JURIKOM (Jurnal Riset Komputer), Vol. 9 No. 4, Agustus 2022
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v9i4.4769
Hal 1128−1135
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Process Mining using Inductive Miner Algorithm to Determine the actual Business Process Model

Muhammad Wanda Wibisono, Angelina Prima Kurniati*, Gede Agung Ary Wisudiawan

School of Computing, Telkom University, Bandung, Indonesia
Email: [1]wandawibisono@student.telkomuniversity.ac.id, [2,*]angelina@telkomuniversity.ac.id, [3] degunk@telkomuniversity.ac.id
Email Penulis Korespondensi: angelina@telkomuniversity.ac.id
Submitted **25-08-2022**; Accepted **30-08-2022**; Published **30-08-2022**

Abstract

At the beginning of 2019, the COVID-19 pandemic entered the country of Indonesia resulting in all learning activities being carried out online in all cities of Indonesia. Likewise, Telkom University concentrates all teaching and learning activities online using the CeLOE Learning Management System. Learning Management System is a system that helps lecturers in managing teaching and learning activities independently in educational institutions. CeLOE is a learning management system of Telkom University developed based on Moodle. In this study, we analyse the CeLOE event log using the process mining method. The goal is to find out the learning patterns of students using CeLOE during the COVID-19 pandemic. This research case study focuses on the activities of students of the Telkom University S1 Informatics study program for the first semester of 2020/2021 in using CeLOE LMS. The analysis of this study conducted a comparison of the performance of three variants of the inductive miner (IM) algorithm through conformance checking values. The results of the analysis obtained are value of conformance checking from the three variants of the inductive miner (IM) algorithm have an average fitness value of up to 1 prove that the inductive miner (IM) algorithm can make a model based on the event log well. Besides that, it has a fairly high precision value with a value range of 0.750-0.850 shows that the inductive miner (IM) makes a process model with relatively many variations of activities outside the event log and the IM process model is "overfit-ting" for all variants of the IM algorithm. Inductive miner (IM) is the best inductive miner (IM) algorithm variant with a fitness value of 1.0, precision value of 0.750, and the generalization value of this algorithm is relatively high (0.984). It is hoped that this research can contribute to the addition of new perspectives related to the implementation of process mining using inductive miner (IM) algorithm in the field of education.

Kata Kunci: Learning Management System; CeLOE; Process Mining; Inductive Miner; Event Log

1. INTRODUCTION

In 2018 Telkom University started to use a platform to help the online learning process called CeLOE LMS (Center for e-Learning and Open Education Learning Management System). CeLOE is a Moodle-based learning management system. Moodle is a software that can convert learning media into web forms. The features provided by Moodle in supporting learning include (1) videos; (2) discussion forums; (3) chat; (4) learning materials; and (5) quizzes [1].

Every activity of students and lecturers in using CeLOE LMS is automatically recorded into a historical record and then stored by the system as event log data. The event log data can be analyzed using the process mining method. The purpose of the process mining of the LMS is to analyze how learning resources and activities can help the process of teaching and learning activities [2]. In the process mining method, the first process is process discovery which requires the main input data in the form of an event log, then processed to find a CeLOE business process model that displays the sequence of student and lecturer activities. Furthermore, conformance checking is checking whether the facts recorded by CeLOE LMS and recorded in the event log are conformed with the process model that has been generated from the discovery process and vice versa. [3].

This research case study focuses on student activities using CeLOE LMS and using data event log of the S1 Informatics study program for the first semester of 2020/2021. The data is used as a representation of the learning process in the study program at Telkom University. In this study, process discovery used the inductive miner (IM) algorithm. This algorithm was chosen because it is able to cope with large event logs and can cope with infrequent activities [4] therefore it is well suited in the analysis of large event logs recorded by CeLOE LMS. The model process resulting from process discovery is then analyzed in conformance checking based on fitness , precision, and generalization.

