

# Discovery and Analysis of the Teaching/Learning Processes using Process Mining Techniques

Guillermo Calderón-Ruiz<sup>1</sup>, Nicolás Cayturo-Silva<sup>2</sup>, Claudia Lazarte-Díaz<sup>3</sup> and Gonzalo Urrutia-Quequezana<sup>3</sup>

<sup>1,2,3</sup> Universidad Católica de Santa María, Urb. San José s/n Umacollo, Arequipa, Perú

## Abstract

Normally it is assumed that the improvement of the teaching-learning process lies in the insertion of new techniques, methods, or technology in the process; but the human factor is left aside, mainly the teacher, perhaps assuming that he/she is properly trained. We want to investigate the relationship between the activities carried out by the teacher and the level of learning achieved, but we want to do it automatically, therefore, as a first step, it is necessary to identify what the teacher does and how it is related to the level of learning.

In this paper, we applied Process Mining techniques to discover (model) and analyze the teaching-learning process in higher education automatically, and the results show that it is possible.

## Keywords

Discovery, Analysis, Teaching, Learning, Process mining

## 1. Introduction

Normally it is assumed that the improvement of the teaching-learning process lies in the insertion of new techniques, methods, or technology in this process; but the human factor is left aside, mainly the teacher, perhaps assuming that he/she is properly trained or that he/she clearly knows his/her activities [1] [2] [3]. We want to investigate the relationship between the activities carried out by the teacher and the level of learning achieved, but we want to do it automatically, therefore, as a first step, it is necessary to identify what the teacher does and how he/she does it to relate it to the level of learning.

The goal of our work is to automatically identify (discover) what the teacher does and also to automatically analyze his activities. To achieve this goal, we use *Process mining* techniques for both activities: discovering and analyzing. *Process mining* is a research discipline that is between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. *Process mining* aims to discover, monitor and improve real processes by extracting knowledge from event logs available in information systems [4]. In its early years, *Process mining* considered three main activities, (i) *process discovery* (i.e., extracting process models from event logs), (ii) *conformance checking* (i.e., comparison of real data, included in event logs, versus an “ideal” model; in order to find deviations), and (iii) *enhancement* (i.e., extract different types of information, e.g., performance of business processes, organizational mining, among others) [4] [5] “Modern” or online *Process mining* considers additional activities as detect, predict and recommend [6]. *Process mining* is applied in several areas [7] and also in education.

In this paper we present a description of the teaching-learning process in a Peruvian university, this process was validated by its participants. At first glance, the process is simple and linear, composed of 5 activities (i.e., Explain activities, Clarify doubts, Class development, Evaluate students and Course summary) and two participants: teacher and student. As a second step, we have created a simulated environment to automate it and generate the necessary data to be able to use the *process mining*

---

CITIE 2022: International Congress of Trends in Educational Innovation, November 08–10, 2022, Arequipa, Peru

EMAIL: gcalderon@ucsm.edu.pe (A. 1); nicolas.cayturo@ucsm.edu.pe (A. 2); 76429718@ucsm.edu.pe (A. 3); 73099453@ucsm.edu.pe (A. 4)

ORCID: 0000-0002-0981-7653 (A. 1); 0000-0003-1656-396X (A. 2); 0000-0002-9978-8605 (A. 3); 0000-0001-5365-8824 (A. 4)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

techniques. The third step has been to use the *process mining* techniques to discover the teaching-learning process and also to analyze it, all this automatically. In discovery we have used, for example, the *Alpha Miner (Alpha++)* algorithm [8] and for analysis the *Replay a Log on Petri Net for Performance/ Conformance Analysis* algorithm [9].

The paper is structured as follows. Section II presents the state of the art of *Process Mining* in education, section III presents the process detail and also the automatic simulation environment, section IV details the application of the *process mining* techniques, finally, section V lists the conclusions of the paper.

## 2. Process Mining in Education

The research we have found does not directly describe the behavior of teachers in the teaching-learning process but focuses on the field of education in general and, in certain cases, the behavior of students.

In [10] the "Subject Validation" and "Teaching and Learning" processes of the Escuela Superior Politécnica del Litoral (EPSOL) are described, modeled and redesigned. For each process, the roles involved in it and their description are identified to generate a model that describes its current state of execution (i.e., the "as-is" model); then a model corresponding to an improvement proposal is generated (i.e., the "to-be" model), based on new requirements. The redesign is oriented to generate a process model that is aligned with the objective of automating educational processes. The results of this research indicate that both processes were modeled in their current state, and then, based on discussions with the owners of the process and the authors' own experiences, a proposal for the redesign of the models ("to-be") was made, as well as the identification of linear and non-linear processes.

