Journal of Artificial Intelligence and Engineering Applications
Website: https://ioinformatic.org/

15th October 2021. Vol. 1. No. 1

Classification Analysis of Student Ability in Learning Using Clustering Method at SMA Tunas Pelita

Nurhayati[1]*, Juliana Naftali Sitompul[2], Tri Kartika Sari[3]

*[*1,2,3]Program of Technical Information, STMIK Kaputama*
**Corresponding author E-mail:* nurhayatiazura059@gmail.com

Abstract

This study aims to classify the assessment of the learning process at SMA Tunas Pelita Binjai T.A. 2018/2019 based on the average grade X, additional subjects applied technology, and student absenteeism classified using Matlab.The data is processed based on learning grouping as much as 2 clusters with different centroids, namely for cluster 1 the average value of even and odd semesters for class X (85.0), additional subjects of applied technology (86.3) and student attendance (2.4) and cluster 2 the average grades of odd-even semesters for class X (68.2), additional subjects of applied technology (70.3) and student attendance (2.4). In the final result, it can be seen that the grouping of learning at SMA Tunas Pelita Binjai with 100 data can be divided into 2 groups, namely group 1 with 62 data with an average value of odd and even semesters and high additional applied technology and student absenteeism. low grades are classified as students with good grades and group 2 as many as 38 data with an average value of odd, even semesters and low values of applied technology and high student absenteeism belonging to students who have poor grades.
Keywords*:data mining, k-means, class X average value*

1. Introduction

SMA Tunas Pelita aims to educate and strengthen attitudes and personalities in everyday life through the implementation of ability-based learning so that students develop according to their talents and abilities. In achieving this goal, the school provides students with expertise in accordance with the study program taken and facilitates the development of these students' abilities. One of the programs provided by SMA Tunas Pelita in developing students' abilities is the classification of students' abilities. Classification is carried out when students will continue their studies from class X (ten) to class XI (eleven) and from class XI (eleven) to class XII (twelve) at SMA Tunas Pelita. To determine the ability of students, including students with good or bad grades.

At SMA Tunas Pelita, the classification process is determined based on the average grade X (ten) grades of odd and even semester report cards, Additional Subjects for Applied Technology for grade X (ten) and Student Attendance. Currently the classification process at SMA Tunas Pelita still does not use the system because it is not yet computerized, so problems arise in the classification process. The problem is that the school determines where the students will be placed in the classification of the learning process. Students will get the results of the scores they got during the first semester and include student data into the classification of abilities.

The study was conducted to analyze the Data Mining Method to help classify students' abilities. The method used is the Clustering method with the K-Means algorithm. The data used for the analysis process of the Data Mining Method is value data. The value taken as the basis for the classification process is the average value of semester I and II in class X (ten), Applied Technology Subjects class X (ten) and Student Attendance. The software used is MatLap (Matrix Laboratory) version 7.7.0, DreamWeaver CS3 and PHP MySQL.

2. Material

2.1 Data Mining

Data mining is a process of extracting and discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Aside from the raw analysis step, it also involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating

