E-learning process analysis to determining student learning patterns using process mining approach

A R Hakim, M A Hasibuan, R Andreswari
Department of Information System, School of Industrial and System Engineering
Telkom University Bandung, Indonesia

Abstract. In Learning Management System (LMS) setting, event log is a historical record which contains a series of user activities that recorded on the system. Process mining is applied to discover e-Learning usage patterns. In this research, heuristic review algorithm is used to create a process model to do analysis. The process mining tools that used are Disco and ProM, their work will be finding the process model and evaluating it. The object of research are two different courses, System Defined Network (SDN) and Basic Telecommunication System (BTS) in different levels (Diploma and Bachelor). In order to know the pattern of using e-Learning at different level, Heuristic Miner can model the event log into the process model well. The measurement of fitness value for SDN and BTS are 0.974 and 0.986. Moreover, SDN course tends to access assignment module as many as 938 times. While BTS course tends to access Forum module with the highest frequency is 639 times.

Keywords—event log, process mining, heuristic Mining, Learning Management System, e-learning

1. Introduction
The rapid growth of data becomes a challenge for every institution. Data is an important component in an institution. Data influence in making decisions to be evaluated. Basically, the data describes a fact that reflects a value. Data processing with analysis and visualization in the learning process at the university can help lecturers know the current student performance, by highlighting the beneficial and supportive information in making decision [1]. Most education institution have used information and communication technology in education, for example Learning Management System (LMS) or virtual learning environment. More than 85% of colleges in the UK use the LMS in the early of 2003. Then in the late 1990s, about 90% of learning management systems have been used in American colleges. As well as in Czech Republic, all the institutions or colleges except the arts-oriented ones have used the LMS [2]. LMS contain a lot of event log related information to the series of uses in the learning process. Event log is a historical record which contains a series of user activities recorded on the system. Process mining is applied to discover the e-Learning usage patterns. The research object of this research are two different courses, System Defined Network (SDN) in Bachelor level and Basic Telecommunication System (BTS) in Diploma level in order to know the pattern of using e-Learning at the same study program in a different level.

2. Literature Review
2.1. Learning Management System
According to the educational point of view, the LMS aims to design and implement various types of quiz-based learning activities. According to the technology point of view, the core components of the LMS are an event log system, containing a set of data from student and lecturer activities in the system...
2. E-Learning is a system that utilizes information technology to provide learning activities through digital media. Currently, E-Learning is developed in various institutions to support the learning process and make it easier, especially for students to get the lesson given by the lecturer. E-Learning provides an information technology-based education tool, focusing on the effectiveness of distance learning in a variety subject. E-Learning as a supporter of the face-to-face teaching process as well as considering student learning patterns. E-Learning can show an improvement in teaching, the overall effectiveness and success gained from E-Learning and continue to be developed [3].

2.2. Process Mining

In general, there are three processes in process mining: discovery, conformance checking, and enhancement, as shown in figure 1. The following are the detail activity for each process.

1. Discovery. This technique uses event log and generates a process model, thus obtaining information or actors behavior involved and events that have occurred based on the event log recorded in the system.

2. Conformance. Conformance checking is testing the suitability between process models and event logs [5].

3. Enhancement. This technique is processing event log with the model to get a new model. The goal is to be able to improve the existing process model by using information based on the event log.

The progress of data mining makes it possible to find valuable patterns in large datasets to make decisions based on that dataset. However, data mining such as Classification, Clustering, Regression, Association Rule, and Sequence Mining do not focus on the process of obtaining such information [6].

2.3. Heuristic Miner

Heuristic Miner is a process mining algorithm developed by Dr. Ton Weitjers, who uses a heuristic approach, heuristic approach is an approach in teaching by displaying data and drawing conclusions using the data. The heuristic approach is used to resolve the deficiencies found in the Alpha algorithm, as it cannot resolve length-one-loop, invisible task, implicit place, and non-free-choice [7].