In [11] the *Process Mining Methodological Framework for Discovery Analysis* is used, focusing on the control flow and data perspectives. The analyzed process is the "Student Admission Process in Universities". The authors use *process mining* techniques and algorithms available in the *ProM* tool for event log preparation and automated discovery of the business process. The analysis in this paper focuses on discovery and performance. Discovery indicates that there are some instances of the process that did not conform to what was assumed. The start and completion events deviated from the expected events. While performance analysis identified potential activities that could be further reviewed. Most of them involve collaboration with external entities. Regarding assessment, more information is needed as this is an activity directly related to the student submitting incomplete applications. Understanding the root causes can help to develop solutions to solve potential problems, for example accessibility to information. The performance analysis also provided an overview of processing times. Although these times appeared to be within the normal range, the authors state that targets and strategies need to be set to account for the increased demand caused by bottlenecks in the process.

Likewise, in [12] the behavior of students is analyzed by identifying specific types of sequences, for this, data is collected from learning systems, from which fields such as the number of activity performed, number of attempt, activity group and different time stamps belonging to the evaluations are rescued, however, for the authors it is necessary to group these activities by the type of action performed, such as: reviewed, analyzed, among others; the research concludes with the finding of different patterns of behavior; despite the fact that the article emphasizes that the findings can benefit the teacher through his didactics, the authors do not analyze the behavior of the teacher in a direct way. In [13] authors work with online learning platforms; however, these leave aside any teacher intervention and focus only on student behavior and self-regulated learning of these, they use only four fields in the event log (time, identifiers, actions, and behavior) then using the *Inductive Miner* algorithm authors find a relationship between student behavior and their different grades.

The research conducted in [14] analyzes the impact of machine learning techniques in combination with *process mining* to measure their effectiveness in improving the learning experience of students in massive open online courses (MOOCs) and decreasing the dropout rate. *Process mining* algorithms were applied to prepare process data (i.e., assessment grades, solution presentation time, video lecture interaction log, participant demographic information, time, and final grades) obtained from a course on the Coursera platform in the summer of 2014; the course is "Principles of Economics". The research

results show that the techniques used in the study can predict student performance at an early stage, which could be analyzed by teachers to improve teaching methodologies and decrease student dropout.

Finally, in [15] an investigation is carried out to predict the learning behavior of students, focusing on identifying the causes of student dropout. Two algorithms in the ProM tool: *Inductive visual Miner* (IvM) and its extension, *Directly Follows visual Miner* (DFvM) were applied, making a comparison between both results, concluding that DFvM presents a more accurate automatic *process discovery* model, allowing teachers and administrators to take effective actions to motivate students to attend classes.

As noted above, these papers do not analyze teacher behavior, but we believe it is necessary to do so because it has an influence on the level of learning achieved; in this paper we will focus on visualizing behavior in order to analyze it later. We leave for a future work to analyze the influence of the teacher's behavior on the level of learning achieved.

### 3. Teaching-learning Process

According to [16], to design a teaching-learning process is a task that every education professional must perform when planning a specific training activity: course, subject, seminar or other.

In this sense, aspects such as the context in which the teaching is to be developed, the contents of the training activity itself, or the evaluation criteria to be considered to determine whether the learning objectives provided have been achieved, should be considered, as well as to identify the factors that influence student performance [11].

#### 3.1. Description of the process

The teaching-learning process, that we are using in this paper, begins with the opening of a new academic semester. On the first day of class, the teacher explains the course content and the different activities that will be developed during the semester. If the content of the course and/or activities is not clear, the students' doubts are clarified. The process continues with the development of the lecture classes, where the contents are explained, doubts are resolved, the activities to be developed are explained, assignments are reviewed, feedback is given, and conflicts in the different activities (i.e., grades, members, among others) are resolved. Once the teaching activities of the academic phase are finished, the evaluation period begins, which is accompanied by the execution of the evaluation, review, clarification of doubts and resolution of conflicts in the evaluation (i.e., grades). The teaching-learning process ends with a summary of the course.

