

## A cluster and association analysis visualization using Moodle activity log data

Andri Reimondo Tamba, Krista Lumbantoruan, Aulia Pakpahan, Samuel Situmeang

Department of Information System, Faculty of Informatics and Electrical Engineering, Institut Teknologi Del, Toba, Indonesia

---

### Article Info

#### Article history:

Received Dec 20, 2021

Revised Mar 16, 2022

Accepted Sep 11, 2022

---

#### Keywords:

Association analysis

Cluster analysis

Course activity log

Data visualization

Moodle

---

### ABSTRACT

The course activity log is where a learning management system (LMS) like Moodle keeps track of the various learning activities. The instructor may directly examine the log or use more complex method such as data mining to conduct a quicker and more in-depth examination of the student's behaviors. Most previous studies on analyzing this log data rely on predictive analysis. Instead of predictive analysis, this study investigates cluster analysis and association analysis. Cluster analysis based on k-means++ is utilized to organize students into groups, given their engagement in the learning course module. Association analysis based on apriori is utilized to extract the relationships between various student activities. A dashboard presentation of the findings is provided to facilitate clearer comprehension. Based on the analysis findings, it can be concluded that the structure of the student cluster is medium. In contrast, the association between student activities is positively correlated and well-balanced. The subjective review of the dashboard reveals that the visualization is already sufficient, but there are some recommendations for making it even better.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

### Corresponding Author:

Samuel Situmeang

Department of Information System, Faculty of Informatics and Electrical Engineering

Institut Teknologi Del

Laguboti, Toba, North Sumatra, Indonesia

Email: samuel.situmeang@del.ac.id

---

## 1. INTRODUCTION

Analyzing and processing user activity data generated from learning management systems (LMS) is a contemporary trend in higher education. In particular, the LMS stores accessed resources or activities metadata in the course log. The substantial amount of data from this platform offers crucial data that can help instructors and students achieve their educational objectives. In order to conduct a more efficient and in-depth examination of the student's actions, the instructor can directly examine the log or use more sophisticated methods such as data mining. Data mining has several functionalities that can be used to analyze the course log. Some of the functionalities include classification [1]–[7], regression analysis [5], [8]–[12] cluster analysis [2], [13] and association analysis [14], [15]. Although previous studies show promising results, most of the proposed approaches to analyze LMS logs have primarily focused on predictive analytics, such as classification and regression analysis. This analytic type overshadows other data mining functionalities like cluster and association analysis.

Research by Hussain *et al.* [2] compare several classification and clustering algorithms to extract user performance patterns during completing the Moodle course, which enables the instructor to detect the low-performance user in advance before the examination. The experimental result shows that the k-means clustering algorithm performs quite well in grouping inactive or active users and poorly performed users

compared to the classification algorithm. In another study, Pradana *et al.* [13] compare the performance of k-means, hierarchical, and louvain clustering algorithms in grouping users based on resource view, course view, assignment view, and upload assignments. The experiment result confirms that the louvain clustering algorithm performed slightly better than the k-means and hierarchical clustering algorithms. K-means will always converge on different solutions depending on the initial centroids. However, this solution may be a local minimum. As a result, the computation is often done several times, with different initializations of the centroids. One method to help address this issue is the k-means++ clustering algorithm [16]. This clustering algorithm initializes the centroids to be distant in general from each other, leading to better results than random initialization.

Research by Aviano *et al.* [14] propose a behavioural tracking model that instructors can use to determine Moodle LMS students learning status. By using this model, an instructor can determine the dominant and most influential behavioural traits in learning using association rules based on the apriori algorithm. Research by Dimić *et al.* [15] suggests the implementation of apriori association analysis algorithm for improving the process of e-testing in blended learning environment. The results of the performed study indicate that, having set particular limits, association analysis could be extremely useful for providing feedback on how to improve and enhance the process of e-testing. In order to generalize the method described in the paper, further study in using data from various courses and different settings of limit values, so as to determine general parameters for the objective approach to the rules' significance assessment. The apriori algorithm achieves good performance gained by reducing the size of candidate sets. In the case of Moodle LMS that the analysis performed at the course level, there are not a large number of frequent patterns or long patterns so this algorithm not necessarily suffer from candidate generation and multiple database scan.

Accordingly, this study focuses on cluster analysis based and association analysis on students activities in LMS (in this case, the widely used open-source system Moodle). These are essential in developing better learning activities or modules. K-means++ is employed to deal with k-means shortcomings in cluster analysis. Besides that, the apriori association analysis algorithm is directly utilized instead of manually defining association rules between behavioral traits or actions. In the preceding research, the author proposes the apriori algorithm to extract association rules. This algorithm generates numerous uninteresting itemsets, leading to various association rules that are of no use.

## 2. RESEARCH METHOD

In general, the research method in this study sequentially consists of data collection, data analysis, data pre-processing, application of data mining functionalities, visualization, evaluation, and conclusions formulation. The application of data mining functionalities consists of cluster analysis and association analysis. Meanwhile, evaluation consists of cluster analysis evaluation, association analysis evaluation, and visualization evaluation. The general overview of the research methodology in this study is shown in Figure 1.



Figure 1. General overview of research method

### 2.1. Data collection

The dataset used in this study originates from three-course learning records in the odd semester of the 2020/2021 academic year at the Del Institute of Technology, Indonesia. The three courses include data visualization, artificial intelligence, and natural language processing. Each course learning process is carried out via the Moodle LMS. The learning record is stored in a course log that is separate from one course to another.

### 2.2. Data analysis

There are nine attributes in the dataset, namely time, user full name, affected user, event context, component, event name, description, origin, and IP address. The dataset has nine columns and 52,460 rows. The time attribute cannot be directly used as timestamp because the format is not followed by the international organization for standardization (ISO). Therefore, the attribute format needs to be changed to conform with the ISO in the preprocessing stage. There are several conditions from the LMS activity log,

such as learning courses, assignments or quizzes given every week, and exams (middle semester examination and final semester examination). An example of the contents of the activity log data can be seen in Table 1.

