

A Comparative Study Of Classification And Clustering Model: The Effect Of Intelligence, Stress And Emotion On Student Learning

Finki Dona Marleny

Sari Mulia School of Health Sciences, Banjarmasin, South Kalimantan

finkidona@stikessarimulia.ac.id

ABSTRACT

Objective: a comparative study for effect of intelligence, stress and emotion on student learning in area department of students Sari Mulia School of Health Sciences.

Technology or Method: This research uses the Experimental method, analyzed data by grouping the existing data based on emotional intelligence, stress level and achievement index as student learning, data analysis using classification and clustering model.

Result: the result of research in this study is a comparative study k-means cluster and nearest neighbor algorithm to produce 3 groups of students based on emotional intelligence, stress level and achievement index.

Conclusion: From the comparison result in this research is to know whether student learning data can be used classification and clustering model, comparative study k-means cluster and nearest neighbor algorithm. The best algorithm can use k-means cluster.

Keywords: Classification, Clustering, intelligence, stress

I. INTRODUCTION

Education is expected to produce quality human beings, responsible and able to progress in the future. In the process of education there is not only on the provision of more adequate infrastructure facilities or improving the quality of educators [1]. Education is a fundamental to improve the quality of human resources. Factors possessed by a person in undergoing the educational process have a considerable influence as in the learning achievement process [2], where the emotional intelligence factor and stress level of a person can also affect the achievement of student achievement.

Emotional intelligence is the ability to connect with others [3]. Low emotional intelligence and stress can make students reluctant to learn, decreasing quality and academic achievement [1]. In the opinion of Philip L. Rice says that stress is an event or environmental stimulus that causes the individual to feel tense [4]. The level of stress also has many influences from the environment in which one learns and the environment outside the lesson. From a person's stress level is also known whether or not someone can control emotions.

Analysis of data for research on emotional intelligence, stress and performance index can be found in every part of the

student. But it needs a process to get useful information according to a specific purpose. This process can use data mining [5]. The goal is to know the pattern of universal patterns of existing data. Emotional intelligence and stress to this learning achievement will be analyzed and grouped so as to form groups that have a hidden similarity of data that has been analyzed. The most popular method of data mining grouping that can be used is classification and clustering [6].

Marleny's research used student data to make 3 group clusters using k-means method. From the results of research on sample data obtained 3 groups of students based on emotional intelligence, level of stress and index achievement, from the results of these studies k-means can be applied to analyze the data [7].

This research will discuss how to make the classification model with classification and clustering algorithm. Model analyze data influence of emotional intelligence and stress to student achievement toward student of Sari Mulia School of Health Sciences

By forming multiple clusters using k-means algorithm can also know the distance between the central clusters on the data to be analyzed. These results form the basis for classifying new data which then appears to be known to the group.

This research uses experimental method which has several stages: (1) data collection method, (2) initial data processing

method, then (3) proposed model, (4) test result and evaluation.

II. RESEARCH METHOD

Data

The data obtained in this study is data on students Sari Mulia School of Health Sciences Banjarmasin, with a random sample. This data is then processed in the form of column numbers that is data about the level of stress that has three levels with high, medium and low values. Data on emotional intelligence divided into moderate, high and very high values and learning achievement data or Achievement Index.

The data will be formed into several columns of numbers that are columns consisting of statements tailored to the questionnaire around emotional intelligence and stress levels.

PROPOSED MODEL

Nearest Neighbour

Nearest Neighbor does not explicit Nearest Neighbor does not explicitly compute decision boundaries itly compute decision boundaries. However, the boundaries form a subs However, the boundaries form a subset of the Voronoi diagram of et of the Voronoi diagram of the training data[8]. Each line of the segment is equidistant between two points of opposite Each line segment is equidistant between two points of opposite class[9]. The more examples that class. The more examples that are stored, the more complex the are

stored, the more complex the decision boundaries can become.