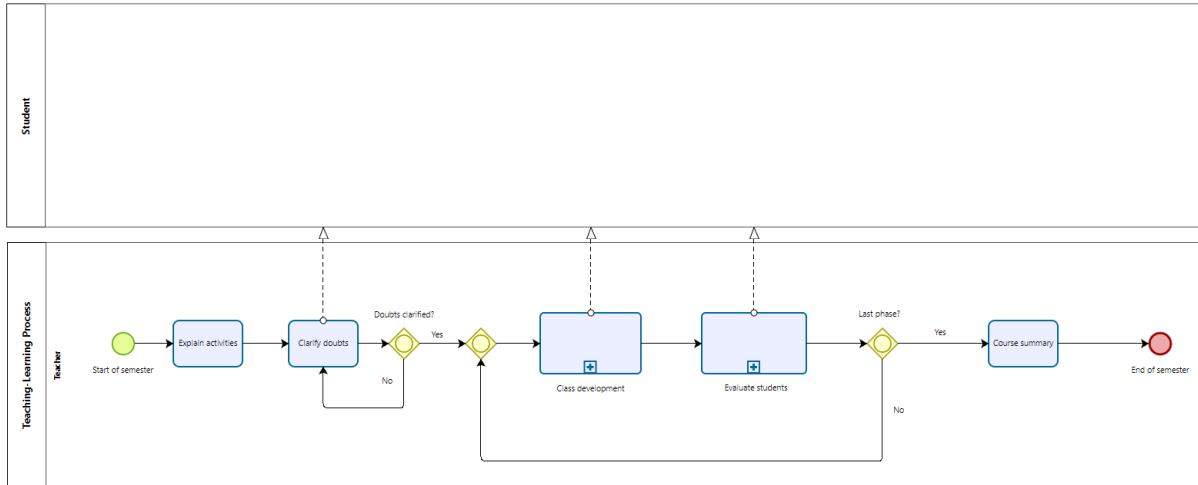
The two participants in the process: teacher and student are detailed in Table 1.

**Table 1**  
Participants in the teaching-learning process

Participant	Head 2
Teacher	The person in charge of teaching students and assessing them through contributions and evaluations.
Student	A person who is acquiring the knowledge imparted by the teacher of a subject and is evaluated according to what he/she has learned.

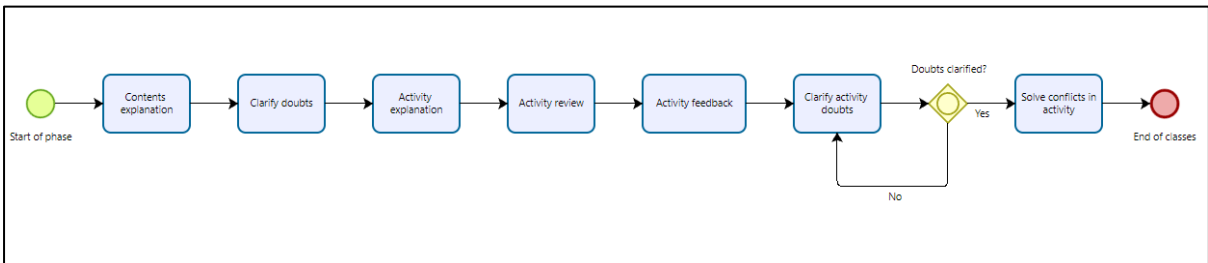
#### 3.2. Modelling of the process

Figure 1 shows the process in its current state, the model has been realized with the BPMN 2.0 standard; according to [17] BPMN is a de facto standard used for business process modeling and aids in the understanding, analysis and communication of business processes.

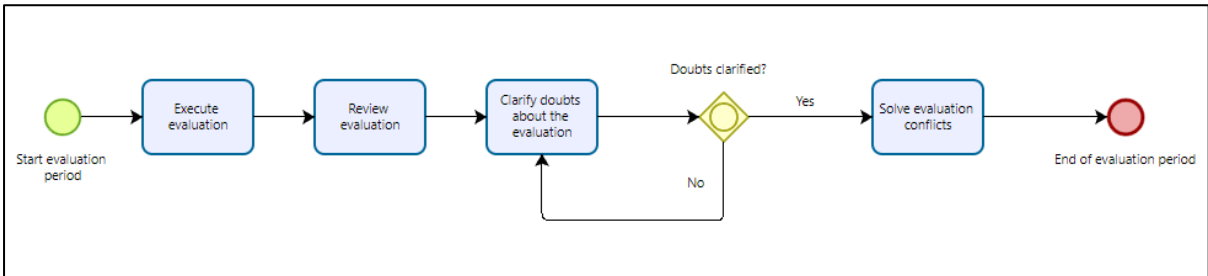


**Figure 1: BPMN model of the Teaching-Learning process**

Figures 2 and 3, respectively, show the Class development and Evaluate students subprocesses.



