Process Mining on Learning Activities in a Learning Management System

Gede Agung Ary Wisudiawan*, Angelina Prima Kurniati*

*School of Computing, Telkom University, Jalan Telkomunikasi Terusan Buah Batu Bandung, West Java, Indonesia
degunk@telkomuniversity.ac.id, angelina@telkomuniversity.ac.id

Abstract—A learning management system is a system that supports the management of teaching and learning activities in an educational institution. Telkom University has its own learning management system called CeLOE which was developed based on the open-source Moodle application. This study analyses an event log, automatically generated by the CeLOE LMS, that records student and lecturer activities in learning. The event log is mined to obtain a process model representing learning behaviours of the lecturers and students during the learning process. The case study in this research is learning in the study program 365 during the first semester of 2020/2021. The results of the analysis obtained are expected to increase understanding of the learning process and produce recommendations for improving the learning management system. The conclusions obtained from the analysis of the process model include the identification of activities and learning content that are used and carried out by students and lecturers, the most frequent learning activities and resources done by students and lecturers during the first semester of 2020/2021 in the study program 365, and the gap of activities and learning resources during the first semester of 2020/2021 in the study program 365.

Keywords—learning management system, event log, process mining, learning process, study program

I. INTRODUCTION

In this modern era, every organization strives to support its business processes with computer-based systems. Universities are no exception; more and more are implementing computer-based learning management systems (LMS). LMS allows universities to manage learning activities carried out by lecturers and students, in the form of material delivery, assignments, quizzes, exams, etc. One of the advantages of implementing LMS in universities is the automatic creation of an event log that is recorded by the LMS. The event log records step-by-step of the LMS usage by lecturers and students in the learning process.

The event log is useful to support process analysis. Process analysis allows understanding of business processes by analysing each activity and the activity sequence, identification of the most common activity sequence carried out by lecturers and students during the learning process, identification of exceptional sequence in certain groups of lecturers and students, and their conformity to learning process implementation guidelines at the university.

One of the main methods for process analysis is process mining. Process mining [1] is a process-based data analysis approach that describes user behaviours based on event log records of an information system. The event log is the main input in the process discovery stage, which is an approach to create a process model that displays the sequence of activities within a business process. The process model that has been produced will then be analysed at the conformance checking stage to check the conformity between the process model and the reference model. The results of process discovery and conformance checking can be used to recommend enhancements to a process that is currently running.

Process mining has been used in several case studies, including manufacturing [2], auditing [3], health [4], and education [5] [6]. In the education domain, process mining can be applied to analyse student and lecturer activities, patterns of learning activities in a study program, and provide recommendations to improve teaching and learning activities in a class. In general, process mining in education focuses on analysing how learning materials and learning activities are used to support the teaching and learning process [6]. There are many algorithms that can use to do process mining like heuristic miner, genetic miner, and fuzzy miner [7]. In this study, heuristic miner algorithm is used because this study need develop Petri net for further analysis. And heuristic miner algorithm can handle the event log with noise [8].

In this study, process mining is used to analyse the event log of CeLOE LMS, an LMS that is used at Telkom University. Based on previous research [8], CeLOE has met the readiness requirement to be analysed using process mining. In this study, an event log from one study program in one semester is used as a representative of learning process in study programs at Telkom University. This study aims to analyse student and lecturer activities, learning activity patterns, and provide recommendations to improve teaching and learning activities based on activities recorded in the CeLOE LMS event log. The algorithm used in the process discovery is the heuristic miner [9]. This algorithm is chosen because of its ability to handle noise, has been implemented in ProM as the main tool for process mining [10], and has been widely applied to various case studies [11]. The resulting process model will be analysed in conformance checking based on fitness, precision, and generalization [1].
II. RESEARCH METHODOLOGY

This research consists of 5 phases, which are: (1) Planning, (2) Extraction, (3) Data processing, (4) Mining and analysis, and (5) Evaluation. Data processing, mining and analysis, and evaluation phases are three iterative phases that need to be carried out repeatedly until sufficient results are found. General methodology is presented in Figure 1. Each stage will be described in the following subsections.