Classify a new example Classify a new example x by finding the training (x_1, y_1) that is nearest to that is nearest to x according to according to Euclidean distance[10]:

$$x - x_i = \sqrt{\sum_j (x_j - x_{ij})^2}$$

guess the class guess the class $\hat{y} - y_1$

K-Means Cluster

The proposed model is k-means cluster. In the application of k-means algorithm there are several steps are as follows [8][9]:

Function direct k-means()

Initialize k prototypes (w_1, \dots, w_k) Such that

$$w_j = i_1, j \in \{1, \dots, k\}, I \in \{1, \dots, n\}$$

Each cluster C_j , is associated with prototype w_j

Repeat

for each input vector i_l , where $l \in \{1, \dots, n\}$,

do

Assign i_l to the cluster C_{j^*} with nearest prototype w_{j^*}

(ie., $|i_j - w_{j^*}| \leq |i_j - w_j|, j \in \{1, \dots, k\}$)

for each cluster $j \in \{1, \dots, k\}$, do Update the

prototype w_j to be the centroid of all samples

currently in C_j , so that $w_j = \sum_{i_1 \in C_j} i_1 / c_j$

Compute the error function:

$$E = \sum_{j=1}^k \sum_{i_1 \in C_j} |i_1 - w_j|^2$$

Until E does not change significantly or cluster membership no longer changes

There are several distance measurement methods that can be used:

- (1) Euclidean distance
- (2) Manhattan distance
- (3) The Minkowski [10].

III. RESULT

Analyzed data by grouping existing data based on emotional intelligence, stress level and achievement index as student learning, data analysis using classification and clustering model. Here are some tables describing the results of grouping based on k-means method;

Table 1. Initial Cluster Centers

Initial	Cluster	
	1	2
Intelligent	3.00	5.00
Stress	1.00	3.00
Grade	3.53	3.25

The result was shown in table 2. The cluster was divided into 2, which means that the data will be split into three groups, with the initial value being the data center. Other data will be defined as group members, if the data has the closest distance to the center of the group.

In the interaction step there is an update of group centers, until the center of the cluster is stable. Members who have the greatest affinity at a particular center point will be recorded as members of the cluster.

Tabel 2. Iteration History^a

Iteration	Change in Cluster Centers	
	1	2
1	1.274	.835
2	.025	.046
3	.000	.003
4	9.605E-006	.000
5	1.883E-007	7.955E-006
6	3.693E-009	4.419E-007
7	7.241E-011	2.455E-008
8	1.420E-012	1.364E-009
9	2.763E-014	7.578E-011
10	1.018E-015	4.211E-012

.Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is 4.202E-012. The current iteration is 10. The minimum distance between initial centers is 2.842.

For three valid variables, we use these variables with the ANOVA method. Table 2 shows the results that the three variables present show significant differences.

Tabel 3. ANOVA

	Cluster		Error		F	Sig.
	Mean	df	Mean	Df		
	Square		Square			
intelligent	.004	1	.127	65	.031	.861
stress	22.780	1	.173	65	131.970	.000
grade	.003	1	.011	65	.249	.619

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this

and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Nearest neighbor uses the principle that data having the same characteristics will have 'nearest' in the vector space. New unknown class data can be predicted how to observe the nearest data class. The distance formula can be used to calculate the proximity of the data. This study used Euclidean distance calculation. Fig.1 show result nearest dataclass.

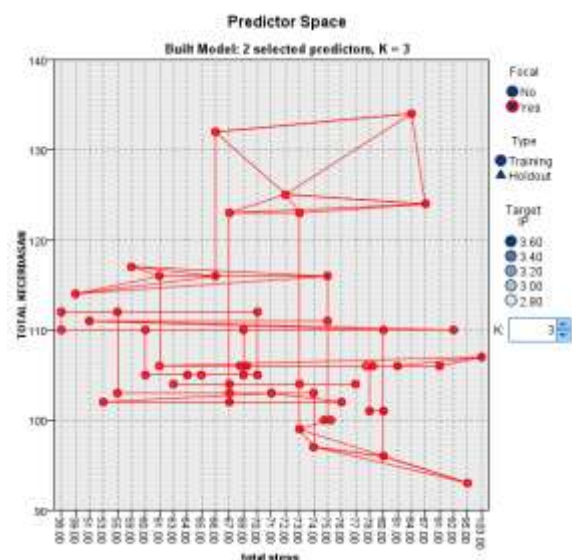


Fig.1 Predictor space NN

Details of the classification results and the parameters used are shown on Fig 2 describes the focal record of the result.