**Figure 2: BPMN model of the Class Development sub-process**



**Figure 3: BPMN model of the Evaluate Students sub-process**

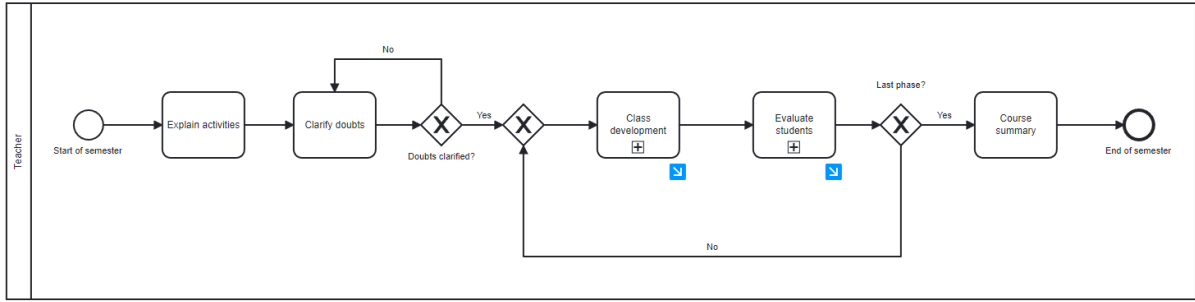
### 3.3. Process simulation

To create the simulated environment, we have used the *BIMP*<sup>2</sup> tool, a free tool to simulate business processes; this tool has been developed by a research team at the University of Tartu. The simulation is based on the configuration of resources and activities of a process based on probability distributions. *BIMP* needs as input a *.bpmn* file extension (i.e., the process must be modeled in a tool that allows to generate a file with this extension), for this purpose we have used the *bpmn.io*<sup>3</sup> tool, which is a web platform that allows to view and edit BPMN, DMN and CMMN diagrams [18].

Figure 4 shows the Teaching-Learning process modeled in *bpmn.io*, which is identical to the process shown in Figure 1.

<sup>2</sup> <https://bimp.cs.ut.ee/>

<sup>3</sup> <https://bpmn.io/>



**Figure 4:** Teaching-learning process modeled with bpmn.io

Once the model was loaded into BIMP, the main elements of the model were configured: resources (roles), activities and gateways (decisions). Figure 5 shows the configuration of the teacher role, highlighting the amount of resources used (one teacher per course) and the cost per hour (which could be used for cost analysis).

Resources +

Name	# of Resources	Cost per Hour	Timetable	Remove
Teacher	1	20.83	Default	

**Figure 5:** Configuration of the teacher role in BIMP

As an example of activity configuration, Figure 6 shows the configuration of the activity *Clarify doubts* about the evaluation, which belongs to the sub process *Evaluate students*. The data we have considered are the resource executing the activity (teacher), the probability distribution associated with the activity (fixed), the time value associated with this activity (2) and the time unit (hours). We have not considered the fixed costs or thresholds since we will not use them in the analysis.

Clarify doubts about the evaluation (Sub-Process: Evaluate students)

Resource  
Teacher

Duration  
Distribution: Fixed to 2 Time unit: Hours

Fixed cost and thresholds  
Fixed cost: \_\_\_\_\_ Cost threshold: \_\_\_\_\_ Duration threshold: \_\_\_\_\_ Time unit: \_\_\_\_\_

**Figure 6:** Configuration of the Clarify doubts about the evaluation activity in BIMP

Finally, gateway configuration was done, in all the cases was required two possible values, i.e., the probability of choosing one path or the other represented by a percentage, which is defined by the process experts.

An important reason for choosing *BIMP* to simulate the Teaching-Learning process is that it allows the generation of an event log (i.e., file that stores the executions of a process, what actually happens in

the process), which is necessary to apply *process mining* techniques. Table 2 shows an extract of the event log generated by the *BIMP* tool after running the simulation. The generated event log consists of 150 cases (complete executions) and 2388 events (a set or sequence of events makes up a case).

According to [19] cited by [11] event logs should only contain event data related to the process under analysis. An event log can be deconstructed to the following elements: Cases, which represent a process instance, therefore the event log would contain several cases. Events, every case is formed by events, these could be understood as a task in the process and every event is part of one and only one case. Event attributes, any extra information related to the process. Common attributes are activity, timestamp, cost and resources [19].

