Journal of Telematics and Informatics (JTI)
Vol.6, No.2, June 2018, pp. 144~151
ISSN: 2303-3703

Optimizing Group Discussion Generation Using K-Means Clustering And Fair Distribution

Alfano Endra Wardhana [1], Imam Much Ibnu Subroto [2], Sri Arttini Dwi Prasetyowati [2]
[1] Teacher of SMK Sore Tulung agung
[2] Master of Electrical Engineering Department, Islamic University of Sultan Agung Semarang
Email : alfanoewa@gmail.com

Abstract

The development of computer-based learning system today can provide a different learning process in a teaching and learning process, but the problems faced by a teacher is the difficulty in grouping discussion group that has a different value of knowledge and skills, because usually this selection of discussion groups in e-learning is done based on the wishes of each student or randomly regardless of the data of knowledge and skills. This research was conducted with the aim of grouping the discussion groups based on the indicators of knowledge and skill by using k-means clustering analysis at SMK Sore Tulungagung. The knowledge and skills scores of class X students in Pekerjaan Dasar Elektromekanik subjects, The Competence of Electricity Installation Engineering will be used as the basic scores. Then, the students of class X were divided into 2 groups, namely the k-means based group and the random based group for further research. The mean score of knowledge and skills are before the learning process and after the results of the evaluation of the discussion group on the k-means based and the random based group. The k-means based class score increases 4,083 from the average. Before the learning, it was 83.292 and it becomes 87.375 after the evaluation, while the random based class only experienced an increase 0,083 from the average. Before the learning it was 81,250 and it becomes 81,333 after the learning evaluation. Based on the result, grouping the discussion group in a fair way in e-learning on the indicators of knowledge and skills using k-means clustering method shows more visible improvement, so k-means clustering is a more optimal method.

Keywords: *discussion group, e-learning, knowledge, skills, k-means clustering.*

1. Introduction

In improving the quality of education, whether in the context of student's resources or learning systems, each SMK has its own management mode. Empowerment of learning system is one of the important components in terms of advancing the quality of the vocational school. If the learning system owned by SMK is able to contribute greatly to the learning activities of students, then the student resources owned by the SMK will also have the good potential and achievement. Communication-based learning system of communication information developed by each SMK has an influence on the availability of information that support the learning activities.

The development of existing computer technology based on learning system can provide different learning process in a teaching and learning process, where the class can still take place even though the teaching staffs with the users do not have to face to face directly. The learning system that based on communication information technology is called e-learning. Students as users can perform various activities such as, downloading learning materials, sending assignments, doing the tasks, doing quizzes and so forth. Furthermore, some of the activities performed by these users are referred as e-learning user behavior.

Each student has different skills and knowledge to each other, so a teacher is responsible for determining the students in the selection of discussion groups in a fair way in e-learning. However, there is a problem occurred, a teacher's difficulties in grouping the discussion groups that have different knowledge and skills, because the discussion group in e-learning is usually done based on the wishes of each student or randomly, without looking at the data of the students' knowledge and skills.

One of the techniques in getting information among large sets of data and quite well known today is data mining. Data mining is defined as a process for finding patterns in the data. This process runs automatically or more often runs semi-automatically. The patterns found should be meaningful and profitable. Data mining can be used for several things, namely estimation, prediction, classification, clustering method and association. This research will apply clustering method with k-means algorithm.

The clustering method is a technique to categorize data by separating data into groups according to certain desired characteristics where the group identity of each data is not known yet. With

this method it is expected to know the data of the group and then give the identity according to the problems encountered.

The previous research which has been published in the journal article about grouping discussion group using k-means clustering was the Application of K-Means Algorithm for Grouping the Student's Score Data by Teguh Hariyadi from Dian Nuswantoro University of Semarang. The clustering of this research only uses a knowledge-ability test.

Based on the problem above, k-means clustering method can be used to group the students in discussion group in a fair way in e-learning based on knowledge and skill ability.