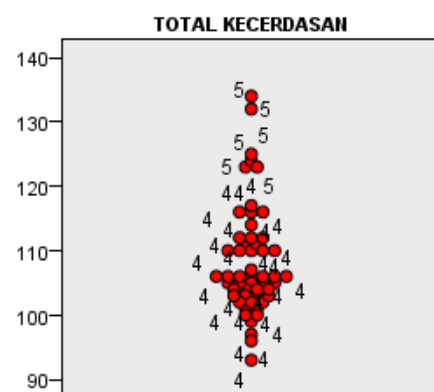


Fig.2 focal record intelligence

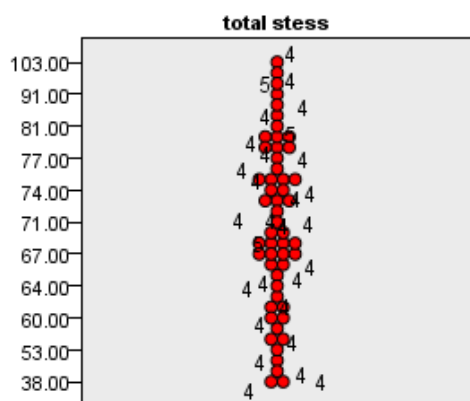


Fig.3 stress focal record stress

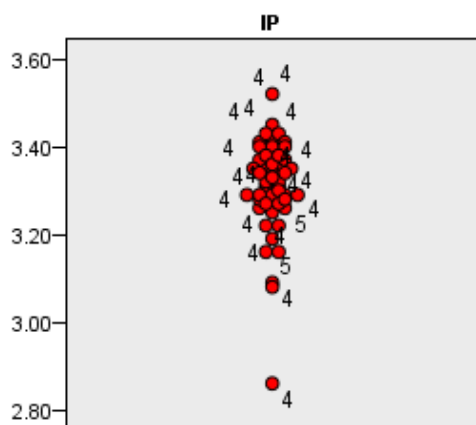


Fig.4 focal record grade

IV. EXPLANATION

In an exhaustive experimental results using the k-means and nearest neighbors method has been formed into several columns of numbers and columns consisting of statements that are adapted to questionnaires about emotional intelligence and the stress level has been adjusted. The Performance Index was used to assess how many times the best result was better than each of the others. The exhaustive results will be the basis of comparison for the heuristic ones. Future work also includes testing different distribution and size of test samples, as well as

allowing different distance measures for each space in K-means and NN parameterizations.

V. CONCLUSION

Classification techniques are used to find models for particular interests. While clustering is a data mining technique to group data based on similarities. The k-means and nearest neighbor approach can be implemented in this research. This method analyzes data that have a high similarity to compare with each other. In this case the data analyzed using k-means method is the data of emotional intelligence, stress level and achievement index student achievement. The data is grouped and it is known there are three groups; this research analyzes data influence student achievement to the level of stress and emotional intelligence.

The result of this comparison is the use of mileseas better than nearest neighbor algorithm. Because the data can be more easily be accepted for the analysis using k-means algorithm.

REFERENCE

- [1] Mulyadi H. Diagnosis Kesulitan Belajar & Bimbingan Terhadap Kesulitan Belajar Khusus. Cet kedua. Yogyakarta: Nuha Litera. 2010.
- [2] Musrofi M. Melesatkan Prestasi Akademik Siswa. Yogyakarta: PT Pustaka Insan Madani. 2010.
- [3] Daniel Goleman, *Emotional Intelligence*. New York: Bantam Dell, 2009.
- [4] Philip L. Rice, *Stress and Health*, 3rd ed.: Cole Publishing, 1999.
- [5] I.H Witten and Eibe Frank, *Data Mining*:

Practical Machine Learning Tools and Techniques., 2005.

- [6] Rui Xu and Donald C. Wunsch, "Partitional Clustering," in *Clustering*, David B. Fogel, Ed. New Jersey: Wiley Publishing, 2009, ch. 4, pp. 67-110.
- [7] Marleny FD, Junaidi HM, Mambang . Penerapan K-Means cluster untuk pengaruh kecerdasan emosi dan stres terhadap prestasi belajar mahasiswa. SEMNASTEKNOMEDIA ONLINE. 2015 Feb 6;3(1):2-1.
- [8] Han Jiawei and Micheline Kamber, *Data Mining: Concepts And Techniques*, 2nd ed. China, Chi.: China Machine Press, 2006.
- [9] Michael J.A. Berry and Gordon S. Linoff, *Data Mining Techniques For Marketing, Sales, Customer Relationship Management*, 2nd ed.: Willey Publishing, 2004.
- [10] Anik K. Jain, "Data Clustering: 50 Years Beyond K-Means," *Pattern Recognition Letters*, pp. 651-666, 2010.
- [11] Eko Prasetyo, *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*, 1st ed., Nikodemus WK, Ed. Yogyakarta: Penerbit Andi, 2012.