**Table 2**

A sample event log extracted from BIMP

Case ID	Event ID	Properties				
		Timestamp	Activity	Resource	Cost	...
1	1	2023-04-28T17	Start of semester	Teacher	0.00	...
	2	2023-05-01T11	Explain activities	Teacher	41.66	...
	3	2023-05-01T13	Clarify doubts	Teacher	41.66	...
	4	2023-05-01T15	Class development	Teacher	41.66	...
	5	2023-05-01T17	Evaluate Students	Teacher	41.66	...
	6	2023-05-02T11	Course summary	Teacher	41.66	...
	7	2023-05-02T11	End of semester	Teacher	0.00	...

## 4. Applying Process Mining techniques

To apply the *process mining* techniques, we have chosen the free tool *ProM*<sup>4</sup>, this tool has more than 1,000 implemented algorithms and allows us to discover and analyze all kind of processes. In this section we explain the use of *process mining* algorithms to discover, compare and analyze the Teaching-Learning process.

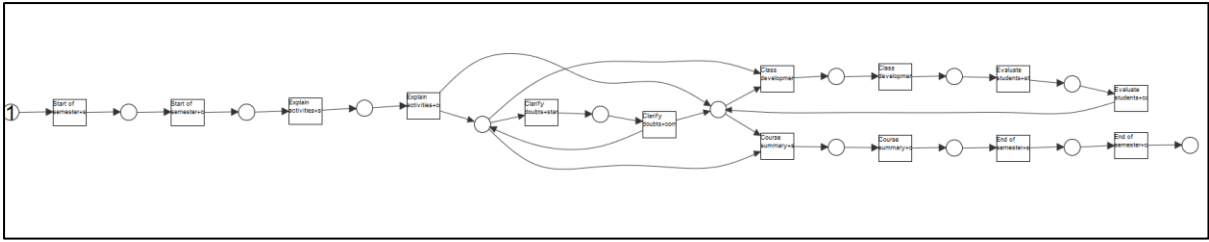
Before applying *process mining* techniques, we need an error-free event log with complete cases. To ensure that the event log is correct we must inspect it, for this we use the *Filter Log using Simple Heuristics* algorithm implemented in *ProM* [8], after this review we obtained an event log with 150 cases (complete executions) and 2388 events (a set or sequence of events makes up a case), see Figure 8 for more details.

### 4.1. Model discovery

To discover (model) the process we have used the *Alpha Miner* (Alpha++) algorithm implemented in *ProM* [8], which allows us to discover processes using as representation a Petri net (i.e., mathematical model that models processes through a flow of events and transitions). The advantage of using this notation is the simplicity of the resulting model and the ease of reading the process. Figure 7 shows the Teaching-Learning process discovered by the *Alpha++* algorithm. The *Mine Petri net with Inductive Miner* algorithm was also applied to this event log, since the Petri net generated by it will be used to perform the *conformance check* [20] in the next step. For didactic and explanatory purposes, this section makes use of the *Alpha++* algorithm due to the outputs it provides, such as a simple and easy to understand Petri net.

In figure 7 you can see that the activities appear twice (for example, Start of semesters and Start of semester+c, the *s* stands for start and the *c* stands for complete) to indicate that the activities have a beginning and an ending that could take time, this combination of activities is known as a class.