Previously related research [5]. This study implements process mining with fuzzy miner to overcome the "spaghetti-like" process model using data from the Website Application Development and Enterprise Systems course for the first semester of 2020/2021. There are also other studies [6], which uses same course data to conduct a comparative analysis of students learning patterns in programming and non-programming courses using heuristic miner algorithms. Other research [7] analyzes the process model obtained from process discovery using the heuristic miner algorithm on student and lecturer learning activities using CeLOE LMS. Based on these studies, process mining in learning activities in the learning management system can help improve the quality of learning.

2. RESEARCH METHODOLOGY

2.1 Event log

Event logs are track record of activities in the system that are stored in database, so that later the administrator can be retrieve it to identify user activities on the system or process that occur [2].

JURIKOM (Jurnal Riset Komputer), Vol. 9 No. 4, Agustus 2022
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v9i4.4769
Hal 1128−1135
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Case id	Event id	Properties				
		Timestamp	Activity	Resource	Cost	...
1	35654423	30-12-2010:11.02	Register request	Pete	50	...
	35654424	31-12-2010:10.06	Examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	Check ticket	Mike	100	...
	35654426	06-01-2011:11.18	Decide	Sara	200	...
	35654427	07-01-2011:14.24	Reject request	Pete	200	...
2	35654483	30-12-2010:11.32	Register request	Mike	50	...
	35654485	30-12-2010:12.12	Check ticket	Mike	100	...
	35654487	30-12-2010:14.16	Examine casually	Pete	400	...
	35654488	05-01-2011:11.22	Decide	Sara	200	...
	35654489	08-01-2011:12.05	Pay compensation	Ellen	200	...
3	35654521	30-12-2010:14.32	Register request	Pete	50	...
	35654522	30-12-2010:15.06	Examine casually	Mike	400	...
	35654524	30-12-2010:16.34	Check ticket	Ellen	100	...
	35654525	06-01-2011:09.18	Decide	Sara	200	...
	35654526	06-01-2011:12.18	Reinitiate request	Sara	200	...
	35654527	06-01-2011:13.06	Examine thoroughly	Sean	400	...
	35654530	08-01-2011:11.43	Check ticket	Pete	100	...
	35654531	09-01-2011:09.55	Decide	Sara	200	...
	35654533	15-01-2011:10.45	Pay compensation	Ellen	200	...
4	35654641	06-01-2011:15.02	Register request	Pete	50	...
	35654643	07-01-2011:12.06	Check ticket	Mike	100	...
	35654644	08-01-2011:14.43	Examine thoroughly	Sean	400	...
	35654645	09-01-2011:12.02	Decide	Sara	200	...
	35654647	12-01-2011:15.44	Reject request	Ellen	200	...
...	...	...	...	...	...	...

Figure 1. Example of event log [3]

Figure 1 is an example of an event log. In an event log in general, there is at least a case that contains some activity, the timestamp the activity occurred, and the id of each event that occurs [3]. Because these three components are the main materials in carrying out the mining process.

2.2 Process Mining

Process mining is a relatively young research that bridges between machine learning and data mining on the one hand and process modeling and analysis on the other [8]. The purpose of process mining is to find, monitor, and improve the actual process by retrieving the necessary information from log files (generally in the form of event logs) recorded by the information system of a company [3]. There are three stages in the mining process:

a. Process Discovery : This stage requires input data in the form of event logs to create a business process model [8].
b. Conformance Checking : Check the comformity between the process model generated from process discovery and the real events in the event log.
c. Enhancement : Improve the existing process model using the information about process actual contained in some event logs [8].

In this study, we will compare the three-dimensional value of conformance checking quality, fitness , precision, and generalization from three variants of inductive miner algorithms.

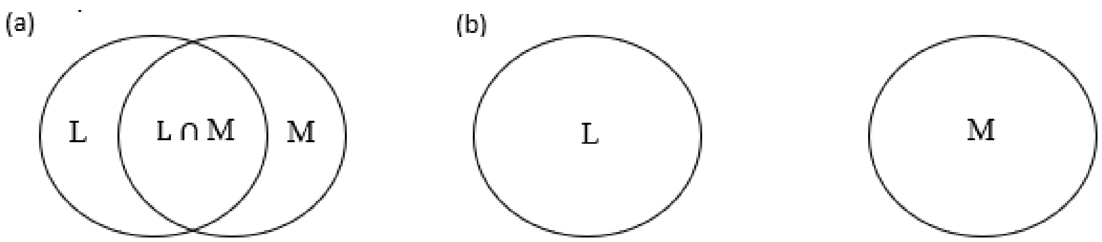

Figure 2. (**a**) Visualization overlap of recorded (L) and Modelled (M) behaviour. (**b**) Visualization of recorded (L) and modelled (M) behaviour [9].