Representations of Heuristic Miners are similar to casual nets. This algorithm mining control-flow perspective from existing process models. The Heuristic Miner algorithm can handle a lot of noise in the event log and can also show the main behavior of a process model where not all details and exceptions are shown.

2.4. Petri Net

Petri Net is a bipartite graph, so there are no arrows connecting two transitions or two places. In graph notation, place is represented by a circle, the transition is denoted by a square, and the connector with
an arrow. Information about each event and state is expressed by transition and place. Petri Net is a tool that aims to model a discrete event system that previously known was automata, each automata can be changed to Petri Net [8].

2.5. Tools Process Mining

Disco is used to analyze and visualize data. Disco is also used to view statistical data processing results. Disco can check and display each case by displaying detailed habits of a process so that, it can be identified. Disco can also export data that has been identified into MXML and XES formats so that this tool is suitable for combination with the use of ProM tools [9].

3. Research Methodology

3.1. Collecting Data

In this phase, the authors took the event log of the System Defined Network (SDN) and Basic Telecommunication System (BTS) courses because the purpose of this study is to see the differences of student patterns towards the use of Telkom University's e-Learning in two different courses with different levels, namely Diploma of Telecommunication Engineering and Bachelor of Telecommunications Engineering, the two event logs that are carried out by adjusting the data requirements in this study will be in the next phase.

| No | Course Name       | Study Program                    | Total log (rows) |
|----|-------------------|----------------------------------|------------------|
| 1  | Basic Telecommunication System | Diploma of Telecommunication Engineering | 34.005 |
| 2  | Software Defined Network | Bachelor of Telecommunications Engineering | 26.361 |

Table 2. List of Raw Data Attributes

| No | Attribute     | Description                                                                 |
|----|---------------|----------------------------------------------------------------------------|
| 1  | Time         | The time when the user accesses is in the form of Timestamp                  |
| 2  | User Full Name | User name used in Telkom University E-Learning                                |
| 3  | Affected User | Other users involved in activities in the main user access                    |
| 4  | Event Context | Activities carried out by users when accessing the menus in Telkom University's E-Learning |
| 5  | Component     | Activity label that is carried out when access to menus in Telkom University E-Learning |
| 6  | Event Name    | Action Status of the user when accessing the menu in the Telkom University E-Learning |
| 7  | Description   | Details of activities carried out by users when accessing menus on Telkom University E-Learning |
| 8  | Origin        | The media used for access, with the contents of the 'web' attribute          |
| 9  | IP Address    | IP addresses used by users when accessing Telkom University E-Learning        |
3.2. Data Preprocessing
Data preprocessing is a process to prepare raw data into a quality input data so that process mining can be done. The system stores data generated when students engage in learning activities. Data in the form of logs stored in the system automatically as a result of learning activities conducted in the database.

3.3. Data processing
Preprocessing on event log in the SDN and BTS courses is carried out first with data cleaning process using Pentaho Data Integration. With the Pentaho Data Integration scheme as in figure 2, the attributes that are not used as input from process mining are discarded.

The important attribute needed is Case ID by analyzing and sorting based on Time and User. After sorting and determining the Case ID based on the time of each activity in one Case ID, the next is to determine the status of each user, by adjusting the name of the user with the name of the lecturer holding the relevant subject, then the data will be deleted which has lecturer status, because at this research focuses on users with status as students, and deleting duplicate data that has the exact same line, deleting users with lecturer status and duplicating data with the Pentaho application data integration again, with the Pentaho Data Integration scheme.

3.4. Discovery
Data discovery is done to find modeling, by using an application to be able to check data from the previous CSV file, using Disco Fluxicon tools. This tool checks the data from CSV files that have passed the preprocessing and sorting stages. The first stage of Discovery data performs CSV input from preprocessing results by setting parameters to determine columns. Then the input to the model checking returns before the data output becomes MXML file format, by specifying Timestamp, Case, Resource, and Activity.