Figure 1. Process mining methodology

A. Planning

Process mining research starts with planning the scope of research, developing research questions, and building research team. The scope of this research is the teaching and learning activities of lecturers and students in the study program 365 during the first semester of 2020/2021, as recorded in the CeLOE LMS. The main research question is “How is the learning process of students and lecturers at CeLOE LMS?” and is divided into:

Q1. How can we use process mining approach to analyse teaching and learning activities in an LMS?
Q2. What is the most common sequence of activities in the teaching and learning process?
Q3. What can we learn from analysing the teaching and learning activities of lecturers and students that have been carried out for one semester in the study program 365?

The research team consists of computer scientists as the process mining experts who are also the authors of this paper and the CeLOE team as the domain experts and the process owners.

The CeLOE LMS was developed from the open-source Moodle application. Lecturers design student learning process using these two-learning media, to support students to achieve an expected learning outcome [12]. Moodle has several learning media that can be used independently or guided, and can be divided into two types, i.e.:

- Resource: learning resources prepared by lecturers to support student learning. Examples of resources are Book, File, Page, URL, and Glossary.

The database of CeLOE LMS consists of 388 tables [13] that are used to support the administration and operation of the CeLOE LMS. One of the tables contains the event log that is used in this study. The event log is stored in the logstore_standard_log table consisting of several attributes [14]:

- Courseid: a unique attribute of a record
- Components: modules or components referred to in an event, e.g., forums, wiki.
- Action: the type of action of an event, e.g., created, updated, viewed
- Target: related submodule information, e.g., post, discussion, page
- Contextinstanceid: a unique attribute that contains the course module, e.g., team message board, team wiki
- Timestamp: date and time

B. Extraction

In this phase, we extracted the CeLOE LMS event log obtained from the logstore_standard_log table. An event log can be defined as:

\[ E = (c, a, t) \]

where \( E \) is a set of events parsed by case id \( c \), activity name \( a \), and timestamp \( t \). An event describes an activity \( a \) that occurs in a particular case \( c \) at timestamp \( t \). Meanwhile, trace \( T \in E \) is a temporal sequence of events that occur in a case, where \( T \in E \). In this study, a case represents a lecturer or student who has a set of events related to their interaction with the CeLOE LMS for one semester.

The initial event log obtained is the event log of the entire study program from the university, then event log was grouped based on the existing study program. To get the event log for each study program, the logstore_standard_log table is matched with the course and course_category tables in the courseid attribute. After getting the event log for each study program, then separating the records based on the lecturer and student users, which is done by matching the logstore_standard_log table with the user table on the userid attribute.

After getting the separated event logs of student and lecturer, then filtering process is carried out based on component attributes. This filtering process sorts out activities and resources such as assignments, quizzes, forums, chat, active quizzes, and feedback. Resources in the form of resources, URL, H5P, folders, pages, glossaries, and books. The filtering process is done using the DISCO application from Fluxicon.

The characteristics of the event logs of lecturers and students are shown in Table 1 and Table 2. In both tables there are traces, events, event classes, variants, first event and last event. Traces are the number of sequences of activities that are formed.
Events are user activities, both lecturers and students, that are recorded in the event log. Event classes are types of activities in the event log (quiz, assignment, forum, etc.). Variants are the number of variations in the sequence of activities in the process. First and last events are activities carried out at the beginning and end of the process.

### Table 1. Characteristics of Lecturer Event Log

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traces</td>
<td>26</td>
</tr>
<tr>
<td>Variants</td>
<td>17</td>
</tr>
<tr>
<td>Events</td>
<td>206</td>
</tr>
<tr>
<td>Event classes</td>
<td>8</td>
</tr>
<tr>
<td>First event</td>
<td>2020-09-02 14:02:58</td>
</tr>
<tr>
<td>Last event</td>
<td>2021-01-27 19:20:45</td>
</tr>
</tbody>
</table>

Table 1 shows that the lecturer log consists of 26 traces that correlate with 26 lecturers, with 17 variants (65%). In addition, there are 206 events that can be distinguished in 8 event classes.

### Data processing

Data pre-processing is the third phase of this research. Data pre-processing is carried out to ensure that the available event logs are ready to be processed with the heuristic miner algorithm. We undertake a role-based data processing to separately analyse lecturer and student records.