a. Fitness

Measures the ability of a process model to reproduce the execution of a process in the event log and vice versa. Fitness can be stated as follows.

$$Fitness = \frac{|L \cap M|}{|L|} \quad (1)$$

b. Precision

Indicates that the process model does not have to show a process that tends to be different from that seen in the event log. Precision can be stated as follows [9].

$$Precision = \frac{|L \cap M|}{|M|} \quad (2)$$

c. Generalization

Assess the process model will be able to reproduce the behavior that has not been seen from the system [10] .

$$Q_G = 1 - \frac{\sum nodes\,(\sqrt{\#executions})^{-1}}{\#nodes\ in\ tree} \quad (3)$$

2.3 Inductive Miner

Inductive miner is one of the process mining algorithms by mining process trees from event logs. There are three variations of the inductive miner algorithm, which are inductive miner, inductive miner infrequet, and inductive miner directly follows [11]. Inductive miner (IM) uses a block-structured process model to introduce a framework that ensures a proper and complete process model [12]. Inductive miner Infrequent (IMi) adds filters to sort out infrequent activities in each inductive miner step [13]. Inductive miner directly-follows (IMd) retrieves the event log and automatically implements a series of steps, allowing users to perform process-based analysis of the logs [14].

3. RESULT AND DISCUSSION

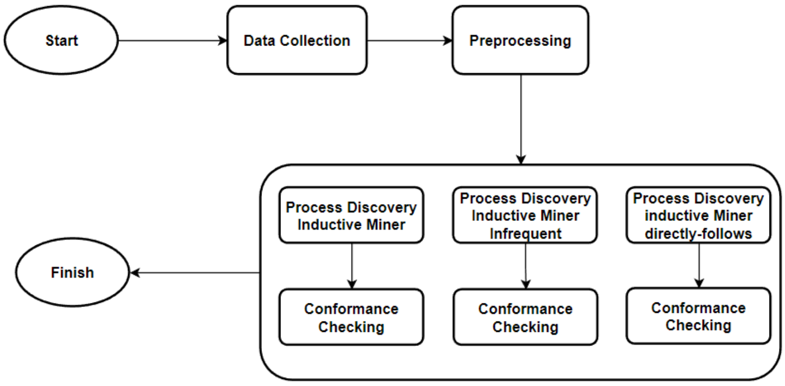

Figure 3. System design flow-chart

3.1 Data Collection

The first step is to collect student activity data using CeLOE which is automatically recorded by the system and stored as event log data. The event log contains a track record of learning activities using Telkom University's CeLOE LMS which was carried out in 2018-2021. The scope of this study uses event log data from the School of Computing, Informatics study program for the first semester of the 2020/2021.

To get the data, you need to match the courseid attribute on the mdl_course table, mdl_course_category table, and mdl_logstore_standard_log table using MySQL Workbench tools. The data consists of 12,575,554 rows and 21 columns or attributes. The event log data is still the original/raw data that needs to be preprocessed for process mining.

Table 1. Student log data attribute using CeLOE with description

No	Columns/Attribute	Description
1	**Id**	Id of the event log
2	**Eventname**	User status action when acces
3	**Component**	Component declare an event
4	**Action**	Actions on the system done by the user
5	**Target**	Target on which the action is taken
6	**Objecttable**	Database table name which represents the event object
7	**Objectid**	Id of the objecttable
8	**Crud**	Indicating ‘c’reate, ‘r’ead, ‘u’pdate, and ‘d’elete operation
9	**Edulevel**	Level of educational value of the event
10	**Contextid**	Id of contextlevel
11	**Contextlevel**	This tells you if it was a course, activity, course category, etc.
12	**Contextinstanceid**	Based on contextlevel this may courseid, course module id, etc.