4. Results and Discussion

4.1. Analysis Fitness Value using Heuristic Miner and Conformance Checker
1) The Result of analysis on Software Defined Network Courses (SDN) Course
Based on the testing result on Relative-to-best Threshold (RT) as shown in table III, the changing of RT has no significant impact on fitness value. The same situation also happens on Dependency Threshold (DT) as seen in table IV, whereas its changing value has no effect on the changing of fitness value. Based on the test results as shown in table V, the value of the Positive Observations Threshold (PT) parameter changed has an effect on the fitness value. The increasing number on PT has an effect on the increasing value of fitness.

| Table 3. Change in RT Value (SDN) |
| RT  | PT | DT | Fitness   |
|-----|----|----|-----------|
| 0.05| 50 | 0.9| 0.9744828 |
| 0.25| 50 | 0.9| 0.9744828 |
| 0.50| 50 | 0.9| 0.9744828 |
| 0.75| 50 | 0.9| 0.9744828 |
| 1.00| 50 | 0.9| 0.9744828 |

| Table 4. Change in RT Value (SDN) |
| RT  | PT | DT | Fitness   |
|-----|----|----|-----------|
| 0.05| 50 | 0.9| 0.9744828 |
| 0.05| 50 | 0.7| 0.9744828 |
| 0.05| 50 | 0.5| 0.9744828 |
| 0.05| 50 | 0.3| 0.9744828 |
| 0.05| 50 | 0.1| 0.9744828 |
2) **Analysis on Basic Telecommunication System (BTS) Course**

Same with the SDN course, the changing value of RT and DT on BTS course has no effect in changing value to fitness value. BTS course has different result with SDN course regarding the effect of changing value on PT to fitness value. If in the SDN course the changing value of PT has an effect on fitness value, while in BTS course there is no effect on fitness value.

| Table 5. Change in RT Values (BTS) |
|-----------------------------------|
| RT   | PT    | DT    | Fitness   |
| 0.05 | 15    | 0.9   | 0.98651683 |
| 0.05 | 15    | 0.7   | 0.98651683 |
| 0.05 | 15    | 0.5   | 0.98651683 |
| 0.05 | 15    | 0.3   | 0.98651683 |
| 0.05 | 15    | 0.1   | 0.98651683 |

| Table 6. Change in DT Values (BTS) |
|-----------------------------------|
| RT   | PT    | DT    | Fitness   |
| 0.05 | 15    | 0.9   | 0.98651683 |
| 0.25 | 15    | 0.9   | 0.98651683 |
| 0.50 | 15    | 0.9   | 0.98651683 |
| 0.75 | 15    | 0.9   | 0.98651683 |
| 1.00 | 15    | 0.9   | 0.98651683 |

| Table 7. Change in PT Values (BTS) |
|-----------------------------------|
| RT   | PT    | DT    | Fitness   |
| 0.05 | 10    | 0.9   | 0.89838064 |
| 0.05 | 15    | 0.9   | 0.98651683 |
| 0.05 | 20    | 0.9   | 0.98651683 |
| 0.05 | 25    | 0.9   | 0.98651683 |
| 0.05 | 50    | 0.9   | 0.98651683 |

### 4.2 Conformance analysis

Using the heuristic miner algorithm, the process models of both courses are developed with considering three parameters (RT, PT and DT). The value with each parameter is different from both courses. SDN course use 50 for PT, 0.05 for RT and 0.9 for DT, while in BTS course use 15 for PT, 0.05 for RT and 0.9 for DT. The result process model for both courses can been seen in figure 2 and figure 3.
4.3. Discussion

The results of Conformance Checking consist of 3 parts: fitness, precision, and structure. The first discussion will discuss the fitness value; the SDN course shows the value of 0.9744828. This value indicates that the process model can model the event log properly, based on the results of the Conformance Checker, there are 5 failed tasks and 8 remaining tasks. Failed tasks and remaining tasks are generated by less than perfect event log replays. BTS shows the value of 0.98651683. This value indicates that the process model can model the event log well, based on the Conformance Checker, there are 4 failed tasks and 10 remaining tasks.

