Clustering student behavior based on quiz activities on moodle LMS to discover the relation with a final exam score

I Waspada, N Bahtiar, and A Wibowo
Department of Informatics, Faculty of Science and Mathematics, Diponegoro University
Jl. Prof. H. Soedharto, SH, Tembalang, Semarang 50275, Indonesia
E-mail: indrawaspada@undip.ac.id

Abstract. One strategy in using the Learning Management System (LMS) is to provide activities that can be accessed by students as learning material. The problem is that the teacher cannot see the students’ online activity directly. In this study, the students' behavior was discovered from the e-learning log file on the quiz activity. Quizzes are performed six times and end with a final exam. Four attributes were proposed as the student behavior model, they are the frequency of repeating quiz work (freq), the duration on each quiz work (dur), increased score when repeating the quiz (perf), and the best score of all attempts (best). The preferable clustering result using K-means algorithm was obtained with 2 (two) clusters. Each cluster was paired with the final exam score which shows that the percentage of failures in cluster_0 is 85%. This result corresponds to the centroid value in cluster_0 that is lower in all attributes compare to the cluster_1. It concludes to evidence that the activity of working on online quizzes has a relation with the final exam score. This information can be used for early intervention by the teacher to prevent the student from failing.

1. Introduction
Moodle as a popular Learning Management System (LMS) provides a quiz feature that can be managed by the teacher for their student. Quizzes are a useful tool for students to test their level of knowledge [1] and commonly used for formative evaluation [2, 3]. The study by Nespereira [4] showed that there was a correlation between student interaction in LMS and the final exam scores, it is in line with the research conducted by López [5] that the average score of the final exam correlated with content interaction, which the quiz module in Moodle classified into the category of content interaction. The impossibility to follow in detail how students behave while learning in LMS may have rather crucial consequences, especially for evaluating students’ online activities (online assessment) [6]. Many possibilities may occur, according to [7] and [8], there are various types of behavior, such as: guessing behavior, sleeping behavior, gaming the system, help-seeking behavior, misusing or underusing help, willingness to collaborate, etc.

It is relatively little attention has been paid to developing analytical techniques to explore unique behaviors of students in online learning environments [6]. Clustering and classification are the methods most frequently used to identify student behaviors [9, 10]. In [11], clustering method was used to produce a cluster model of students based on cognitive types and their performance. López [12] has done clustering based on some activities on assignment, forums, courses, and resources modules that produce 4 (four) clusters with one of them being identified as a cluster with the highest pass.
probability. Recent work by [6] on analysis of student interactions uses process mining methods allowing for mapping and modeling the process of completing quizzes by students.

In this research, the clustering was focused on getting the student’s behavior model in answering on the quiz module. The target was to discover the cluster (patterns) that have a fail student. Thus it is expected that the detected pattern can be used by the teacher to give an early intervention to prevent that failure.

There is some evidence to support in revealing the fail students behavior, they are (1) behaviour of repeating the exercise as much as possible to get perfect score [13], (2) behavior of gaming the system, which is defined as attempting to succeed in an educational task by systematically taking advantage of properties and regularities in the system used to complete that task, rather than by thinking through the material [14], and (3) behavior that unmotivated examinees will answer too quickly [15]. It is confirmed by [16] that higher motivation to learn has been linked not only to better academic performance, but to greater conceptual understanding, satisfaction with school, self-esteem, social adjustment, and school completion rates.

2. Methodology
The research was carried out by referring to the CRISP-DM framework which includes business understanding, understanding data, data preparation, modeling, evaluation, and deployment stages.

2.1. Business understanding
The case study in this study was taken from the class of Computer Network given to students of the Library and Information Vocational Study Program at Diponegoro University. The teacher made use of Undip’s e-learning at kulon.undip.ac.id (kulon) to provide material files and organize quizzes.