Table 1. Sample of activity log contents

Time	User full name	Affected user	Event context	Component	Event name	Description	Origin	IP address
1/11/20, 11:23	12S170** J***	-	Course: 12S4056-visualisasi data (S1SI, 2020/2021, <i>gasal</i> )	Logs	Log report viewed	The user with id '2,123' viewed the log report for the course with id '1,213'.	web	36.71.141.20
1/11/20, 11:22	12S17*** Me***	-	Course: 12S4056-visualisasi data (S1SI, 2020/2021, <i>gasal</i> )	Logs	Log report viewed	The user with id '2,123' viewed the log report for the course with id '1,213'.	web	36.71.141.20

### 2.3. Data preprocessing

Data preprocessing carried out in this research consists of removing rows with all duplicate values, feature selection, timestamp data type conformation, and sequence activity extraction. Selected features include time, event context, component, and event name. Sequence activity extraction is done by grouping events based on user id and specific intervals.

### 2.4. Cluster analysis

Based on the first business question, the authors propose clustering with k-means++. Students is grouped based on predetermined attributes to measure student activity. To produce a business question formulation, the researcher first reviewed the reflection indicators that could measure the implementation of learning. Reflection indicators affect latent variables or indicators that can reflect, represent, and observe the effects of latent variables. Therefore, the latent variables need to be determined beforehand. Latent variables cannot be observed directly unless measured by one or more indicators. The latent variables to obtain indicators of reflection on the implementation of learning can be seen in Table 2 [17], [18].

Table 2. Latent variables for reflection of learning implementation indicators

No.	Variable latent	Description
1	Rate of attendance	Teachers and students meet from the entire specified class schedule.
2	Disciplines of attendance	Teachers and students are present at all meetings according to the specified schedule.
3	Disciplines of lecture official record reporting	The teacher makes a teaching report after the implementation of the lesson.
4	Quality of work	Students have a balanced understanding and achievement.
5	Capability	Students are satisfied with the material presented by the teacher.
6	Communication	Teachers are able to facilitate communication between students and teachers and students with students.
7	Initiative	Teachers are able to increase students' creativity in achieving achievements.
8	Competition	Teachers are able to motivate students to participate in competitions related to learning.
9	Student activity	Teachers can see student activity while at LMS based on student activities.
10	Relationships between online activities	Teachers can see student behavior while using the LMS based on the relationship between activities.

Based on the latent variables described in Table 1, we determine "student activity" and "relationships between online activities" as latent variables to obtain reflection indicators. This is because the dataset used in this study only fulfills the requirements for processing these latent variables. To process all latent variables other than "student activity" and "relationships between online activities", it takes several attributes that are not contained in the dataset used in this study, such as value attributes and class schedules. Based on this, "student activity" and "relationships between online activities" are the most appropriate latent variables to use.

After determining "student activity" and "relationships between online activities" as latent variables, the next step is to determine reflection indicators. Reflection indicators are used to measure latent variables. Reflection indicators that can be used to measure "student activity" and "relationships between online activities" are summarized in Table 3.

Table 3. Reflection indicator

No.	Reflection indicator	Description
1	Total contributions made by students in each forum.	"Student activity" can be seen based on contributions made by students since the forum was opened. The more contributions a student makes, the student is said to be active in the forum. The number of student contributions can be calculated using the formula: discussion created+post created-post deleted. To meet these needs, the event name, component, and user_id attributes are used.
2	The average number of pages that students access during learning in the LMS.	"Student engagement" can be seen based on the average number of pages students visit each week on the LMS. The more pages a student visits, the more active the student is. To meet these needs, the time, user_id, and event name attributes are used.
3	The average activity per session that students do in each event while using the LMS.	"Student activity" can be seen based on the number of activities that students do sequentially while using the LMS. The more activities students do in each event at the LMS, the more active the students are. To meet these needs, the time, user_id, and event name attributes are used.
4	The average time used by students during the course of study.	"Student activity" can be seen based on the time students spend during the course learning. The longer the time used by students to carry out learning during lectures, the students are said to be active in lectures. The total time used by students during course learning can be calculated by the formula: lesson ended-lesson started in the event name column. To meet these needs, the time, user_id, and event name attributes are used.
5	Activity transactions based on sequential events.	"The relationship between online activities" can be measured from the relationship between one activity and another. This is done to see the interrelated activities that students often do. To obtain these indicators, it takes duration, date, time, event name, and user id. To meet these needs, the attributes of time, user full name, event context, event name, and description are used.

The evaluation of the clustering model is carried out by calculating the algorithm's performance based on a predetermined clustering model. In determining the accuracy of the k-means++ clustering model, it is done by measuring the quality and strength of a cluster that combines the values of cohesion and separation. One of the most common methods used to test the results obtained from the clustering method is the silhouette coefficient method. The size of the silhouette coefficient is found in the values of -1 to 1. If the value of the silhouette coefficient is closer to the value of 1, the better the quality of grouping objects into a cluster, on the contrary, if the value is closer to the value of -1, the worse the quality of the object will be. It was grouping the data in the cluster. The assessment criteria or measurement of whether or not the clustering results are based on the silhouette coefficient according to Kaufman and Roesseeuw [19] can be seen in the Table 4. In this study, researchers used the silhouette coefficient as a method for calculating the accuracy of a cluster generated by the k-means++ clustering method.