1. **Lecturer data processing**

   The next step is filtering from the logstore_standard_log table attribute, namely CRUD (create-read-update-delete) by selecting records that have C and U. The selection of these 2 types of CRUD is because when the lecturer adds and changes activities or resources it will be recorded in the CRUD attribute with type C and U. Then do the merge subsequent events process assisted by ProM tools using the merge subsequent events plugin to ensure that there are no redundant records.

2. **Student data processing**

   Student log data is processed by filtering event name attributes using Disco Fluxicon tools. This process selects several activities and one sub activity that is most relevant to represent the activities of the students. This filtering is done to minimize the looping of existing activities. For example, a quiz activity has sub-activities attempt viewed, attempt submitted, and course module viewed then in the log three records are recorded with the same quiz activity but different sub-activity. In this case, the selected sub-activity was attempt submitted which stated that the student worked on the quiz until it was submitted. The results of the analysis of the selected activities and sub-activities can be seen in Table 3 where there are activities owned by student users, selected sub-activities to represent the activity and a description containing a description of the selected sub-activity.

### Table 3. Filtering Activity & Sub Activity Student Log

<table>
<thead>
<tr>
<th>Activity</th>
<th>Sub Activity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forum</td>
<td>Course_module_viewed</td>
<td>View activity</td>
</tr>
<tr>
<td>Glossary</td>
<td>Course_module_viewed</td>
<td>View resource</td>
</tr>
<tr>
<td>HSP</td>
<td>Course_module_viewed</td>
<td>View resource</td>
</tr>
<tr>
<td>Page</td>
<td>Course_module_viewed</td>
<td>View resource</td>
</tr>
<tr>
<td>Quiz</td>
<td>Attempt_submitted</td>
<td>submitted</td>
</tr>
<tr>
<td>Resource</td>
<td>Course_module_viewed</td>
<td>View resource</td>
</tr>
<tr>
<td>URL</td>
<td>Course_module_viewed</td>
<td>View resource</td>
</tr>
<tr>
<td>Folder</td>
<td>Course_module_viewed</td>
<td>View resource</td>
</tr>
<tr>
<td>Feedback</td>
<td>Course_module_viewed</td>
<td>View activity</td>
</tr>
<tr>
<td>Choice</td>
<td>Answer_created</td>
<td>submitted</td>
</tr>
<tr>
<td>Chat</td>
<td>Message_sent</td>
<td>Message sent</td>
</tr>
<tr>
<td>Book</td>
<td>Course_module_viewed</td>
<td>View resource</td>
</tr>
<tr>
<td>Assignment</td>
<td>Assesable_submitted</td>
<td>submitted</td>
</tr>
<tr>
<td>Active_quiz</td>
<td>attempt_started</td>
<td>Attempt started</td>
</tr>
</tbody>
</table>

### Mining and analysis

This is the main part of the study and consisting of process discovery, conformance checking, and analysis steps. Process discovery was done using the heuristics miner on ProM tools [2]. Heuristics miner was chosen in this study because this algorithm is one of the most used algorithms for process discovery that can deal with noise and can be used to express the main behaviour recorded in an event log. In ProM, we used the interactive Data-aware Heuristics Miner (iDHM) plugin that allows interactive settings including several options of graphs to present the resulted process models.

The process model discovered using heuristics miner was then checked for its conformance to the traces in the event log, and vice versa. Conformance checking was done by measuring fitness, precision, and generalization [1], following the formulas:

1. **Fitness** ($Q_f$) indicates how well the resulting model represents traces in the event log.

   $$Q_f = \frac{C_{me}}{\min(C_{me}, C_{em})}$$

   Where
   - $C_{me} = \text{cost of aligning model to the event log}$
   - $C_{em} = \text{cost of aligning event log to the model}$

2. **Precision** ($Q_p$) indicates how accurate the model describes the traces in the event log. A model with low precision causes underfitting.

   $$Q_p = \frac{n_{ae}}{\text{count}(n_{am})}$$

   Where
   - $n_{ae} = \text{number of activities in the log}$
   - $n_{am} = \text{number of activities in the model}$
3. **Generalization** \( Q_g \) indicates the ability of the model to be generalized, that is, it can describe traces that are not in the event log. A model with low generalization causes overfitting.