JURIKOM (Jurnal Riset Komputer), Vol. 9 No. 4, Agustus 2022
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v9i4.4769
Hal 1128−1135
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

13	**Userid**	Id of user
14	**Courseid**	Id of course
15	**Relateduserid**	Other users associated with the main user
16	**Anonymous**	Actions other the unknown main user
17	**Other**	Any other fields needed for event description
18	**Timecreated**	Time when the event was triggered
19	**Origin**	Media used to access the CeLOE LMS
20	**IP**	User IP address when using CeLOE LMS
21	**realuserid**	Original user id

3.2 Preprocessing

Preprocessing is a method to process raw data into quality data. Quality data can improve the performance of an organization [15]. Quality data in process mining is useful for obtaining optimal results [16] by reducing the dimensions of the data without losing the characteristics of the data. In this study, data preprocessing was carried out by reducing columns or attributes that were not needed during the implementation of process mining, adding new columns or attributes needed when implementing process mining, and reducing data size by deleting row with missing values and redundant data.

Table 2. Sample of event log CeLOE

id	eventname	component	action	target	objecttable	objectid	crud
8093748	\core\event\course_created	core	created	course	course	2823	c
8093750	\core\event\course_section_created	core	created	course_section	course_sections	55701	c
8093751	\core\event\course_section_created	core	created	course_section	course_sections	55702	c
8093752	\core\event\course_section_created	core	created	course_section	course_sections	55703	c

Table 3. Sample of event log CeLOE

edulevel	contextid	contextlevel	contextinstanceid	userid	courseid	relateduserid	anonymous	other
1	425713	50	2823	5	2823		0	a:2:{s:9:"shortname"; s:7:"PRAK-SD";s:8:"fullname"; s:23:"Praktikum Struktur Data";}
1	425713	50	2823	5	2823		0	a:1:{s:10:"sectionnum";i:0;}
1	425713	50	2823	5	2823		0	a:1:{s:10:"sectionnum";i:1;}
1	425713	50	2823	5	2823		0	a:1:{s:10:"sectionnum";i:2;}

Table 4. Sample of event log CeLOE

timecreated	origin	ip	realuserid
1.59E+09	web	36.79.249.118	NaN
1.59E+09	web	36.79.249.118	NaN
1.59E+09	web	36.79.249.118	NaN
1.59E+09	web	36.79.249.118	NaN

Tables 2, 3, and 4 are an example of raw data from the learning process event log data using the CeLOE LMS. The data has many columns or attributes and also data that does not meet the criteria for implementing process mining.

In this study, the first step taken in data preprocessing is to delete unused columns and remove missing values and data redundancies. The next step is to select several activities with one sub-activity that is most relevant to student activities, the goal is to minimize the repetition of existing activities using DISCO tools [7]. For example, Forum activities have several sub-activities, for example viewed course modules, created discussions, and uploaded assessable. In the log, four of those sub-activities are logged with the same forum activity. In this case, the selected forum sub-activity is the viewed course module. Table 5 shows activities with selected sub-activities along with their descriptions.

Table 5. Filtering activity and sub-activity student log.

Activity	Sub Activity	Description
Forum	viewed_course_module	View activity
Glossary	viewed_course_module	View Resource
H5P	viewed_course_module	View Resource
Page	viewed_course_module	View Resource
Quiz	submitted_attempt	Submitted
Resource	viewed_course_module	View Resource
URL	viewed_course_module	View Resource
Folder	viewed_course_module	View Resource
Feedback	viewed_course_module	View activity

JURIKOM (Jurnal Riset Komputer), Vol. 9 No. 4, Agustus 2022
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v9i4.4769
Hal 1128−1135
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Choice	created_answer	Submitted
Chat	sent_message	Message Sent
Book	viewed_course_module	View Resource
Assignment	submitted_assessable	Submitted
Active_quiz	started_attempt	Attempt Started

3.3 Process Discovery

Process discovery requires event log input data to create a process model that represents the actual process running. Figure 4 shows the simple process of student learning using DISCO tools.