Table 4. Kaufman and Roesseeuw assessment criteria [19]

Silhouette coefficient value	Evaluation
0.71–1.00	Strong structure
0.51–0.70	Medium structure
0.25–0.50	Weak structure
≤0.25	No structure (bad structure)

## 2.5. Association analysis

Based on the second business question, the researcher processes the data by applying data mining through the association analysis method. In this case, the researcher compares several association analysis algorithms used in previous studies based on LMS activity log cases. Based on several related studies, such as the application of apriori to obtain association rules from Moodle for case studies of *UN IPA* values and association analysis on behavioral tracking in LMS, apriori gives good association results. This study uses a log of activities carried out by students during the implementation of learning in the LMS. The attributes and characteristics of the dataset used in this study are the same as the dataset used in this study, which contains information on the time and activities of students. In addition, the data type in the attributes used is categorical. Therefore, the researcher uses the apriori association rules algorithm to analyze the activity log in the LMS.

To obtain association rules, it is necessary to determine the frequent itemset from the student activity log in the LMS which contains the sequence of student activities, but to obtain association rules that can be used to develop teaching methods, only a few events are used. Events that do not support the development of teaching methods is removed from the frequent itemset. Teaching methods can be developed based on the activities carried out by students in the LMS. Based on the analysis conducted, students spend more time on discussion forums and viewing content pages. Therefore, the events used to get association rules are course

module viewed, course viewed, discussion created, subscription created, discussion viewed, a post created, quiz attempt reviewed and submission created.

An interesting association rule is an item set that meets the minimum support and minimum confidence limits. This is done to eliminate itemsets that have a support value smaller than the minimum support and itemsets that have a confidence value smaller than the minimum confidence [4]. The minimum value of support and minimum value of confidence is determined by the author. Table 5 shows the minimum support value based on the analysis in previous research.

Table 5. Minimum support value in previous research

No.	Research title	Minimum support
1	Data mining in course management system [20]	0.3
2	Standardizing interestingness measures for association rules [21]	0.3
3	A data mining framework to analyze road accident data [22]	0.3
4	A literature survey on association rule mining algorithms [23]	0.3
5	Set-oriented mining for association rules in relational databases [24]	0.3
6	Analysing road accident data using association rule mining [25]	0.3
7	Analysis of road traffic fatal accidents using data mining techniques [26]	0.4
8	The application of improved association rules data mining algorithm apriori in CRM [27]	0.4
9	Pattern discovery using association rule mining on clustered data [28]	0.5

Based on previous research, six studies are using a minimum support value of 0.3 in several research cases so that this value can be relevant to use in each dataset. Therefore, this study uses a minimum support value of 0.3. Table 6 shows the minimum confidence value based on the analysis in previous research.

Table 6. Minimum confidence value in previous research

No.	Research title	Minimum confidence
1	A guide for association rules mining in Moodle course management system [29]	0.5
2	Association rule mining using k-map model in data mining [30]	0.5
3	Finding frequent pattern with transaction and occurrences based on density minimum support distribution [31]	0.5
4	Association rule algorithm with FP growth for book search [32]	0.5
5	Relevant association rule mining from medical dataset using new irrelevant rule elimination technique [33]	0.5
6	Pattern discovery using association rule mining on clustered data [28]	0.5
7	Set-oriented mining for association rules in relational database [24]	0.7
8	Arules-a computational environment for mining association rules and frequent item sets [34]	0.8
9	Association rules mining on forest fires data using FP-growth and ECLAT algorithm [35]	0.8
10	An improved vertical algorithm for frequent itemset mining from uncertain database [36]	0.9

Based on previous research, six studies use a minimum confidence value of 0.5 in several association cases to use this value in each relevant dataset. Therefore, this study uses minimum confidence of 0.5. In rules generation, if the "relationship between online activities" carried out by students at the LMS has met the minimum support and minimum confidence, then the association rules can be determined. A good evaluation for association rules can be seen in the interestingness measure both subjectively and objectively.

Subjective assessment is generally carried out by the parties concerned in determining the rules of an attractive association. Therefore, this assessment is limited to the responses of certain parties. Meanwhile, objective assessment is carried out by statistical calculations that produce numbers to produce actual measurements for association rules. Based on this information, this study evaluate the association rules by using an interesting measure objectively [4].

Objective assessment is done by measuring the null-invariant and the non-null-invariant. Datasets that have null-transactions use null-invariant evaluations, conversely for datasets that do not have null-transactions use non-null-invariant evaluations. A dataset is called a null-transaction if a transaction has items with a much smaller number of occurrences than other transactions or has a non-existent number of occurrences [5]. The information of these occurrences in student activity log shown in Table 7.

Table 7. Matrix support itemset in student activity log

	Course module viewed	Not course module viewed	Total
Course viewed	23	374	397
Not course viewed	357	1,379	1,736
Total	380	1,753	2,133

In the lift calculation that has been carried out, it is found that the rule (course viewed, course moduled viewed) has a lift value of 3.33, which is greater than 1, which indicates that the transaction (course viewed, course moduled viewed) has a positive correlation while the number of transactions the item (course viewed, course modulated viewed) is of little value. This proves that there is no effect of null transactions in determining the correlation between items. Therefore, the evaluation used is an objective assessment by calculating the null-invariant.

One of the null-invariant evaluation metrics is the Kulczynski measurement which is good for use in datasets that have null transactions. In addition, the balance ratio, which is used to see the balance of the itemset on the association rules and the results of the balance ratio, is highly correlated with the Kulczynski measurement [4]. Therefore, this study uses a null-invariant evaluation metric with Kulczynski and the imbalance ratio.

## 2.6. Clusters and association rules visualization

The data mining process applied in this study produce some information that is presented in a dashboard. Information that is presented in this visualization is information obtained from reflection indicators in obtaining clustering results for information on "student activity" and results of association analysis for "relationships between online activities". Analysis for the visualization of information used to display information in the dashboard can be seen in Table 8.