2. Literature Review

2.1 E - Learning

E-learning consists of two words, namely 'e' which stands for 'electronica' and 'learning' which means 'learning'. So, e-learning means learning by using the help of electronic devices. So, in practice, e-learning uses the services of audio, video or computer equipment or a combination of all of them. In other words, e-learning is a learning which is supported by technology services such as telephone, audio, videotape, satellite transmission or computer. In line with that, Onno W. Purbo explains that the term "e" in e-learning is all the technology used to support teaching efforts through internet electronic technology. Internet, satellite, audio / video tapes, interactive TV, and CD-ROM are some of the electronic media used [1].

2.2 Group Discussion

Group discussion is a meeting of two or more people, it is done to exchange experiences and opinions, and usually to reach a joint decision. According to the opinions above, it can be concluded that group discussion techniques are a form of activity characterized by an attachment to a subject matter or question, in which the members or participants of the discussion are honestly trying to come to a conclusion after listening and studying, as well as considering the opinions delivering during the discussion [5].

2.3 K-Means Clustering

The process of grouping a set of objects into the same object classes is called clustering [4]. Clustering is one of the functions in data mining processes to find the groups or identification of almost identical groups of objects. Clustering is an attempt to identify similar groups of objects and help in finding patterns of dispersion and relationship patterns in large sets of data. An important point in the clustering process is to declare a set of patterns to appropriate groups that are useful for finding similarities and differences to produce valuable conclusions.

K-means is one method of non-hierarchical data clustering that seeks to partition existing data into one or more clusters / groups. This method groups the data into the cluster / group so that the data which have the same characteristics (High intra class similarity) are grouped into the same cluster and the data which have different characteristics (Law inter class similarity) are grouped into another cluster [3]. The clustering process begins by identifying the data to be in the cluster, Xij (i = 1, ..., n; j = 1, ..., m) where n is the amount of data to be in the cluster and m is the number of variables. At the beginning of the iteration, the center of each cluster is set free (arbitrary), Ckj (k = 1, ..., k; j = 1, ..., m). Then we calculate the distance between each data with each cluster center. To calculate the distance between the i-th data (xi) and the k-cluster center (ck), we give the name (dik), Euclidean formula [2] can be used as in (1), that is:

$$d_{ik} = \sqrt{\sum_{j=1}^{m} \left(C_{ij} - C_{kj}\right)^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

A data will be a member of the k-cluster if the distance to the center of the k-cluster is at least small when compared to the distance to the center of the other cluster. This can be calculated by using equation (2). Next, we group the data that belongs to each cluster. [2]

$$Min \sum_{k=1}^{k} d_{ik} = \sqrt{\sum_{j=1}^{m} \left(C_{ij} - C_{kj}\right)^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

The score of the new cluster center can be calculated by finding the average score of the data that belongs to the cluster, using the formula in equation 3. [2]

$$c_{kj} = \frac{\sum_{i=1}^{p} x_{ij}}{p}; \quad \text{..... (3)}$$

3. Research Methods

The type of research methods being used is modeling and experimenting. Experimental research is a research where the researchers manipulate one or more variables in a certain way so that it affects one or more other variables that are measured. In this final project, the writer applies k-means algorithm to an auxiliary program to group students' discussion group in a fair way using indicator of knowledge and skill.

To do the distribution of grouping the discussion groups with k-means method, the first step that must be done is collecting the data of knowledge and skills scores of the subjects "Pekerjaan Dasar Elektromekanik" class X, The Competence of Electricity Installation Engineering, 2016 / 2017. The second step is determining the groups as k-means group and random group. The third step is processing the student 's data using k-means clustering method. After obtaining the clustering results based on the students' knowledge and skills, the writer continues the fourth step. That is grouping the students' discussion groups in a fair way by dividing the first cluster students evenly in each group followed by the second and the third cluster students.