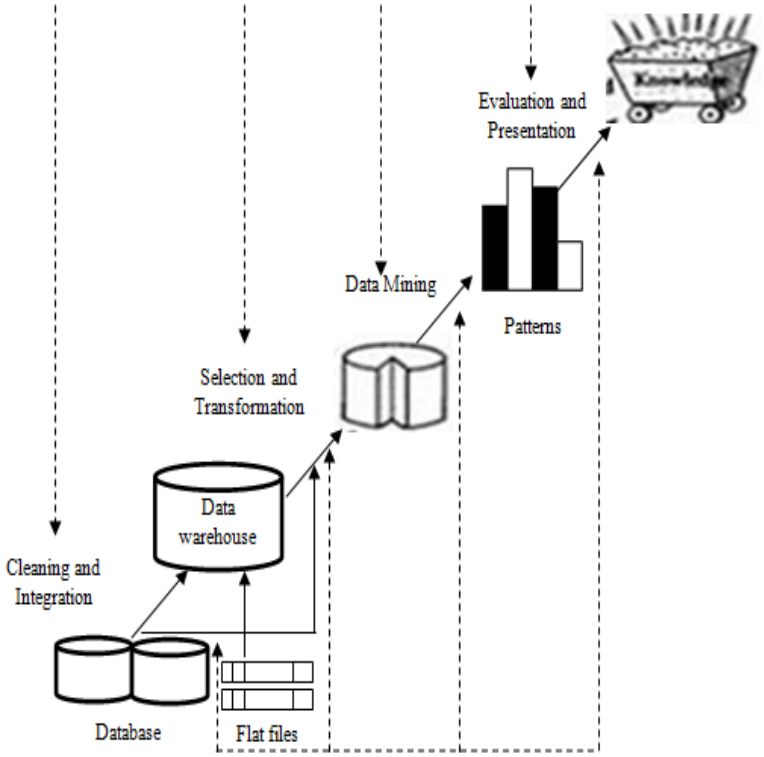

Fig. 1 Data Mining Process

2.2 Algoritma K-Means

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of theclusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more.

2.3 Student Learning Outcomes

Mark Battersby (1999) of the Learning Outcomes Network explains that learning outcomes are more than simply several sentences appended to existing lesson plans or curricula; instead, the development of learning outcomes and their use within a unit of instruction shapes learning and assessment activities and can enhance student engagement and learning.

Because of their ability to benefit many groups in postsecondary education, the development of learning outcomes has become an increasing priority for instructors and institutions over the course of the last decade. Establishing a focus on integrated, generalizable, and transferable skills complements contemporary demands on graduates and builds a foundation for lifelong learning. As government and public attention on the products of higher education increases, learning outcomes help to define the goals and essential aspects of higher education within the institution, to students, and to the general public.

3. Methodology

3.1. Research Stages

Systematic steps are organized as illustrated in Fig. 3 to make it easier in performing research.

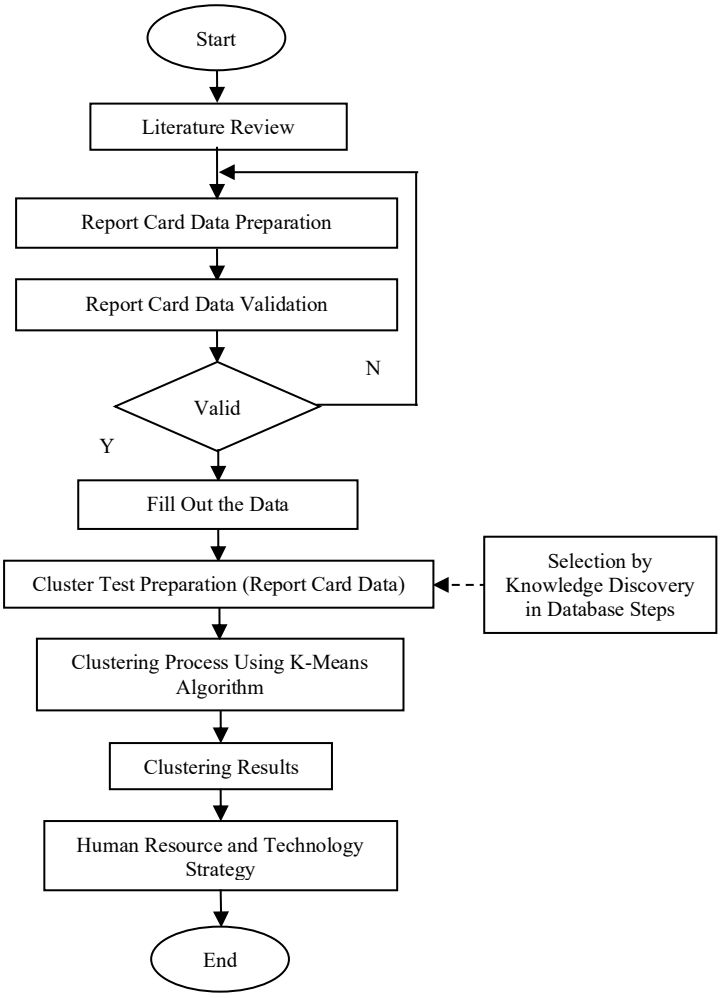

Fig. 2: Research steps

Data begins to be inputted from into Microsoft Excel. If the input value is valid (true), then the next process is imported (entered) in matlab (matrix Laboratory), if the input value is not valid then return to re-check the value data. After being entered into MATLAB, the results from the MATLAB must be analyzed using the k-means method. Finished