Table 8. Information visualization analysis

No.	Visualized information	Visualization form	Reason
1	Clustering implementation results	Scatter plot	The information obtained in the implementation of this clustering shows a comparison of student activity based on the defined reflection indicators. The attribute type used to obtain this information is numeric. See the relationship between the attributes used to get the grouping results.
2	The results of the implementation of association analysis	Graph	The information generated in the implementation of this association analysis shows the distribution of student behavior carried out in the LMS.

Based on the information in Table 6, the resulting dashboard contains two data visualization panels, namely for clustering and association analysis. Scatter plot visualization is used to present information on "student activity," while graph visualization displays "relationships between online activities" performed on the LMS activity log. Information obtained from the application of data mining is visualized in various forms of idioms. The visualization is validated using the qualitative-summative technique. The results of this evaluation are in the form of ratings, for example, strongly agree, agree, moderate, disagree, and strongly disagree. The form of evaluation that can be done is to distribute surveys. Surveys can be carried out by giving questionnaires to teachers to obtain and collect assessments of the resulting visualizations.

Nielsen argues that only five respondents have been able to find problems as much as 80% in testing the usability of a system [37]. Therefore, this study uses 10-15 respondents. The distributed survey contains several statements and questions posed by the researcher. The questions and statements submitted have been fulfilled to help researchers find the information they want to obtain from respondents to evaluate the resulting visualization.

## 3. RESULTS AND DISCUSSION

In this section, it is explained the results of the research and, at the same time, is given a comprehensive discussion. Results can be presented in figures, graphs, tables, and others that make the reader understand easily [38], [39]. The discussion can be made in several sub-sections.

### 3.1. Cluster analysis result

Based on the implementation of the k-means++ algorithm with grid search hyperparameter optimization, students are grouped into three clusters, where each cluster shows a measure of student activity. Table 9 shows a snippet of the clustering results obtained. To determine clusters that indicate groups of students who are active, less active, and very inactive, an analysis is carried out on the grouping results based on the cluster center obtained. The resulting centroid value has a negative value, that's because the defined indicator value has a value of 0. The centroid for each cluster is shown in Table 10.

Table 9. Snippet of clustering result

ID students	Total contributions on the forum	Average page accessed	Average activity per session	Average usage time of LMS	Cluster
1,979	10	6.54	56.00	18.05	2
1,980	6	5.43	72.38	7.86	2
1,981	16	6.38	52.85	4.12	2
1,982	10	5.20	38.47	14.75	0
1,983	4	6.00	64.93	11.95	2
1,985	4	5.60	60.38	5.76	2

Table 10. Centroid for each cluster

Indicator	Cluster		
	0	1	2
Total contributions on the forum	-0.45	-0.70	0.84
Average page accessed	-0.31	-0.37	0.54
Average activity per session	-0.59	-0.48	0.91
Average usage time of LMS	-0.56	1.87	-0.13

Table 10 shows the cluster center data obtained in grouping using the clustering method. Based on the table, it can be determined the label of student activity. The centroid value containing a negative value means that one of the indicators in the cluster has a value of 0.

- Cluster 2 is an active student group. This is evidenced by the acquisition of cluster centers in cluster 2 which is greater than in cluster 1 for almost every indicator, namely the total contribution in the forum, the average page accessed, and the average activity per session, and the overall indicator is greater than cluster 0. In addition, the values obtained for almost every indicator in this cluster are positive.
- Cluster 1 is a group of students who are quite active. In this cluster, the centroid obtained by almost all indicators has a lower value than cluster 2, the centroid in this cluster only excels in one indicator, namely, average duration class per week. Meanwhile, when compared to cluster 0, the centroid in cluster 0 is higher for the three indicators. The centroid value in cluster 1 is almost entirely negative, where only one indicator, namely average duration class per week, is positive while the rest is negative.
- Cluster 0 is a group of students who are less active. This is because the cluster center in cluster 0 has the lowest value for all indicators in cluster 2 and three indicators in cluster 1. In addition, all the results obtained in this cluster are negative.

Based on the results of the analysis on the acquisition of the cluster center, cluster 2 was determined for the active student group, cluster 1 for the moderately active student group, and cluster 0 for the less active student group. The results for the clustering implementation obtained information on student activity, namely from 55 students. There were 21 active students, nine quite active students, and 25 fewer active students.

Then, evaluate the grouping results used by applying the silhouette coefficient method. Based on the results of the clustering evaluation obtained, the silhouette score generated in this study was 0.54. Based on the assessment criteria of Kaufman and Roesseeuw, the assessment obtained is a medium structure.

### 3.2. Association analysis result

Based on the sequence of these activities, an itemset is obtained, which contains the sequence of activities carried out by students. In performing association rules, support and confidence are needed, both of which are determined from the itemset that has been obtained. Based on the support and confidence that has been obtained, association rules can be produced that show the relationship between activities carried out by students while using the LMS. The resulting association rules are as many as two rules in the Table 11.

Table 11. Result of association analysis

No.	Antecedents	Consequents	Antecedent support	Consequent support	Support	Confidence
1	(Course module viewed)	(Course viewed)	0.92	0.80	0.73	0.91
2	(Course viewed)	(Course module viewed)	0.92	0.80	0.73	0.79

In Table 9, information on the relationship between activities in the form of antecedents and consequents and the support and confidence of each generated rule. Based on the results of the association rules obtained, the strong rules generation events are course viewed and course module viewed. Thus, the activity is a pattern of student behavior that meets the minimum value of support and minimum confidence. The evaluation results obtained through applying the Kulczynski and the imbalance ratio metric that can be seen in the Table 12.