After a fair group discussion is established, the fifth step is learning in groups. The sixth step is evaluating the learning outcomes in groups. If you have not obtained the desired results, then back to the third step of data processing, if data evaluation of research is in accordance with the desired hence obtained an increase in the score of the students' knowledge and skills. So, we can conclude the differences of grouping the discussion group in a fair way on the students' ability of knowledge and skills using k-means clustering method.

3.1. Data Grouping

a. Data Clustering With K-Means

The steps of the K-means Clustering method consist of several stages. First, specify the score of k as the number of clusters to be formed. Second, specifies the initial centroid (cluster central point) at random. Next the third step, calculates the distance of each data to each centroid using the formula of correlation between two objects (Euclidean Distance). Fourth, group each data by the closest distance between the data with its centroid.

Fifth, determine the new centroid position by calculating the average value of existing data on the same centroid as the new cluster center. Sixth, if the position of the new centroid with the old centroid is not the same as the calculation then back again to the third step. Finally, save the clustered data with the same cluster center.

b. Group Grouping Discussion

The steps of Pseudo-code modeling consist of several steps. First, specify the result of clustering students' score data using k-means clustering method. Second, divide the first cluster member into each discussion group. After that, the first cluster member is divided into the second cluster until the second cluster member is divisible, and also the third cluster and so on. Next, if all the cluster members are divided into group discussions, then the learning process in groups can begin.

The steps of forming the discussion group are modeled using the algorithm in the form of Pseudo-code as shown in Figure 1

```
        From a number of clustered groups
step 1: create K where K is the number of discussion groups that will be filled by the
        members
step 2: start from the i-st cluster where i = 1
step 3: take 1 member from i cluster and put him into each discussion group one by one
step 4: if there is some rest of step 3 then back to step 3
step 5: repeat step 3 for i = i + 1 to n
```

Figure 1 Modeling in Pseudo-code form

3.2. Evaluation

This evaluation is done after implementing the learning process in groups using the students' knowledge and skills. The test of knowledge uses a written test with the subject matter of "Pekerjaan Dasar Elektromekanik" that has been discussed, while the skill test is the result of the student's practice.

If the results of the data evaluation are in accordance with what have been desired, then an increase in the score of the students' knowledge and skills has been reached. So that it can be concluded that grouping the discussion group in a fair way using k-means clustering method shows great differences on the students' knowledge and skills.

To measure the success rate of k-means clustering based on discussion group method, it is tested on two methods namely random method and k-means method. After the learning process with the discussion of each group has been done, then the results will be assessed based on the students' test scores of both methods.

4. Result and Discussion

After a discussion group in a fair way is established, the next step is learning in groups. The evaluation of the outcomes of group learning during four meetings with the theoretical and practical subjects of "Pekerjaan Dasar Elektromekanik" was conducted with both written and practice tests.

Comparison of scoring data after discussion grouping in a fair way on students' knowledge and skill using k-means clustering method or k-means based group and randomized or random based group, as shown in table 1 and 2.

Table 1 Data Evaluation Result of K-Means Based Group

NUM.	STUDENT'S NAME	REG. NUMBER	Knowledge		Skills	
			Before	After	Before	After
1	ABDULATIP NGALIMI	23749/177.010	76	80	77	80
2	ACHMAD BAHTIAR FADILAH	23750/178.010	77	81	76	80
3	ADE ALMADUN	23751/179.010	90	92	89	91
4	ADE KURNIAWAN	23752/180.010	90	92	93	94
5	ADIN ANDRE FABELAN	23753/181.010	78	82	77	82
6	ADITIA BAGASKORO	23754/182.010	87	90	86	89
7	AGNES ADE PUTRA PAMUNGKAS	23755/183.010	75	80	75	79
8	AGUNG SETIAWAN	23756/184.010	82	87	83	89
9	AHMAD KUSHAIRI	23757/185.010	79	85	80	87
10	AHMAD QOZI PRATAMA	23758/186.010	81	86	79	84
11	AHMAD RIDUWAN	23759/187.010	84	87	85	87
12	AKBAR DWI WAHYUDI	23760/188.010	75	79	75	80
13	ALDO ANANTA TRI SUTOYO	23761/189.010	87	92	86	90
14	ALFIAN ARWIS SANTOSO	23762/190.010	89	93	90	94
15	ALWI BURHANUDDIN ZUHRI	23763/191.010	79	83	79	85
16	ANDI CAHYO	23764/192.010	92	94	95	96
17	ANDI MARCOPOLO	23765/193.010	78	86	77	85
18	ANDITO EKO SAPUTRA	23766/194.010	80	90	89	94
19	ANGGA SAPUTRA	23767/195.010	86	92	90	94
20	ANISA NIRMALA DEWI D.	23768/196.010	91	93	84	88
21	ANJUN RAMANDA SEPTIAN	23769/197.010	77	81	75	80
22	ARDI EKO MAHENDRO	23770/198.010	92	95	93	95
23	ARDIANTO	23771/199.010	75	80	75	79
24	ARI KURNIAWAN	23772/200.010	95	96	95	96

