification of Materials and Its Changes. *Journal of Science Education Research*, *6*(1), 10–15. https://doi.org/10.29303/jppipa.v6i1.194

Sirait, W., Defit, S., & Nurcahyo, GW (2019). K-Means Algorithm for Clustering Students' Final Projects Based on Expertise. *Journal of Information Systems and Technology*, *1*(3), 25–30. https://doi.org/10.35134/jsisfotek.v1i3.6

Siregar, MH (2018). Data Mining Clustering Sales of Building Tools Using the K-Means Method (Case Study at Adi Building Stores). *Journal of Technology And Open Source*, *1*(2), 83–91. https://doi.org/10.36378/jtos.v1i2.24

Sulistiyawati, A., & Supriyanto, E. (2021). Implementation of the K-means Clustering Algorithm in Determining Excellent Class Students. *Journal of Tekno Kompak*, *15*(2), 25. https://doi.org/10.33365/jtk.v15i2.1162