**Table 12. Result of evaluation**

Evaluation	Kulczynski	Imbalance ratio
Total rules	2	2
Mean	0.85	0.12
Minimum value	0.85	0.12
Maximum value	0.85	0.12

There are association rules containing 2-itemset with values where these association rules have Kulczynski values in the range of values  $[>0.5, 1]$  and IR values in the range of values  $[0, <0.5]$ . This range of values shows that the association rules in rules generation have a positive and interesting correlation with balanced antecedent and consequent support. Based on the list of results of the evaluation of the resulting association rules, the average Kulczynski score is 0.85 (85%).

**3.3. Visualization result**

The data mining functionalities implementation result is visualized in a dashboard. The dashboard consists of a cluster analysis visualization panel and an association analysis visualization panel. Figure 2 shows the cluster analysis visualization panel for data visualization courses, while Figure 3 shows the association analysis visualization panel for data visualization courses.

The cluster analysis results are visualized in an interactive scatter plot. The interactive mechanism in the visualization allows the user to choose the cluster or indicator axis to be shown. Three clusters are produced, namely cluster 0, cluster 1, and cluster 2; each is shown in purple, blue, and red. One plot represents one student with student information and cluster. The visualization also allows the user to see the indicator's value.

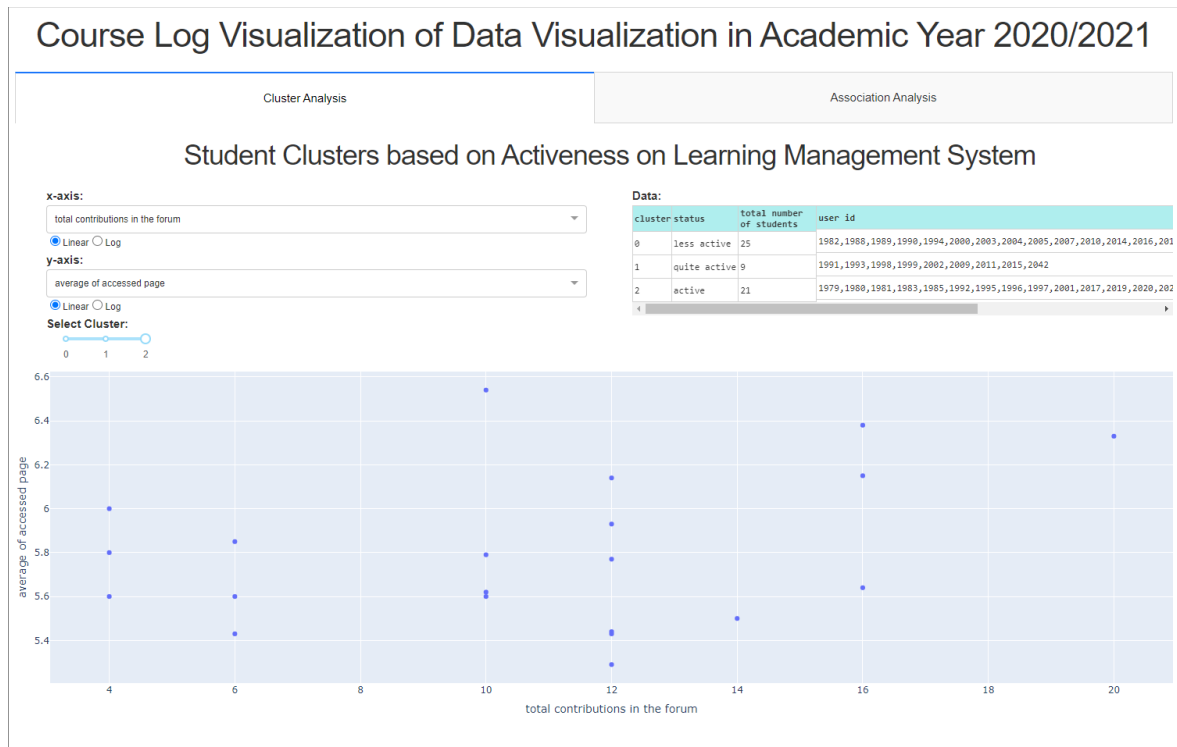


Figure 2. Cluster analysis panel

The association analysis results are visualized in an interactive graph. The visualization shows the relationship between activities that students often carry out. The interactive mechanism in the visualization allows the user to choose the layout type, node form, and node color to be shown.