Table 1. Mapping of Targeted Behavior to Quiz Work Indicators

| No | Targeted Behavior                                                                 | Quiz Work Indicators of Fail Student                                                                 |
|----|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| 1  | Students in the online tutorial repeat the exercise as much as possible to get a perfect score [13]. | Rarely, no time, or lazy to practice  
• Not targeting perfect score, they stop to attempt although the score is not perfect yet |
| 2  | Gaming the system [14]                                                            | Work hastily, brute force, or cheat  
• The pattern of many repetitions but the value obtained decreases  
• Not targeting perfect score, they stop to attempt although the score is not perfect yet |
| 3  | Unmotivated examinees will answer too quickly [15]                                 |                                                                                                       |

Table 2. Mapping of Quiz Work Indicators to the Measurement Attributes.

| No | Quiz Work Indicators of Fail Student                                                                 | Measurement                                                                 | Attribute Name |
|----|----------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|----------------|
| 1  | Rarely, no time, or lazy to practice                                                                | The frequency of repeating a quiz work (low repetition)                     | freq           |
| 2  | Work hastily, brute force, or cheat                                                                  | Duration on each quiz work (low duration)                                   | dur            |
| 3  | The pattern of many repetitions but the value obtained decreases                                     | Increased score when repeating the quizzes, as performance (low performance) | Perf           |
| 4  | Not targeting perfect score, they stop to attempt although the mark is not perfect yet                | The best score of all attempt (low score)                                   | best           |
We analyzed 73 students data from a total of 6 quizzes. Data was taken from the official E-learning website, kulon.undip.ac.id (Moodle LMS). Quiz activity is chosen as the subject of student behavior research which is expected to have a relationship with the final exam result. Quizzes were configured to be able to assessed repeatedly (unlimited). Quiz question randomly picked from question bank.

The hypothesis raised in this study to discover our targeted behavior is that students who fail the final exam, are likely to have specific indicators when working on quiz activity as listed in table 1. The next activity is mapping the indicators to the measurement variable as the attribute of quiz work which is listed in table 2. It is expected that all attributes will support each other to predict the fail student behavior.

2.2. Data understanding

The data used is obtained through the Moodle logs, for more detailed information is presented in table 3.

Table 3. Quiz activity log file from Moodle LMS.

| No | Filename                           | Material                                   |
|----|------------------------------------|--------------------------------------------|
| 1  | Log aktivitas kuis materi 1.csv    | Network type                               |
| 2  | Log aktivitas kuis materi 2.csv    | Binary - decimal number conversion         |
| 3  | Log activities kuis materi 3.csv   | Network Protocol                           |
| 4  | Log aktivitas kuis materi 4.csv    | Introduction of packet tracer              |
| 5  | Log aktivitas kuis materi 5.csv    | Hub and Switch                             |
| 6  | Log aktivitas kuis materi 6.csv    | Computer network and internet              |

Each quiz has 10 default attributes as shown in table 4, but only 5 which are useful for quiz activity analysis purposes, they are attributes number 6 to 10. These attributes cannot be used immediately, but require the preprocessing stage discussed in the next stage.

2.3. Data preparation

At this stage, preprocessing data is carried out which includes extraction, transformation, cleaning (handling of missing value), and normalization stages to get 4 (four) required attributes as stated in table 1. It is necessary to extract from the raw data with several methods as presented in table 5.

Table 5. Extraction and transformation methods to get the required attributes.

| No | Attribute Name | Extraction and Transformation Methods                                                                                                                                 |
|----|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | freq           | Counting all log data of quiz work based on the same name or id, and must have finished status                                                                       |
| 2  | dur            | Standardize time to minutes, then subtracting “Completed” attribute with “Start on” attribute.                                                                        |
| 3  | perf           | Compare the last score with the earliest repetition (0 = don’t work on quiz; 0.25 = the last score is smaller, 0.5 = the same value, 1 = the last score is greater)             |
| 4  | best           | Take the best score from all repetitions working on the quiz                                                                                                            |

After obtaining the appropriate attributes, then the data needs to be cleaned because some data have missing values, among others, because the students did not work on the quiz. The missing value then fills with a zero value. The next thing to deal with is data normalization using the min-max with range 0 to 1. Finally, the data is ready to be modeled as presented in table 6.
2.4. Modeling

In the modeling stage, the K-means algorithm is used to do clustering with the determination of K values referring to the Elbow method combined with the value of the Davies-Bouldin index. This stage was done by using Rapidminer tool.

| No | Attribute      | Information                      |
|----|----------------|----------------------------------|
| 1  | Surname        | Last name                        |
| 2  | First name     | First name                       |
| 3  | Institution    | Institution name                 |
| 4  | Department     | Department name or Study program |
| 5  | Email address  | Email address                    |
| 6  | State          | Quiz status is working or Done   |
| 7  | Started on     | Time to start working on the quiz|
| 8  | Completed      | Time to complete the quiz        |
| 9  | Time took      | Duration of quiz work            |
| 10 | Grade          | Quiz score                       |

Table 4. Available attributes in the quiz log file.