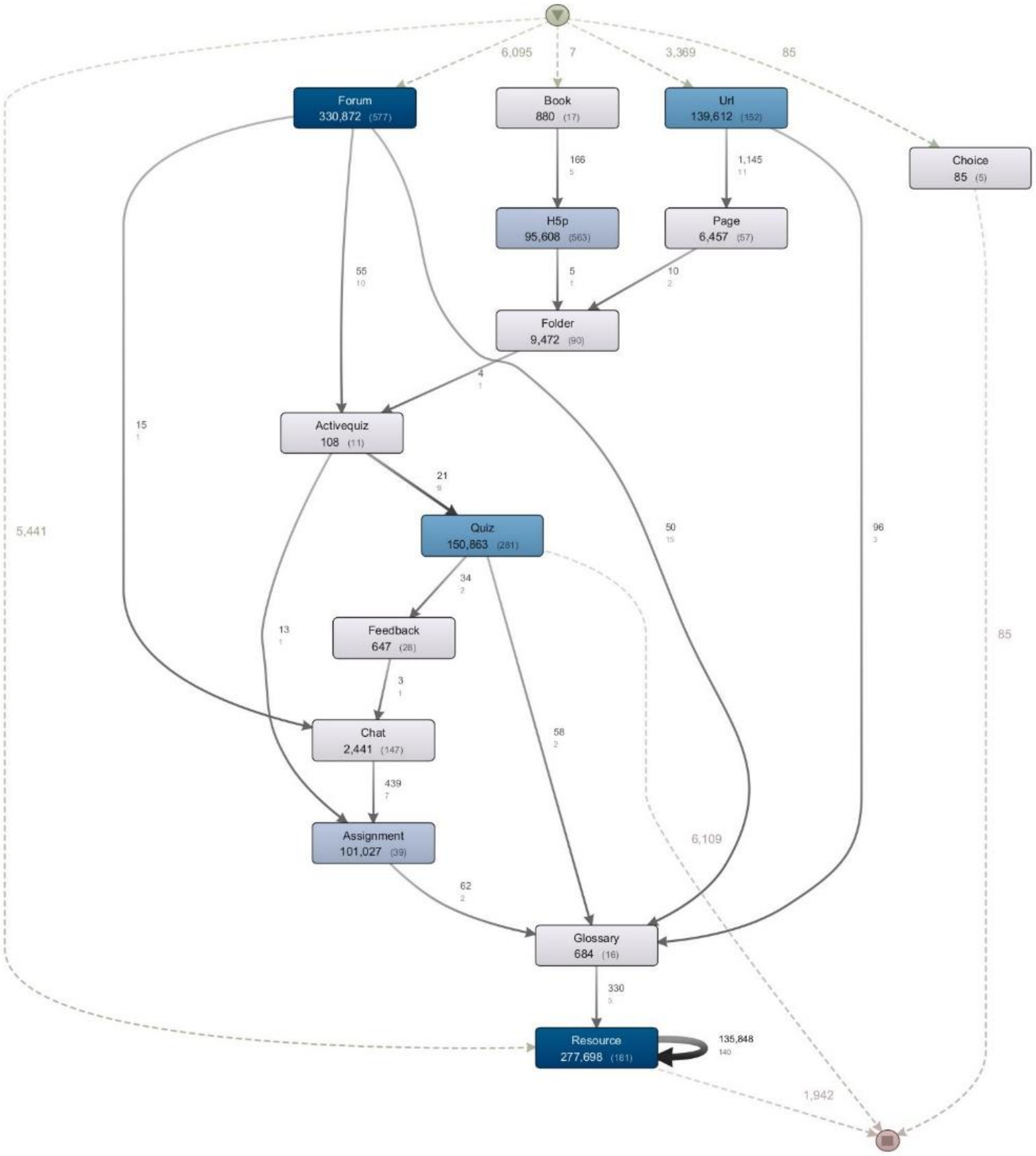

Figure 4. Student learning process model in using CeLOE LMS

From the student learning process model, it is known that students have more frequent access to forums compared to other activities. Table 6 shows the frequency of all student activities when accessing CeLOE LMS.