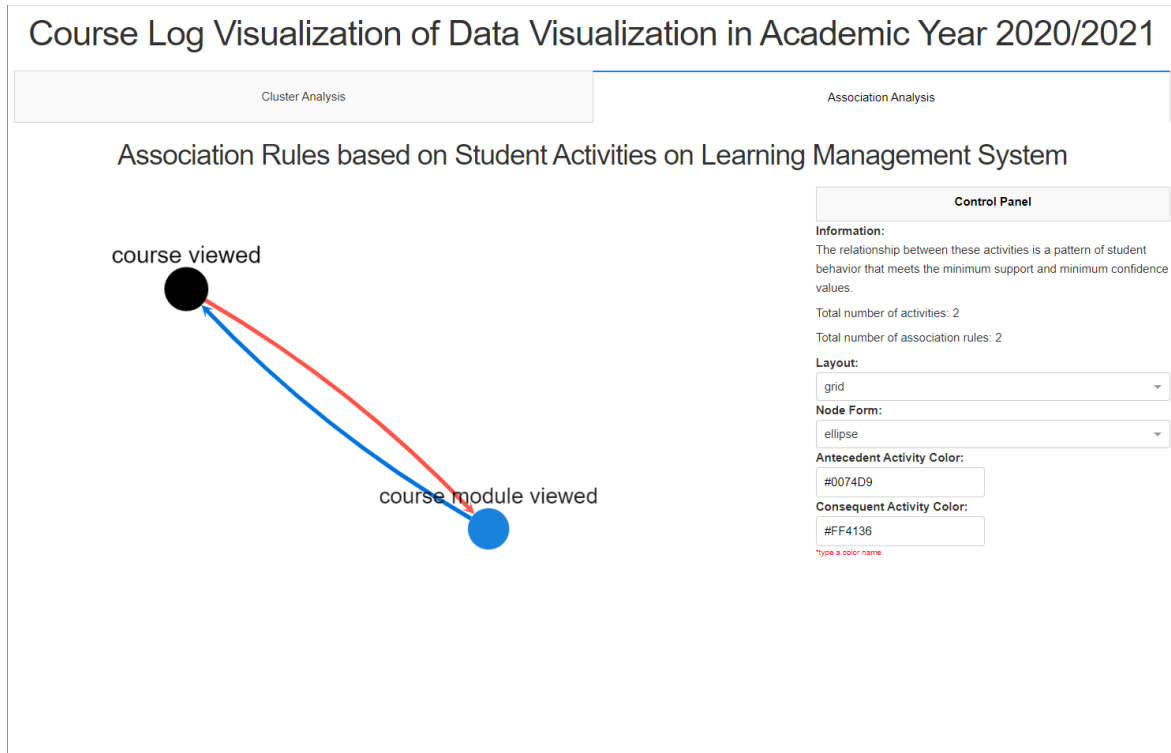


Figure 3. Association analysis panel

Subjective evaluation on visualization is carried out using a questionnaire. Evaluation is intended to assess the effectiveness of the developed dashboard to support the learning records analysis process. Questionnaires were distributed to 12 respondents consisting of lecturers and teaching assistants at the Del Institute of Technology. The value scale ranges from 1 (strongly agree) to 5 (strongly disagree). Figure 4 shows that of the 12 respondents, 3 gave an assessment that strongly agreed, 2 agreed, and 4 quite agreed. This result indicates that the cluster analysis visualization panel can be used in the learning records analysis. Meanwhile, Figure 5 shows that of the 12 respondents, 5 agreed, 5 quite agreed, 1 disagreed, and 1 strongly disagreed. Further examination of the comments section in the questionnaire shows that the disagreement indicates that the delivery of information in the survey could be better.

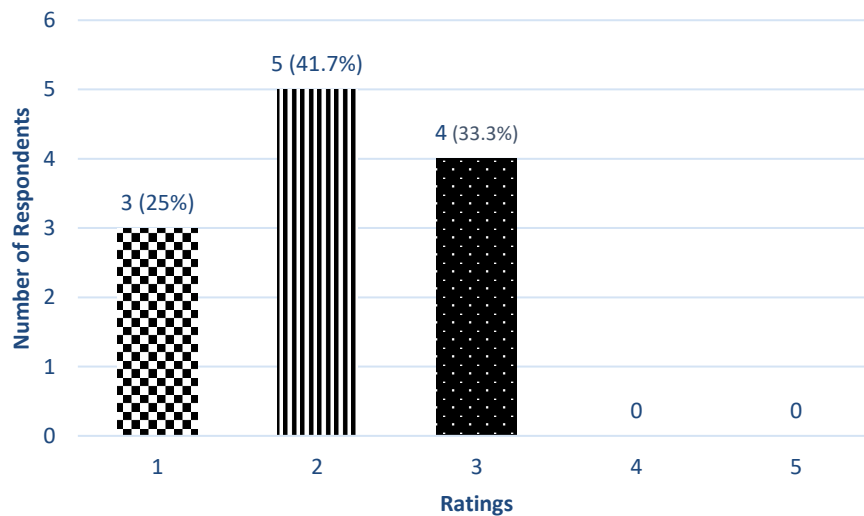


Figure 4. Respondents assessment of the results of clustering visualization

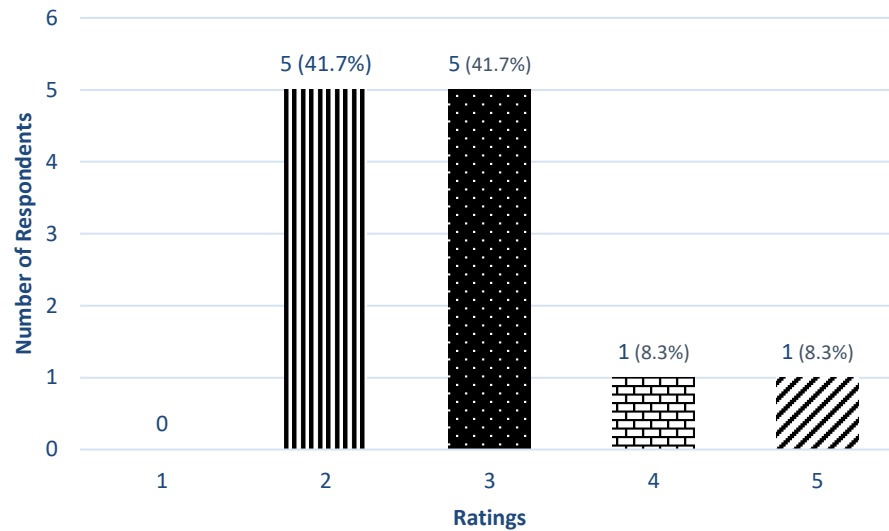


Figure 5. Respondents assessment of association analysis visualization results

#### 4. CONCLUSION




Based on the results of the study, it could be the data mining approach for visualizing activity log data in an LMS could be made with two data mining functionalities, namely clustering to obtain information about "student activity" and association analysis to obtain information about "relationships between online activities" carried out by students. The grouping of "student activity" is obtained by applying the k-means++ algorithm. The evaluation value on the results of clustering using the silhouette coefficient algorithm is 0.54 with medium-structure criteria according to the assessment of Kaufman and Roesseeuw. Meanwhile, information on "relationships between online activities" is obtained from association rules generated using apriori. The association rules were validated using Kulczynski and the imbalance ratio. The range of evaluation values is  $[>0.5, 1]$  for Kulczynski, and the IR value is in the field of values  $[0, 0.5]$ . This range of values indicates that the association rules in rule generation have a positive and interesting correlation. The two pieces of information obtained from the data mining results are visualized on a dashboard. The dashboard was validated using the qualitative-summative technique by distributing the survey, while the purposive sampling technique was used to determine the respondents. The results of the dashboard visualization evaluation generated based on the assessment by the respondents showed that the visualization was quite good in helping to review learning analytics in LMS more quickly. But to make the survey more accurate, it needs to get better at giving out information.