Table 6. Preprocessing results data.

| id | freq1 | ... | freq6 | ... | dur1 | ... | dur6 | ... | perf1 | ... | perf6 | ... | best1 | ... | best6 |
|----|-------|-----|-------|-----|------|-----|------|-----|-------|-----|-------|-----|-------|-----|-------|
| 1  | 0.3   | ... | 0.4   | ... | 0.5  | ... | 0.1  | ... | 1.0   | ... | 1.0   | ... | 1.0   |     | 1.0   |
| 2  | 0.3   | ... | 0.0   | ... | 0.6  | ... | 0.0  | ... | 1.0   | ... | 0.0   | ... | 1.0   |     | 0.0   |
| 3  | 0.2   | ... | 0.6   | ... | 0.5  | ... | 0.2  | ... | 0.5   | ... | 1.0   | ... | 1.0   |     | 1.0   |
| ...| ...   | ... | ...   | ... | ...  | ... | ...  | ... | ...   | ... | ...   | ... | ...   | ... | ...   |
| 73 | 0.3   | ... | 0.0   | ... | 0.1  | ... | 0.0  | ... | 1.0   | ... | 0.0   | ... | 1.0   |     | 0.0   |

2.4.1. Clustering with K-Means method

The K-Means algorithm uses the process repeatedly to get the cluster database. It takes the desired number of initial clusters as input and produces the number of end clusters as output. If the algorithm is needed to generate cluster K, then there will be an initial K and a final K. The K value chosen as the initial center will be calculated using the Euclidean Distance formula, which is to find the closest distance between the centroid point and the data/object. Data that has a short distance or closest to the centroid will form a cluster.

The main problem in the K-Means algorithm is to determine the value of k. There are several methods to determine the best k value, including the most often used methods, are the Davies-Bouldin Index and Elbow [17].

2.4.2. Davies-Bouldin Index

Davies-Bouldin Index (DB) is a validity index that does not depend on the number of clusters and the clustering algorithms [18], this is based on the idea that good separation between clusters is a partition with intra-cluster homogeneity and high cohesiveness [19]. Then to determine the DB index, we need to define the size of the dispersion and the size of the cluster similarity [18]. Dispersion of $S_i$ from cluster $C_i$ moreover, separation $D_{ij}$ between clusters $i^{th}$ and $j^{th}$ are defined [19] using formula 1 and formula 2.

$$S_i = \left( \frac{1}{|C_i|} \sum_{x \in C_i} D^p(x, c_i) \right)^{\frac{1}{p}}$$

$p > 0$ (1)
with \(|C_i|\) is the number of data points in the cluster \(C_i\), and \(c_i\) is the center of the cluster \(C_i\),

\[
D_{ij} = \left( \sum_{l=1}^{d} \left| v_{il} - v_{jl} \right|^t \right)^{1/t}, \quad t > 1
\]

(2)

with \(v_i\) and \(v_j\) are centroids each from cthe luster \(C_i\) and \(C_j\). Then the DB index is defined as follows:

\[
V_{DB} = \frac{1}{k} \sum_{i=1}^{R_i} R_i
\]

(3)

with \(k\) is the number of clusters and \(R_i\) defined as follows:

\[
R_i = \max_{i\neq j} R_{ij}
\]

(4)

with \(R_{ij}\) is a benchmark of similarity between cluster \(C_i\) and \(C_j\) which is defined as follows:

\[
R_{ij} = \frac{S_i + S_j}{D_{ij}}
\]

(5)

Since the goal is to achieve minimum intra-cluster dispersion and maximum inter-cluster separation, the number of cluster \(c\) which minimizes \(V_{DB}\) Taken as the optimal value of \(c\) [5].

2.4.3. Elbow method

Elbow method is a method used to produce information in determining the best number of clusters by looking at the percentage of the comparison between the number of clusters that will form an elbow at a point. This method provides ideas/ideas by selecting cluster values and then adding the value of the cluster to be used as a data model in determining the best cluster.