Table 6. Student Activities

Activity	Frequency
Forum	330872
Resource	277698
Quiz	150863
URL	139612

JURIKOM (Jurnal Riset Komputer), Vol. 9 No. 4, Agustus 2022
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v9i4.4769
Hal 1128−1135
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Assignment	101027
H5P	95608
Folder	9472
Page	6457
Chat	2441
Book	880
Glossary	684
Feedback	647
Active_quiz	108
Choice	85

This study used an inductive miner algorithm to discover the model of the student learning process using CeLOE LMS. The algorithm was chosen because the inductive miner can cope with relatively large event logs and can cope with infrequent activities .

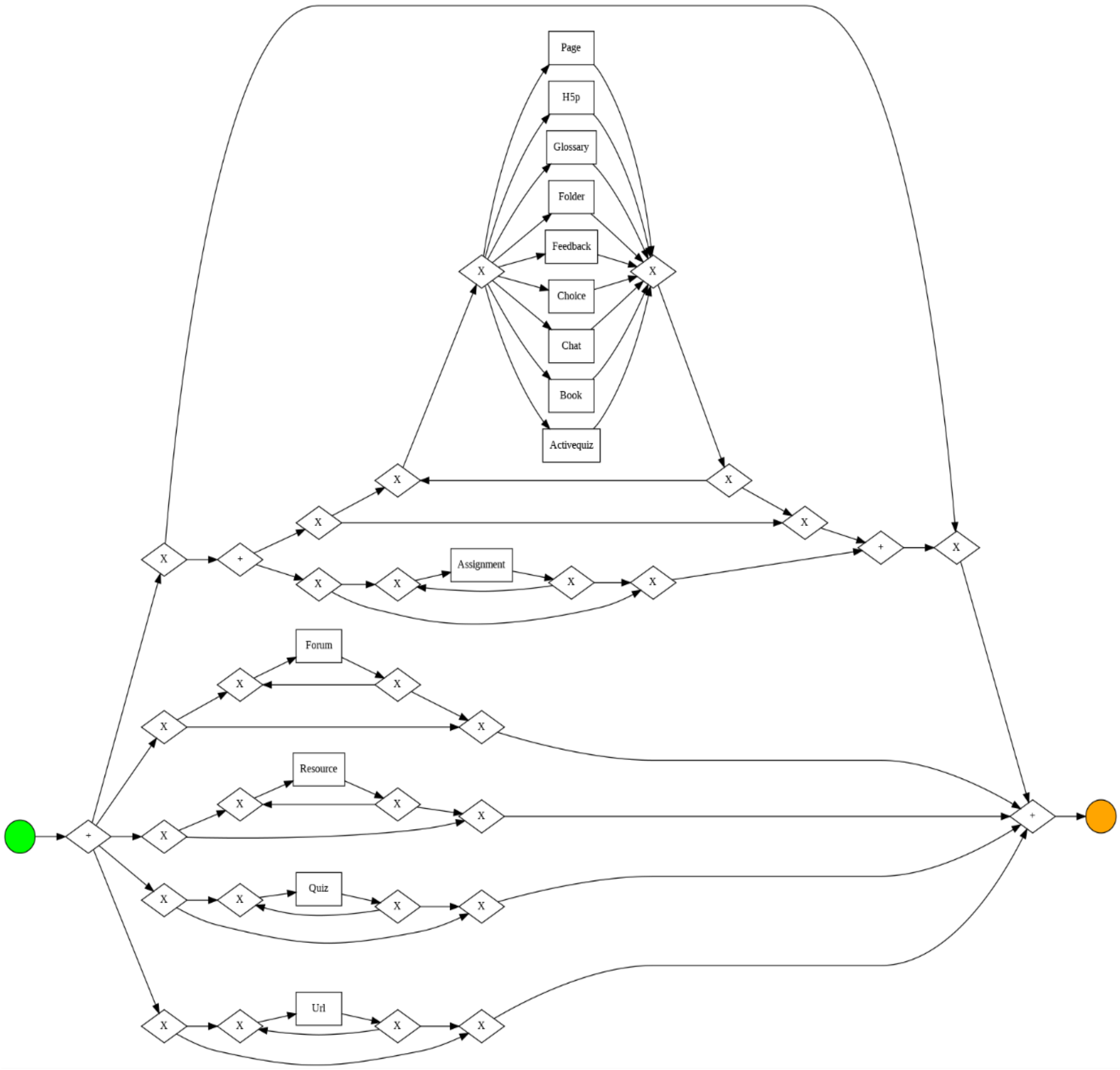

Figure 5. Student learning process model using Inductive miner

Figure 5. shows the results of process discovery using an inductive miner algorithm showing multiple activities running concurrently at a time. Furthermore, we analyze the performance of the inductive miner algorithm in making the model process.

3.4 Conformance Checking

Conformance checking method performs an analysis of the relationship between the activities depicted in the process model and the event log [17]. This study examined the comformity between the student learning process model and the CeLOE LMS event log and analyzed the performance of the inductive miner algorithm based on fitness, precision, and generalization. Conformance checking technique used by this study is the footprints technique. Because the technique is fundamental (but scalable) for comparing entities (such as event logs, DFG, Petri nets, process trees, other types of models).

Table 7. Result of conformance checking inductive miner algorithm

	Noise Threshold	Fitness	Precision	Generalization
Inductive miner	0.2	1.0	0.750	0.984
		0.9972	0.849	0.946
Inductive miner infrequent	0.4	0.9985	0.842	0.922
	0.6	0.9987	0.750	0.981
	0.8	0.9871	0.750	0.981
Inductive miner directly-follows		1.0	0.825	0.981

Table 7 shows the results of conformance checking. Based on these results, IM and IMd have a fitness value of 1.0 which means that the two algorithms can model the event log well, besides that the precision values of the two algorithms are 0.750 and 0.825 The precision value of IM is lower than the IMd showing less activity variation outside the IM event log compared to the IMd, and the generalization