Table 1 shows the scoring data before the learning process and the result score of group discussion evaluation grouped by k-means clustering. The score of knowledge and skills are all equally improved after members of the discussion group are divided in a fair way.

Students with better knowledge and skills will assist students with less understanding of learning materials so that all members of the discussion group can be better at evaluation after learning.

Table 2 Data Evaluation Result of Random Based Group

NUM.	STUDENT'S NAME	REG. NUMBER	Knowledge		Skills	
			Before	After	Before	After
25	ATOEK FATTUR ROHMAN	23773/201.010	86	88	86	89
26	BAGAS ANGGA PERMADHI	23774/202.010	84	84	79	83
27	BAGAS ISMOYO	23775/203.010	75	78	75	78
28	BAGAS SATRIO UTOMO	23776/204.010	89	90	86	86
29	BAMBANG PRAYOGO	23777/205.010	82	84	83	83
30	BERNAT DWI CAHYO PURNOMO	23778/206.010	79	80	80	82
31	CHOIRUL DWI PRASETYO	23779/207.010	84	84	84	83
32	DANANG DWI SAPUTRO	23780/208.010	78	78	78	77
33	DEVANO ROMADHON	23781/209.010	76	77	76	76
34	DHANAR KHESIT PANGANGKAH	23782/210.010	79	80	80	80
35	DIKKY FEBRIAN	23783/211.010	87	85	89	85
36	DIMAS ALWI NUR ROZEK	23784/212.010	80	80	78	80
37	DIMAS IMAM SAPUTRO	23785/213.010	77	75	77	76
38	DIMAS SANTOSO	23786/214.010	78	76	75	75
39	DOMAS SUHARTA	23787/215.010	76	75	76	75
40	DONNIE SETIAJI	23788/216.010	81	80	88	85
41	DORI QURNIAWAN	23789/217.010	79	80	81	80
42	DUWI FEBRIANTO	23790/218.010	81	82	77	78
43	EDWIN PRASETIYO	23791/219.010	79	80	81	82
44	EKO CAHYONO	23792/220.010	77	76	77	77
45	ELYAS ADIP NURYAHYA	23793/221.010	82	80	83	81
46	FEBI SETYAWAN	23794/222.010	86	85	86	85
47	FICKI QOIRUL IDZAM	23795/223.010	84	85	83	84
48	FIRMAN ANDI SYAHPUTRA	23796/224.010	89	90	94	92

Table 2 shows the scoring data before the learning process and the scoring result of the random group discussion. After group learning, some scores of knowledge and skills increases, but some of them also decrease. Even, the scores increase very little or tend to remain.

Discussion groups with no students with better knowledge and skills will still be less understanding of the learning materials. As a result, there are some discussion groups that cannot improve their knowledge and skills during evaluation after the lesson.