The following is a measurement of the number of intra-cluster distances between points using a euclidian distance formula with the number of clusters \(C_k\) containing \(n_k\) observation points [20]:

\[
D_k = \sum_{i \in C_k} \sum_{j \in C_k} \left\| x_i - x_j \right\|^2 = 2n_k \sum_{i \in C_k} \left\| x_i - \mu_k \right\|^2
\]

(6)

By summing the normalized intra-cluster square, it gives a measure of compactness of the cluster with the following formula:

\[
W_k = \sum_{k=1}^{K} \frac{1}{2n_k} D_k
\]

(7)

Variant quantity \(W_k\) is the basis of the procedure for determining the optimal number of clusters which are then better known as Elbow method [20]. The higher the number of clusters \(K\) then the value of \(W_k\) will be smaller.

2.5. Evaluation

Evaluations carried out on the modeling results include their suitability with the planned targets at the beginning, namely getting a pattern of fail student behavior.

2.6. Deployment

The finished model can then be implemented into an online quiz activity clustering application program that can be used by teachers to detect and handle potentially fail students.

3. Experimental result

In this section outlined the results of experiments that discuss the determination of \(K\) values, analysis of the results of clustering as a whole, analysis of the results of clustering per attribute, and analysis of the distribution of the final exam results on the cluster found.

3.1. Determination of K value in K-means

The desired number of clusters is as minimal as possible with the best (optimal) performance. Therefore, the value of \(k\) up to 10 is used, each cluster quality is calculated using the DB Index and Elbow Method. The test results are illustrated in figure 1.
Based on Figure 1, the graph of DB Index and Elbow shows that 2 and 5 are the preference choice of k, the conclusion goes to k=2 as through some experiments and observations that fail student behavior cluster always farthest compare the others, so the smallest number of k is the best choice for analysis purpose.

3.2. K-Means with k=2
Detailed centroid values from the clustering process using K-Means with a value of k = 2 are presented in figure 4 which illustrates the comparison of the centroid value of each attribute between cluster_0 and cluster_1. By observing the Figure 4, it appears that all attributes of cluster_1 are higher then cluster_0

3.3. Analyze each Attribute
The purpose of the per attribute analysis is to see in more detail the comparisons, primarily related to the attributes discussed in the 1st to 6th quizzes.

Figure 1. Determination of the number of k.

Figure 2. Centroid value per cluster.
3.3.1. The frequency of Quiz Repetition
In the frequency attribute, it is expected to be able to model the activeness and enthusiasm of student learning. Based on figure 3, generally, cluster 0 shows lower than cluster 1. In the second quiz, all clusters experienced a decrease in repetition frequency, this may be caused by the material studied was mathematical, making it difficult for vocational students in the library and information to work on it, this lowers the spirit to repeat.

![Figure 3. Frequencies of quizzes.](image)

3.3.2. Duration of Quiz Work
Duration attribute is modeled as students who work seriously with only trial and error or cheating. In general, when students work online quizzes with the desire to understand better and try to get good grades, they will undoubtedly be careful and make sure the answers are correct. Thus it is possible for the student to re-examine the notes and reference books used. While students who are not dangerous, their activity patterns will do brute force or cheat because they are only concerned with value, do not try to understand or remember.
Based on figure 3, it seems cluster_0 has behavior to do quizzes faster than cluster_1; there is a trend of decreasing duration from quiz 1 to 6, this is maybe because students are increasingly familiar with strategies and environments working on online quizzes. Cluster_0 does quiz 2, and quiz 5 very quick which indicates the possibility of working with trial and error, using shortcuts or cheating.

![Figure 4. Centroid of Finishing Duration of Quizzes.](image)

3.3.3. Repetition Quiz Performance
The performance attribute aims to model students' efforts naturally in improving their understanding by repeating the quiz. Good performance is when repeating the quiz work produces better value.
Based on figure 5, cluster_0 is high enough in quiz 1 but then decreases on the next quizzes, as the opposite of cluster_1 which shows the trend of increasing performance from quiz 1 to 6, this trend indicating student at cluster_1 have the motivation to get good grades by trying to repeat and learn from mistakes done before.