value of the two algorithms are 0.984 and 0.981 These results indicate that the two algorithms are "overfit-ting", it means accuracy of the models will changes when using different data.The fitness value of IMi using different noise threshold parameters has the same tendency to go towards 1.0. For this reason, IMi can still be said to be good in modeling event logs. The precision value of the IMi tends to decrease as the noise threshold gets higher. The higher the noise threshold, the less the process model makes activity variations outside the event log, and the generalization value of the IMi is relatively high, which is above 0.9, this indicates that this algorithm is overfit as well as the IM and IMd algorithms. Overall, the IM algorithm can model the event log well, and the IM process model relatively makes the variation of activities outside the event log, judging from the precision value is still relatively high and the IM process model also includes an "overfit-ting" for all variants of the IM algorithm.

4. CONCLUSION

Based on this research, it can be concluded that the learning pattern of S1 Informatics students in the first semester of 2020/2021 in using CeLOE LMS based on the CeLOE menu which is often accessed by students as follows: Forum (330872 (29.64%) total access), Resource (277698 (24.87%) total access), Quiz (150863 (13.51%) total access), URL (139612 (12.5%) total access), Assignment (101027 (9.05%) total access), H5P (95608 (8.56%) total access), Folders (9472 (0.85%) total access), Pages (6457 (0.58%) total accesses), Chat (2441 (0.22%) total accesses), Book (880 (0.08%) total accesses), Glossary (684 (0.06%) total accesses), Feedback (647 (0.06%) total accesses), Active quiz (108 (0.01%) total accesses), and Choice (85 (0.01%) total accesses). IM can model based on event logs just fine but there is still relatively much variation of the activity beyond the modeled event log. Based on the three variants of the IM algorithm, the IM has the best fitness value (1.0) with a fairly low precision value (0.750) compared to the IMi and IMd, although, the generalization value of this algorithm is relatively high (0.984). It is hoped that this research can contribute to the addition of new perspectives related to the implementation of process mining using inductive miner algorithm in the field of education.