#### REFERENCES




- [1] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Computers & Education*, vol. 113, pp. 177–194, Oct. 2017, doi: 10.1016/j.compedu.2017.05.007.
- [2] M. Hussain, S. Hussain, W. Zhang, W. Zhu, P. Theodorou, and S. M. R. Abidi, "Mining moodle data to detect the inactive and low-performance students during the moodle course," in *Proceedings of the 2nd International Conference on Big Data Research*, Oct. 2018, pp. 133–140. doi: 10.1145/3291801.3291828.
- [3] S. Helal *et al.*, "Predicting academic performance by considering student heterogeneity," *Knowledge-Based Systems*, vol. 161, pp. 134–146, Dec. 2018, doi: 10.1016/j.knosys.2018.07.042.
- [4] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: a tutoring action plan to prevent academic dropout," *Computers & Electrical Engineering*, vol. 66, pp. 541–556, Feb. 2018, doi: 10.1016/j.compeleceng.2017.03.005.
- [5] A. Alshaqiti and A. Namoun, "Predicting student performance and its influential factors using hybrid regression and multi-label classification," *IEEE Access*, vol. 8, pp. 203827–203844, 2020, doi: 10.1109/ACCESS.2020.3036572.
- [6] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Computers in Human Behavior*, vol. 104, pp. 1–33, Mar. 2020, doi: 10.1016/j.chb.2019.106189.
- [7] S. Shrestha and M. Pokharel, "Educational data mining in moodle data," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 10, no. 1, pp. 9–18, Apr. 2021, doi: 10.11591/ijict.v10i1.pp9-18.
- [8] I. Mwalumbwe and J. S. Mtebe, "Using learning analytics to predict students' performance in moodle learning management system: a case of Mbeya University of Science and Technology," *The Electronic Journal of Information Systems in Developing Countries*, vol. 79, no. 1, pp. 1–13, Mar. 2017, doi: 10.1002/j.1681-4835.2017.tb00577.x.