![Figure 5. Centroid of Repetition Performa of Quizzes.](image)

3.3.4. The best value from quiz work
Figure 6 shows that cluster_0 is a much lower motivation than the cluster_1.

![Figure 6. Centroid on Best Score of Each quiz.](image)

3.4. Analysis of the distribution of the final exam results to the clusters found
To see the relation between the behavior of students working on an online quiz with the final exam results, table 7 is arranged as a matrix between clusters found with the status of the exam results, namely fail and pass. The results in table 7 show cluster_0 are quite convincing to fail (85%).

|       | cluster_0 | cluster_1 |
|-------|-----------|-----------|
| fail  | 11 (85%)  | 9 (15%)   |
| pass  | 2 (15%)   | 51(85%)   |

![Table 7. Fail and pass ratio in each cluster.](image)
The average values obtained in each cluster are presented in Table 8, showing the alignment between the behavior in the cluster and the ratio of fail and pass.

| Cluster Index | Average Final Exam Score |
|---------------|--------------------------|
| Cluster_0     | 44.00                    |
| Cluster_1     | 80.64                    |

4. Conclusion

The clustering with four attributes that are proposed able to discover the cluster_0 to fail student behavior, they have lower value in all attributes. Based on these results it is also evidence that the activity of working on online quizzes has a relation with the results of the final exam. This information can be used for early intervention by the teacher.

References

[1] Martín-blas T, Serrano-fernández A 2009 Computers 52 35
[2] Kibble J 2007 UAJP Adv Physiol Educ 31 253
[3] Peat M, Franklin S 2002 Br J Educ Technol 33 15
[4] Nespereira CG, Dai K, Redondo RPD, Proc. Second Int. Conf. Technol. Ecosyst. Enhancing Multicult. - TEEM '14, pp 283–290
[5] López SLS, Redondo RPD, Vilas AF 2016 Is interpersonal participation relevant to pass ? In: Fourth Int. Conf. Technol. Ecosyst. Enhancing Multicult. - TEEM’16. ACM, Salamanca, pp 319–325
[6] Juhaňák L, Zounek J, Rohliková L 2017 U Comput Human Behav. doi: 10.1016/j.chb.2017.12.015
[7] Peña-Ayala A 2014 Expert Syst Appl 41 1432
[8] Baker RSJ d., Inventado PS 2014 Educational Data Mining and Learning Analytics. In: Learn. Anal. From Res. to Pract. pp 73–83
[9] Jovanović J, Gašević D, Dawson S, Pardo A, Mirriahi N 2017 Internet High Educ 33 74
[10] Dutt A, Ismail MA, Herawan T 2017 A Systematic Review on Educational Data Mining. IEEE Access 5:15991–16005
[11] Jovanovic M, Vukicevic M, Milovanovic M, Minovic M 2012 Int J Comput Intell Syst 5 597
[12] López SLS, Redondo RPD, Vilas AF 2017 Discovering knowledge from student interactions. In: Proc. 5th Int. Conf. Technol. Ecosyst. Enhancing Multicult. - TEEM 2017. pp 1–8
[13] Cleaver TG, Elbasyouni LM 2005 Student online assessment behaviors. IEEE Trans Educ 48:400–401
[14] Baker RSJ d., Corbett AT, Koedinger KR, Evenson S, Roll I, Wagner AZ, Naim M, Raspat J, Baker DJ, Beck JE 2006 Adapting to When Students Game an Intelligent Tutoring System. In: ITS’06 Proc. 8th Int. Conf. Intell. Tutoring Syst. pp 392–401
[15] Wise SL, Kong X 2005 R Appl Meas Educ 18 163
[16] Usher A, Kober N 2012 Cent Educ Policy 4 53
[17] Kodinariya TM, Makwana PR 2013 RInt J Adv Res Comput Sci Manag Stud 1 2321
[18] Gan G, Ma C, Wu J 2007 Data Clustering: Theory, Algorithms, and Applications, ASA-SIAM Series of Statistics and Applied Probability. American Statistical Association, Alexandria, Virginia
[19] Pal NR, Biswas J 1997 C Pattern Recognit 30 847
[20] Tibshirani R, Walther G, Hastie T 2001 J R Stat Soc B 63 411