On the precision value of the SDN course shows the value of Advanced Behavioral Appropriateness 0.80877197 and the value of Degree of Model Flexibility 0.3035714. The value of the Precision Basic Telecommunication System shows the value of Advanced Behavioral Appropriateness 0.76736844 and the value of Degree of Model Flexibility is 0.30555555. This value indicates that the process model is less flexible so that only a few allow task variations to be allowed on the log. Seen in the process model that has enough paths to display or represent sequences of tasks that
take place according to the event log. If the process model shows results that are not flexible at all, then the results of the Advanced Behavioral Appropriateness show a value of 1 and the results of the Degree of Model Flexibility show a value of 0.

5. Conclusion
The conclusions obtained from the analysis in this study are
1. The pattern of student learning behavior in Telkom University e-Learning is as follows:
   a. Activities for courses Software Defined Network (SDN): Login, interactive content, files, forums / assignments, logout.
   b. Activities for Basic Telecommunication System (BTS) courses: Login, forum, assignment, quiz, file / interactive content, logout.
   c. To access the quiz, SDN courses do not become active activities used by students
2. Heuristic Miners can model event logs into the process model quite well, the fitness value for SDN is 0.9744828 and BTS is 0.98651683. There are three parameters that affect the fitness value of SDN and BTS. The parameter with the biggest influence is Positive Threshold observations, because the frequency of the two courses is different.
3. The frequency obtained by analyzing the originator with task matrix can provide an overview of student patterns, by looking at the trend of using e-Learning in Telkom University, the subject of SDN tends to access Assignment with the highest total frequency of 938 times access. In the course of BTS the tendency to access the Forum with the highest total frequency is 639 times.
4. Raw log events that are used are not structured to be applied to mining processes, by determining the case_id based on the timestamp and user, from every user who accesses Telkom University e-learning, using the access assumption with the access time delay or the session timeout not more than 2 hours and after that using assume login and logout from the first and last activity in the case_id. As well as preprocessing on the event log, each case from access is made more general to help analysis in this study.

References
[1] Cristóbal R and Sebastián V 2010 Educational Data Mining : A Review of the State of the Art, 40(6) pp 601–618
[2] Daellenbach, Hans G, McNickle, Donald C, Palgrave Macmilan 2005 Management science, Decision making through systems thinking
[3] Juhaňák L, Zounek J, Rohliková L 2017 Using process mining to analyze students’ quiz-taking behavior patterns in a learning management system. Computers in Human Behavior.
[4] Hubackova S 2015 History and Perspectives of Elearning. Procedia - Social and Behavioral Sciences p 191 pp 1187–1190.
[5] Aalst W 2011 Process Mining: Discovery, Conformance, and Enhancement of Business Processes
[6] Aalst W, Weijters A, Maruster L Workflow Mining : Discovering process models from event logs
[7] Wen L, Van Der Aalst W M P, Wang J, Sun J. Mining Process Models with Non-Free-Choice Constructs Online:http://wwwis.win.tue.nl/~wvdaalst/publications/p394.pdf pp 1-32
[8] Darmawan A T, Kurniati A P, Atastina I 2014 Evaluasi Proses Bisnis ERP dengan Menggunakan Process Mining Studi Kasus : Fresh Food Inventory LOTTEMART BANDUNpp 1–6.
[9] Günther C W and Rozinat A 2012 Disco: Discover your processes CEUR Workshop Proceedings 936 pp 40–44.
[10] Dongen B Alves de Medeiros K, Verbeek H M W, Weijters J M M, van der Aalst W 2005 The ProM framework: A new era in process mining tool support Application and Theory of Petri Nets 3536 pp 444–454.