3.2.Training Data

Table 1 : Calculation Results on Reports

Student's name	Average Semester Report Card I/II Class X	Semester Applied Technology Supplementary Subjects.I/II Class X	Student Attendance
A	80.5	76.5	2
B	75.4	65.1	1
C	78.2	68.2	2
D	82.5	71.1	4
E	63.4	61.4	4
F	71.9	70.5	3
G	78.2	75.8	3
H	84.3	81.1	1
I	75.3	75.4	2
J	76.5	75.5	2

Table 2 Result of Group Determination Example

Student's name	Average Semester Report Card I/II Class X	Semester Applied Technology Supplementary Subjects I/II Kls X	Student Attendance	Distances		Grup
	X	Y	Z	C1	C2	
A	80.5	76.5	2	0	2.6	1

B	75.4	65.1	1	12.5	11.1	2
C	78.2	68.2	2	8.6	7.8	2
D	82.5	71.1	4	6.0	6.5	1
E	63.4	61.4	4	22.9	21.2	2
F	71.9	70.5	3	10.5	8.2	2
G	78.2	75.8	3	2.5	0	2
H	84.3	81.1	1	6.0	8.2	1
I	75.3	75.4	2	5.3	3.0	2
J	76.5	75.5	2	4.1	1.8	2

Table 3 Results of the 2nd Iteration Example Group Determination

Student's name	Average Semester Report Card I/II Class X	Semester Applied Technology Supplementary Subjects I/II Kls X	Student Attendance	Distances		Grup
	X	Y	Z	C1	C2	
A	80.5	76.5	2	1.9	9.0	1
B	75.4	65.1	1	13.2	5.3	2
C	78.2	68.2	2	2.0	9.1	1
D	82.5	71.1	4	5.2	8.5	1
E	63.4	61.4	4	23.2	13.9	2
F	71.9	70.5	3	12.0	2.3	2
G	78.2	75.8	3	4.4	7.0	1
H	84.3	81.1	1	5.4	15.2	1
I	75.3	75.4	2	7.2	5.4	2
J	76.5	75.5	2	6.0	5.9	2

Table 4 Results of the 3rd Iteration Example Group Determination

Student's name	Average Semester Report Card I/II Class X	Semester Applied Technology Supplementary Subjects I/II Kls X	Student Attendance	Distances		Grup
	X	Y	Z	C1	C2	
A	80.5	76.5	2	2.1	3.9	1
B	75.4	65.1	1	10.8	5.3	2
C	78.2	68.2	2	6.8	5.9	2
D	82.5	71.1	4	4.0	10.4	1
E	63.4	61.4	4	21.9	12.4	2
F	71.9	70.5	3	9.7	1.3	2
G	78.2	75.8	3	2.9	6.8	1
H	84.3	81.1	1	7.6	16.6	1
I	75.3	75.4	2	5.5	6.6	1
J	76.5	75.5	2	4.3	7.3	1

Table 5 Results of the 4th Iteration Example Group Determination

Student's name	Average Semester Report Card I/II Class X	Semester Applied Technology Supplementary Subjects I/II Kls X	Student Attendance	Distances		Grup
	X	Y	Z	C1	C2	

A	80.5	76.5	2	1.2	13.3	1
B	75.4	65.1	1	11.6	3.5	2
C	78.2	68.2	2	7.9	6.3	2
D	82.5	71.1	4	5.8	11.5	1
E	63.4	61.4	4	21.8	10.2	2
F	71.9	70.5	3	9.4	4.4	2
G	78.2	75.8	3	1.4	11.3	1
H	84.3	81.1	1	7.2	19.2	1
I	75.3	75.4	2	4.3	9.7	1
J	76.5	75.5	2	3.1	10.3	1