- [9] E. Alqurashi, "Predicting student satisfaction and perceived learning within online learning environments," *Distance Education*, vol. 40, no. 1, pp. 133–148, Jan. 2019, doi: 10.1080/01587919.2018.1553562.
- [10] M.-D. Dascalu, M. Dascalu, S. Ruseti, M. Carabas, S. Trausan-Matu, and D. S. McNamara, "Cohesion network analysis: predicting course grades and generating sociograms for a Romanian moodle course," 2020, pp. 174–183, doi: 10.1007/978-3-030-49663-0\_21.
- [11] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, "Predicting student performance from LMS data: a comparison of 17 blended courses using moodle LMS," *IEEE Transactions on Learning Technologies*, vol. 10, no. 1, pp. 17–29, Jan. 2017, doi: 10.1109/TLT.2016.2616312.
- [12] M.-D. Dascalu *et al.*, "Before and during COVID-19: a cohesion network analysis of students' online participation in moodle courses," *Computers in Human Behavior*, vol. 121, pp. 1–19, Aug. 2021, doi: 10.1016/j.chb.2021.106780.
- [13] C. Pradana, S. S. Kusumawardani, and A. E. Permanasari, "Comparison clustering performance based on moodle log mining," *IOP Conference Series: Materials Science and Engineering*, vol. 722, no. 1, pp. 1–11, Jan. 2020, doi: 10.1088/1757-899X/722/1/012012.
- [14] D. Aviano, B. L. Putro, E. P. Nugroho, and H. Siregar, "Behavioral tracking analysis on learning management system with apriori association rules algorithm," in *2017 3rd International Conference on Science in Information Technology (ICSITech)*, Oct. 2017, pp. 372–377, doi: 10.1109/ICSITech.2017.8257141.
- [15] G. Dimić, B. Predić, D. Rančić, V. Petrović, N. Maček, and P. Spalević, "Association analysis of moodle e-tests in blended learning educational environment," *Computer Applications in Engineering Education*, vol. 26, no. 3, pp. 417–430, May 2018, doi: 10.1002/cae.21894.
- [16] D. Arthur and S. Vassilvskii, "K-Means++: the advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, pp. 1027–1035.
- [17] E. J. Dommert, B. Gardner, and W. van Tilburg, "Staff and students perception of lecture capture," *The Internet and Higher Education*, vol. 46, pp. 1–10, Jul. 2020, doi: 10.1016/j.iheduc.2020.100732.
- [18] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. U. Sarker, and M. U. Sattar, "Predicting student performance in higher educational institutions using video learning analytics and data mining techniques," *Applied Sciences*, vol. 10, no. 11, pp. 1–20, Jun. 2020, doi: 10.3390/app10113894.
- [19] L. Kaufman and P. J. Rousseeuw, Eds., *Finding groups in data*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1990, doi: 10.1002/9780470316801.
- [20] C. Romero, S. Ventura, and E. Garcia, "Data mining in course management systems: moodle case study and tutorial," *Computers & Education*, vol. 51, no. 1, pp. 368–384, Aug. 2008, doi: 10.1016/j.compedu.2007.05.016.
- [21] M. R. Shaikh, P. D. McNicholas, M. L. Antonie, and T. B. Murphy, "Standardizing interestingness measures for association rules," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 11, no. 6, pp. 282–295, Dec. 2018, doi: 10.1002/sam.11394.
- [22] S. Kumar and D. Toshniwal, "A data mining approach to characterize road accident locations," *Journal of Modern Transportation*, vol. 24, no. 1, pp. 62–72, Mar. 2016, doi: 10.1007/s40534-016-0095-5.
- [23] P. Yazgana and A. O. Kusakci, "A literature survey on association rule mining algorithms," *Southeast Europe Journal of Soft Computing*, vol. 5, no. 1, pp. 1–14, May 2016, doi: 10.21533/scjournal.v5i1.102.
- [24] M. Houtsma and A. Swami, "Set-oriented mining for association rules in relational databases," in *Proceedings of the Eleventh International Conference on Data Engineering*, 2002, pp. 25–33, doi: 10.1109/ICDE.1995.380413.
- [25] S. Kumar and D. Toshniwal, "Analysing road accident data using association rule mining," in *2015 International Conference on Computing, Communication and Security (ICCCS)*, Dec. 2015, pp. 1–6, doi: 10.1109/CCCS.2015.7374211.
- [26] L. Li, S. Shrestha, and G. Hu, "Analysis of road traffic fatal accidents using data mining techniques," in *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, Jun. 2017, pp. 363–370, doi: 10.1109/SERA.2017.7965753.
- [27] P. Gong, C. Yang, H. Li, and W. Kou, "The application of improved association rules data mining algorithm apriori in CRM," in *2007 2nd International Conference on Pervasive Computing and Applications*, Jul. 2007, pp. 95–99, doi: 10.1109/ICPCA.2007.4365419.
- [28] H. Z. Oo, N. Saing, and M. Kham, "Pattern discovery using association rule mining on clustered data," *International Journal of New Technology and Research (IJNTR)*, vol. 4, no. 2, pp. 7–11, 2018.
- [29] G. Avlijaš, M. Heleta, and R. Avlijaš, "A guide for association rule mining in moodle course management system," in *Proceedings of the International Scientific Conference - Sinteza 2016*, 2016, pp. 56–61, doi: 10.15308/Sinteza-2016-56-61.
- [30] D. A. Abdullah, "Association rule mining using k-map model in data mining," *Eastern Mediterranean University (EMU) - Doğu Akdeniz Üniversitesi (DAÜ)*, 2015.
- [31] P. Khare and H. Gupta, "Finding frequent pattern with transaction and occurrences based on density minimum support distribution," *International Journal of Advanced Computer Research*, no. 3, pp. 3–8, 2012.
- [32] T. Andi and E. Utami, "Association rule algorithm with FP growth for book search," *IOP Conference Series: Materials Science and Engineering*, Dec. 2018, pp. 1–6, doi: 10.1088/1757-899X/434/1/012035.
- [33] K. Rameshkumar, M. Sambath, and S. Ravi, "Relevant association rule mining from medical dataset using new irrelevant rule elimination technique," in *2013 International Conference on Information Communication and Embedded Systems (ICICES)*, Feb. 2013, pp. 300–304, doi: 10.1109/ICICES.2013.6508351.
- [34] M. Hahsler, B. Grün, and K. Hornik, "Arules—a computational environment for mining association rules and frequent item sets," *Journal of Statistical Software*, vol. 14, no. 15, pp. 1–25, 2005, doi: 10.18637/jss.v014.i15.
- [35] N. Ariney and I. S. Sitanggang, "Association rules mining on forest fires data using FP-growth and ECLAT algorithm," in *2015 3rd International Conference on Adaptive and Intelligent Agroindustry (ICAIA)*, Aug. 2015, pp. 274–277, doi: 10.1109/ICAIA.2015.7506520.
- [36] J. Yang, Y. Zhang, and Y. Wei, "An improved vertical algorithm for frequent itemset mining from uncertain database," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Aug. 2017, pp. 355–358, doi: 10.1109/IHMSC.2017.87.
- [37] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93*, 1993, pp. 206–213, doi: 10.1145/169059.169166.
- [38] J. Sadowski, "When data is capital: datafication, accumulation, and extraction," *Big Data and Society*, vol. 6, no. 1, pp. 1–12, Jan. 2019, doi: 10.1177/2053951718820549.
- [39] J. R. Saura, B. R. Herráez, and A. Reyes-Menendez, "Comparing a traditional approach for financial brand communication analysis with a big data analytics technique," *IEEE Access*, vol. 7, pp. 37100–37108, 2019, doi: 10.1109/ACCESS.2019.2905301.




**BIOGRAPHIES OF AUTHORS**

**Andri Reimondo Tamba**    he received his bachelor's degree from the Information System Study Program, Faculty of Informatics and Electrical Engineering, Del Institute of Technology. His interest areas are software quality and artificial intelligence. He can be contacted at email: andriraymond98@gmail.com.






**Krista Lumbantoruan**    she received her bachelor's degree from the Information System Study Program, Faculty of Informatics and Electrical Engineering, Del Institute of Technology. Her interest areas are software quality and artificial intelligence. She can be contacted at email: kristasihombing1999@gmail.com.



**Aulia Pakpahan**    she received her bachelor's degree from the Information System Study Program, Faculty of Informatics and Electrical Engineering, Del Institute of Technology. Her interest areas are software quality and artificial intelligence. She can be contacted at email: auliapakpahan12@gmail.com.



**Samuel Situmeang**    he received his bachelor's degree from the Information Technology Study Program, Faculty of Computer Science and Information Technology, University of North Sumatra and master's degree from the Department of Computer Science and Information Engineering, College of Engineering, National Chung Cheng University. His interest areas are artificial intelligence, multimedia data analysis, and data warehouse. He can be contacted at email: samuel.situmeang@del.ac.id.