REFERENCES

[1] I. K. Suartama dan I. D. K. Tastra, E-learning berbasis moodle, Yogyakarta: Graha Ilmu, 2014.
[2] I. Atastina dan A. P. Kurniati, "Student registration process evaluation using process mining case study: IT Telkom," Ninth International Conference on Digital Information Management (ICDIM 2014, pp. 189-183, 2014.
[3] W. M. P. van der Aalst, "Discovery, Conformance and Enhancement of Business Processes," dalam Process Mining, Springer Berlin, Heidelberg, 2011, pp. XVI, 352.
[4] S. Leemans, D. Fahland dan W. van der Aalst, "Process and deviation exploration with Inductive visual Miner," BPM Demo Sessions 2014 (co-located with BPM 2014, Eindhoven, The Netherlands, September 20, 2014, pp. 46-50, 2014.
[5] L. Valensia, R. Andreswari dan R. Fauzi, "Implementation of Process Mining to Discover Student Learning Patterns using Fuzzy Miner Algorithm (Case Study: Learning Management System (LMS) Telkom University)," pp. 126-131, 2021.
[6] S. Chanifah, R. Andreswari dan R. Fauzi, "Analysis of Student Learning Pattern in Learning Management System (LMS) using Heuristic Mining a Process Mining Approach," pp. 121-125, 2021.
[7] G. A. A. Wisudiawan dan A. P. Kurniati, "Process Mining on Learning Activities in a Learning Management System," pp. 476-482, 2022.
[8] W. van der Aalst, "Data Science in Action," dalam Process Mining, Springer Berlin, Heidelberg, 2016.
[9] J. Carmona, B. v. Dongen dan M. Weidlich, "Relating Processes and Models," dalam Conformance Checking, Springer Cham, 2018, pp. XIV, 270.
[10] J. Buijs, B. Dongen dan W. van der Aalst, "Quality Dimensions in Process Discovery: The Importance of Fitness, Precision, Generalization, and Simplicity," International Journal of Cooperative Information Systems, vol. 23, p. 1440001, 2014.
[11] I. Nuritha dan E. R. Mahendrawathi, "Structural Similarity Measurement of Business Process Model to Compare Heuristic and Inductive Miner Algorithms Performance in Dealing with Noise," Procedia Computer Science, vol. 124, pp. 255-263, 2017.

JURIKOM (Jurnal Riset Komputer), Vol. 9 No. 4, Agustus 2022
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v9i4.4769
Hal 1128−1135
http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

[12] S. J. Leemans, D. Fahland dan W. M. P. van der Aalst, “Discovering Block-Structured Process Models from Event Logs - A Constructive Approach,” dalam Application and Theory of Petri Nets and Concurrency, Springer Berlin Heidelberg, 2013, pp. 311-329.
[13] S. J. Leemans, D. Fahland dan W. M. P. van der Aalst, “Discovering Block-Structured Process Models from Event Logs Containing Infrequent Behaviour,” vol. 171, pp. 66-78, 2014.
[14] S. J. J. Leemans, E. Poppe dan M. T. Wynn, “Directly follows-based process mining: A tool,” dalam Proceedings of the ICPM Demo Track 2019 (CEUR Workshop Proceedings, Volume 2374).Vol. 2374., Sun SITE Central Europe, 2019, pp. 9-12.
[15] I. P. A. Eka Pratama, Handbook data warehouse: teori dan praktik berbasiskan open source, Informatika Bandung, 2018.
[16] I. A. Fitriansah, R. Andreswari dan M. A. Hasibuan, “ANALISIS DAN IMPLEMENTASI PROCESS MINING PADA AKTIVITAS MAHASISWA BERDASARKAN EVENT LOG PENGGUNAAN APLIKASI BIMBINGAN TUGAS AKHIR,” 2019.
[17] J. Carmona, B. v. Dongen, A. Solti dan M. Weidlich, “Relating Processes and Models,” dalam Conformance Checking , Springer Cham, 2018, pp. XIV, 270.
[18] L. Mardhatillah, M. Er dan R. P. Kusumawardani, “IDENTIFIKASI BOTTLENECK PADA HASIL EKSTRAKSI PROSES BISNIS ERP DENGAN MEMBANDINGKAN ALGORITMA ALPHA++ DAN HEURISTICS MINER,” vol. 1, 2012.