4. Results

4.1. Clasifikation Analysis Methodology

The determination of the results of the number of centroids for each cluster / group is as follows:
Centroids. 1 = Total Group / Number of Groups 1
C1 = 5276.0000/62 = 85.0
C2 = 5354,7000 / 62 = 86.3
C3 = 153,0000/62 = 2.4
Centroid 2 = Total Group / Number of Groups 2
C1 = 2592.7000 / 38 = 68.2
C2 = 2673.6000 / 38 = 70.3
C3 = 92.00000 / 38 = 2.4
The calculation of the distance of the object to the centroid is carried out 5 repetitions where the cluster is determined as much as 2, namely cluster 1 and cluster 2, so the total iteration is 24 times, namely:
5 iterations, total sum of distances = 788.2
4 iterations, total sum of distances = 785.2
5 iterations, total sum of distances = 788.2
5 iterations, total sum of distances = 788.2
5 iterations, total sum of distances = 788.2

Result analysis :

1. For cluster / group 1 consists of 62 data. When viewed from cluster 1, it can be analyzed that the high average grades of odd and even semesters of class X, the value of additional applied technology subjects and low student absenteeism, this shows that the students in this group are classified as students who still have grades. the good one.

2. . For cluster/group 2 consists of 38 data. When viewed from cluster 2, it can be analyzed that the low average scores for odd and even semesters, the value of additional applied technology subjects and high student absenteeism will affect students. This indicates that students in this group belong to students who have high grades which is not good.

5. Conclusion

Based on the results of research and discussions that have been carried out, conclusions can be drawn including: 1.The data is processed based on learning grouping as much as 2 clusters with different centroids, namely for cluster 1 the average value of even and odd semesters for class X (85.0), additional subjects of applied technology (86.3) and student attendance (2.4) and cluster 2 the average grades of odd-even semesters for class X (68.2), additional subjects of applied technology (70.3) and student attendance (2.4). 2. In the final result, it can be seen that the grouping of learning at SMA Tunas Pelita Binjai with 100 data can be divided into 2 groups, namely group 1 with 62 data with an average value of odd and even semesters and high additional applied technology and student absenteeism. low grades are classified as students with good grades and group 2 as many as 38 data with an average value of odd, even semesters and low values of applied technology and high student absenteeism belonging to students who have poor grades.

Acknowledgement

We would like to thank seluruh staff, Dosen dan pihak sekolah di Kota Binjai for supporting the completion and publication of this research.

References

[1] Asriningtias, Y and Rodhiyah Mardhiyah.2014. Data Minung Aplication to display Information Student Graduation Rate. media.neliti.com. Jurnal Informatika Vol.8, No.1 Januari 2014

[2] Azevedo, A. and Santos, M. F. KDD, SEMMA and CRISP-DM: a parallel overviewArchived 2013-01-09 at the Wayback Machine. In Proceedings of the IADIS European Conference on Data Mining 2008, pp 182–185.

[3] Dudita, Windha Mega Pradnya. 2015. Using the K-Means Clustering to Determine the Nutritional Status of Toddlers. Jurnal Informatika, Vol. 15, No. 2, Desember 2015.

[4] *Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases" (PDF). Retrieved 17 December 2008.*

[5] *Han, Jiawei; Kamber, Micheline (2001). Data mining: concepts and techniques. Morgan Kaufmann. p. 5. ISBN 978-1-55860-489-6. Thus, data mining should have been more appropriately named "knowledge mining from data," which is unfortunately somewhat long*

[6] Piatetsky-Shapiro, Gregory; Parker, Gary (2011). "Lesson: Data Mining, and Knowledge Discovery: An Introduction". *Introduction to Data Mining*. KD Nuggets. Retrieved 30 August 2012.

[7] Pratama, Anindito Yoga and Yuhilza Hanum. 2016. Implementasi of Data Mining Techniques to Determine the Results of SMAN 99 Jakarta entry Selection for Students of SMPN 9 Jakarta Using Decision Tree. ejournal.gunadarma.ac.id. Vol 21. No. 1 2016

[8] Sibuea, Fitri Larasati and Andi Sapta. 2016. Mapping of Outstanding Students Using the K-Means Clustering Method. Jurnal.stmikroyal.ac.id. VolIV No.1 Desember 2017. hlm.85-92.