\[
Q_g = 1 - \frac{\sum (\sqrt{e})^{-1}}{n_m}
\]

Where
- \( n \) = number of nodes
- \( e \) = number of executions
- \( n_m \) = number of nodes in the model

**E. Evaluation**

Evaluation was done through an intensive discussion with the CeLOE team and further interview with representative of lecturers and questionnaire students to verify the findings. The detail about evaluation, and the finding discussed at section 4 about evaluation. More details about the research stages including the evaluation are described in Section 3.

**III. RESULTS AND DISCUSSION**

In the mining and analysis phase, process discovery is carried out using the heuristic miner algorithm on ProM tools which produces a process model. After getting the process models, then the results are converted into Petri net to make sure the models are consistent for conformance checking.

**A. Mining and analysis of Lecturer activities**

The discovered process model from the lecturer event log is shown as a directly-follow graph in Figure 2. Conformance checking was done and shown that the process model has a fitness of 87.196%, precision 37.472%, and generalization 99.9%. Those values indicate that the process model has a high conformance to the event log (high fitness) and can be generalized well (high generalization), but precision is low.

Figure 2 also shows that lecturers performed several activities, with the three most frequent activities [frequency; percentage] being Assignment [14,410; 85.459%], Quiz [1,349; 8.001%], and Forum [1,040; 6.168%], as shown in Table 4. It is also shown in Table 4 that Chat and Data are rarely used by the lecturers during their teaching activities through the LMS. This finding suggests potential recommendation to promote Chat and Data to be used more often by the lecturers.

**TABLE 4. LECTURER ACTIVITIES**

<table>
<thead>
<tr>
<th>Lectures Activities</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment</td>
<td>14,410</td>
<td>85.459%</td>
</tr>
<tr>
<td>Quiz</td>
<td>1,349</td>
<td>8.001%</td>
</tr>
<tr>
<td>Forum</td>
<td>1,040</td>
<td>6.168%</td>
</tr>
<tr>
<td>Chat</td>
<td>11</td>
<td>0.065%</td>
</tr>
<tr>
<td>Data</td>
<td>3</td>
<td>0.018%</td>
</tr>
</tbody>
</table>

Deeper analysis was done by checking the first and the last activities of lecturer traces, as presented in Figure 3. The most frequent first activities are Forum (21; 81%), Assignment (4; 15%), or Quiz (1; 4%). This finding reflects the reality based on interview to the lecturers (show on section 4. Evaluation), they are started using the LMS during a semester by a forum, assignment, or quiz. The most frequent last activities are Assignment (17; 65%), Forum (6; 23.1%), or Quiz (3; 11.5%). This finding also reflects the reality where lecturers are mostly used Assignment, Forum, or Quiz to end their teaching during the semester.

![Figure 3](image-url)  
**Figure 3.** The first and last activities of lecturers

More detailed analysis also can be done by examining the dotted chart, as presented in Figure 3. It is shown in the dotted chart the most frequent activity is Assignment (light blue dots). This reflects the fact that most lecturers use LMS mostly for managing assignments for the students. The dotted chart also shows a high variance of the sequence of activities done by lecturers. The duration ranges from a few days up to nearly 5 months. This reflects the real situation where some lecturers have used LMS much more that the other lecturers.
B. Mining and analysis of Student activities

The resulted model process of student log is presented as a directly-follow graph in Figure 5. Conformance checking of this model shows a fitness of 97.0%, precision 15.4%, and generalization 99.9%. Those values show that the process model is highly represent the event log (high fitness) and can be generalizable (high generalization) but is low in precision.

Deeper analysis was done by checking the first and the last activities done by students, as presented in Figure 6. The top three most frequent first activities [frequency; percentage] are Forum [1,509; 38%], Resource [1,081; 27%], and URL [801; 20%]. This finding reflects the reality from questionnaire to students (show on section 4. Evaluation), where most students accessed Forum, Resource, or URL in the beginning of their learning activities in the LMS. Other activities are infrequent (less than 10%), which are: Assignment, Quiz, H5P, Page, Chat, Glossary, and Book.