Table 3 Comparison of the mean scores of the k-means based group and random based group

NUM.	GROUP	Mean Score of Knowledge and Skills		Enhancement Score
		Before	After	
1	K-MEANS	83,292	87,375	4,083
2	RANDOM	81,250	81,333	0,083

Figure 2 Comparison of the mean scores of the k-means based group and random based group

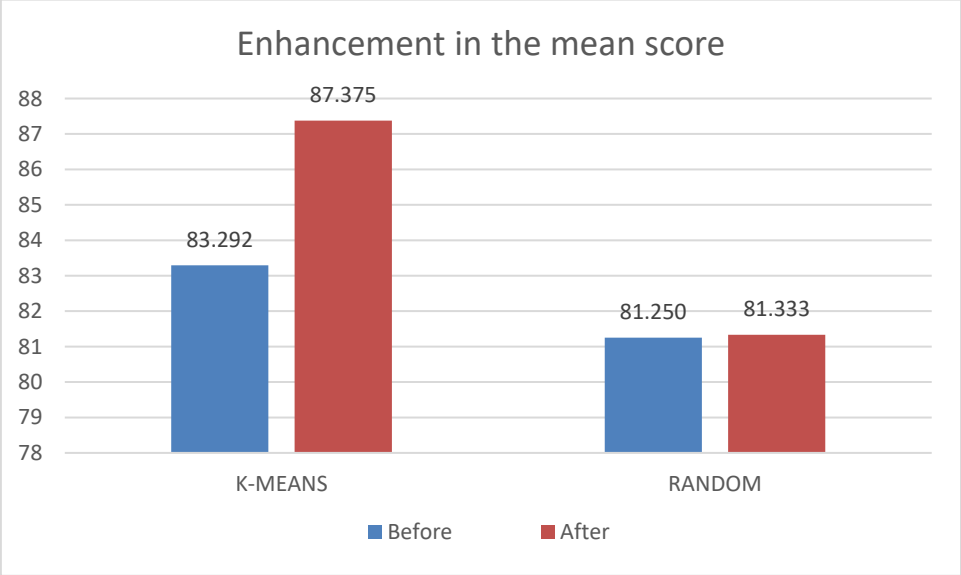

Table 3 and figure 3 show the mean score of knowledge and skills before the learning process and after the results of the discussion group's evaluation on k-means based group and random based group. K-means based group score increases 4,083 from the average, before the learning it was 83.292 and it becomes 87.375 after the evaluation, while random based group only experienced an increase 0,083 from the average, before the learning it was 81,250 and it becomes 81,333 after the learning evaluation.

Based on the results above, grouping the discussion group in a fair way in e-learning on the indicators of knowledge and skills using k-means clustering method shows more visible improvement.

5. Conclusion

From the results of analysis and discussion in this study, the written can reach some conclusions. First, the increase score of discussion groupings using k-means clustering analysis is 4,083. Second, the score of discussion group randomly increased only 0,083. And third, grouping of discussion groups using k-means clustering analysis with knowledge and skill indicators can further improve student's evaluation results compared with randomly grouping.

6. References

[1]. A. S. Aribowo, "E-learning Cerdas Dengan Personalisasi Menggunakan Teknik Data Mining Dan Decision Support System," in Seminar Nasional Informatika 2010 (semnas IF 2010), Jogyakarta, 2010.

[2]. A. P. D. J. ,. H. F. Fernandez, "E-learning and Educational Data Mining in Cloud Computing: An Overview," Int. J. Learning Technology, vol. 9, no. 1, pp. 25-52, 2014.

[3]. Giyanto, Heribertus. 2008. *Penerapan algoritma Clustering K-Means, K-Medoid, Gath Geva*. Tesis Tidak Terpublikasi. Yogyakarta: Universitas Gajah Mada.
[4]. Han, Jiawei; & Kamber, Micheline. 2001. *Data Mining Concepts and Techniques Second Edition*. San Francisco: Morgan Kauffman.
[5]. Sukardi, Dewa Ketut. 2002. *Pengantar pelaksanaan program bimbingan dan konseling di sekolah*. Jakarta: Rineka Cipta.