The top three most frequent last activities [frequency; percentage] are Assignment [1,540; 39%], Quiz [1,427; 36%], and Forum [411; 10%]. This finding reflects the reality where most students accessed Assignment, Quiz, or Forum in the end of their learning activities during the semester. It also related to the fact that most lecturers give assignments or quizzes to assess student understanding in the end of semester, or forum to discuss further topics in the learning activities. Other activities are infrequent (less than 10%), which are Resource, URL, Page, H5P, Folder, and Feedback.

Comparing the last activities of the lecturers (Figure 3) and the students (Figure 6), we have got similar list of the three most frequent last activities, which are Assignment, Quiz, and Forum. But, comparing the first activities of the lecturers and the students, we found that there are differences in their top infrequently used by students (frequency less than 1%) are Page, Folder, Chat, Choice, Activequiz, Glossary, Book, and Feedback.

### Table 5. Student Activities

<table>
<thead>
<tr>
<th>Students Activities</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource</td>
<td>48,207</td>
<td>27.027%</td>
</tr>
<tr>
<td>Assignment</td>
<td>36,312</td>
<td>20.358%</td>
</tr>
<tr>
<td>URL</td>
<td>35,902</td>
<td>20.128%</td>
</tr>
<tr>
<td>Forum</td>
<td>28,639</td>
<td>16.056%</td>
</tr>
<tr>
<td>Quiz</td>
<td>20,835</td>
<td>11.681%</td>
</tr>
<tr>
<td>H5P</td>
<td>6,204</td>
<td>3.478%</td>
</tr>
<tr>
<td>Page</td>
<td>1,004</td>
<td>0.563%</td>
</tr>
<tr>
<td>Folder</td>
<td>953</td>
<td>0.534%</td>
</tr>
<tr>
<td>Chat</td>
<td>134</td>
<td>0.075%</td>
</tr>
<tr>
<td>Choice</td>
<td>91</td>
<td>0.051%</td>
</tr>
<tr>
<td>Activequiz</td>
<td>45</td>
<td>0.025%</td>
</tr>
<tr>
<td>Glossary</td>
<td>29</td>
<td>0.016%</td>
</tr>
<tr>
<td>Book</td>
<td>10</td>
<td>0.006%</td>
</tr>
<tr>
<td>Feedback</td>
<td>2</td>
<td>0.001%</td>
</tr>
</tbody>
</table>

It can also be seen in Figure 5 that students performed several activities, with the top 6 most frequent activities are Resource, Assignment, URL, Forum, Quiz, and H5P. More detailed list of student activities is presented in Table 5. The three most frequent activities [frequency; percentage] are Resource [48,207; 27.027%], Assignment [36,312; 20.358%], and URL [35,902; 20.128%]. The other eight activities that are
three most frequent first activities. Most lecturers used Assignment, Quiz, and Forum, while students used Resource, Assignment, and URL. In the list of the most frequent first activities of students, Forum is the fourth. This can be explained rationally that students needed to check on the Resource and URL to help them done Assignment, Quiz, and discuss further in the Forum.

More detailed analysis was done by examining the trace variant diagram. Figure 7 shows the top 8 most frequent trace variants covers 488 out of 3,994 traces (12.22%). Each row in a trace variant diagram represents a trace variant. The rows in the trace variant diagram are ordered from the most frequent variant to the least frequent one. Each coloured shape represents an activity. Figure 5 show that the top four variants consist of one activity each. There are 186 students (4.66%) only used Forum, 74 students (1.85%) only used URL, 64 students (1.6%) only used Quiz, and 50 students (1.25%) only used Resource.

![Figure 7. Top 8 trace variants of student activities](image)

**IV. EVALUATION**

Evaluation was done through interview and questionnaire to lecturer and student. The evaluation was especially done by questionnaire to 34 students’ respondent and interview to five lecturers. Questionnaire containing questions about what their activities at the beginning and the end of their class for each semester. Students can choose more than one activity for each question from the questionnaire. The lecturer respondents interviewed about whether the process model (figure 2) is compatible with their activities in the LMS and how they start and ending their semester.

From interview to five lecturers, they agree with the process model was developed use process mining show on figure 2. Three lectures initiate their lecturing activity in LMS with forum. In their first post at the forum, they can inform the student how they run the class. For the example at their first post in the forum, the lecturer usually publishes their email or video conference link so the student can contact their class lecture at the time. And the two other lecture they use quiz or assignment at the beginning of the semester. Quiz or assignment containing pre-test. The pre-test is used to know the student knowledge about each topic. So, at the class meeting using video conference the lecture can initiate the material subject based on the student pre-test result.

At the end of semester, the lecturer agree that they used assignment and quiz to take final score at that semester. And forum used to know the student’s opinion about student learning experience in that semester. The last forum is usually used to discuss about every topic in the class meet or the other topic.

Based on the 1st question result, the first activity that students most accessed is Forum (25%) then followed by resource file (25%), quiz (19%), assignment (17%), H5P (10%) and URL (3%). Both result between the most first students activity describe in process model (figure 5 and figure 6) and first activity from questionnaire result (figure 8) is forum. So between first activity in process model and the evaluation is match.

![Figure 8. Students activities that accessed at the beginning of semester](image)

The result from 2nd question about the activities that access at the end of semester is assignment (30%), then followed by quiz (27%), resource file (19%), H5P (10%), forum (8%), and URL (6%) that shown on figure 9. From the result of questionnaire that assignment is the most activity was accessed by student at the end of the semester. The questionnaire result match to process model from figure 5 and last activities on figure 6.

![Figure 9. Students activities that accessed at the end of semester](image)
V. CONCLUSIONS

Based on the study, it can be concluded that: (1) Process mining can be used to analyse the teaching/learning activities carried out by lecturers and students during a semester. In this study, the most frequent activities carried out by lecturers in the Study Program 365 are assessments, among other activities including chats, quizzes, and forums. Meanwhile, the most frequent activities carried out by students in the Study Program 365 are assignments and quizzes. (2) Process mining can also be used to analyse the sequence of activities during the teaching/learning process. We analysed the durations of lecturer activities and found that the durations ranged from a few days up to five months. We also analysed the first and last activities carried out by the lecturers and students. The most frequent first activities of the lecturers are the Assignment, Forum, and Quiz; while the most frequent first activities of the students are Forum, Resource, and URL. (3) The findings of the study experiments can be used to improve understanding on how the process has been carried out and suggest improvements. Based on the findings of the most frequent activities carried out by lecturers and students, we can suggest the University to promote other functionalities to be used by the lecturers and students. For example, the lecturers should be introduced to the LMS functionalities other than assignments, forums, and quizzes.

The evaluation about the finding in this study is done using interview five lecturer about the process model. And the other done by the questionnaire to 34 students about this study finding student process model. And the result is booth match between the finding and the evaluation.

This study can potentially be reproduced and further developed to implement process mining in a larger scope of study. Potential future work is to apply this approach to analyse and compare teaching/learning process in several study program; to improve data processing approach by generating ontology mapping on the activities in the process; and to recommend improvements in the teaching/learning process using the LMS.

ACKNOWLEDGMENT

This research was funded by an internal grant with the contract number KWR.0037/PNLT3/PPM-LIT/2021 from the Telkom University.

REFERENCES


Gede Agung Ary Wisudiawan, M.T. is a lecturer at Telkom University, Bandung Indonesia. He graduated from Bandung Institute of Technology (2012). He teaches at the Bachelor of Informatics, Software Engineering, and Information Technology study programs. His research area is software engineering specialty at learning management system. He is also a member of Tel-C (Technology Enhanced Learning Center Research) research laboratories.

Angelina Prima Kurniati, Ph.D. is a senior lecturer at Telkom University, Bandung, Indonesia. She graduated from her doctoral degree in Computer Science at the School of Computing, University of Leeds (2020). She is a member of IEEE Indonesia section. She teaches at the Bachelor of Informatics, Software Engineering, and Information Technology study programs. The majors she delivered are data mining, IT project management, information quality, among others. Since 2020, she is also the Head of Secretay at the School of Computing, Telkom University. Her current research projects are on process mining, project management, and data science.