<sup>4</sup> <https://www.promtools.org/doku.php>



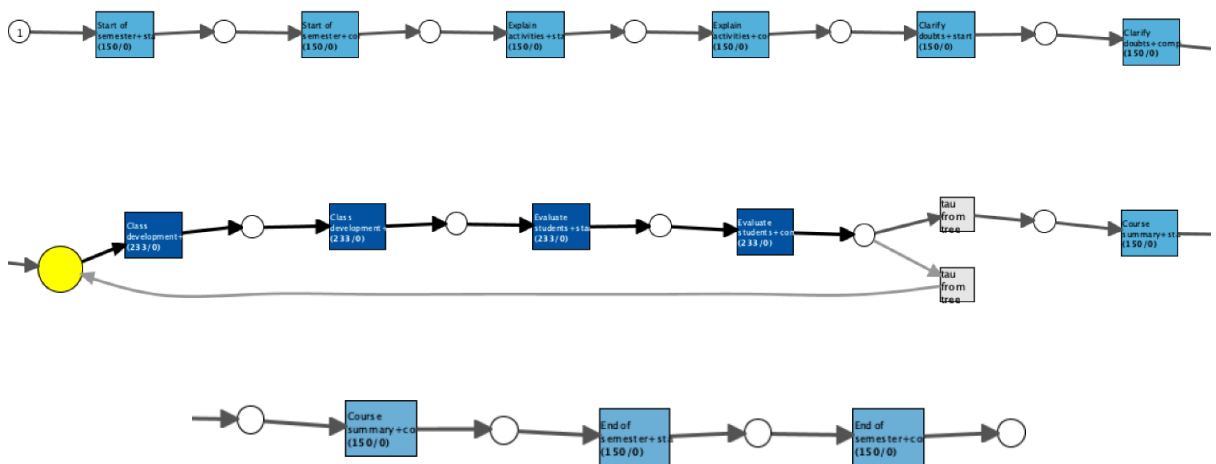
**Figure 7:** Petri net of the Teaching-Learning process, discovered by the Alpha++ algorithm

The extracted process shows the most common flow - *start of semester - explain activities - clarify doubts - class development – evaluate students - course summary - end of semester*, which agrees with the description of the process explained above and shown in figure 1. The three activities with the highest number of occurrences are *evaluate students*, *class development*, and *clarify doubts* with 211, 211 and 172 occurrences respectively. The mined process considers "Start of semester" as the main starting activity (100.0% of the cases) and "End of semester" as the main ending activity of the process (100.0% of the cases).

## 4.2. Conformance

Conformance techniques in *process mining* use two inputs: (i) an event log and (ii) a process model (Petri net) [21], the conformance result shows information about the differences between the process model and the behavior recorded through the event log. For our case, the Petri net (the process model) is obtained by applying the discovery techniques mentioned in the previous section using the original event log. Then, to have a new event log (i.e., a different behavior), the process simulation is run again in the *BIMP* tool; with these two inputs we ensure that the comparison allows finding differences.

Process conformance was carried out with the algorithm *Replay a Log on Petri Net for Conformance Analysis* of the *ProM* tool. This algorithm is based on reproducing each of the cases of the event log on the process model [22] and thus, comparing the activities of the process model (Petri net) against the activities of the new event log [20]. This conformance algorithm displays the results in graphical and numerical format.

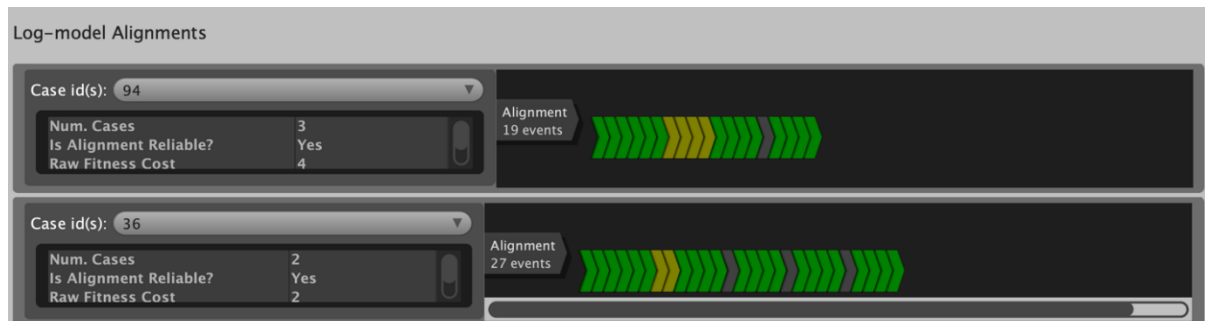


**Figure 88:** Resulting Petri net after applying the Replay plug to Log on Petri Net for Conformance Analysis

Figure 8 shows the result in graphical format, a Petri net with activities in different shades of blue; the darkest activities (e.g., evaluate students or class development) are the ones that are executed in a

similar way in both models, this means that there is no difference in those activities. On the other hand, the lighter activities (e.g., explain activities or clarify doubts) are not performed with the same frequency in both models, here there are differences [22].

The applied algorithm also allows us to see the differences at the case level. In Figure 9 we can appreciate cases 94 and 36 in these cases it can be seen that the activities painted in green are executed in the event log and in the model, on the other hand, the activities painted in yellow are only executed in the event log [23]; these differences should be thoroughly analyzed to determine if there are problems in the model or in the way of doing things.



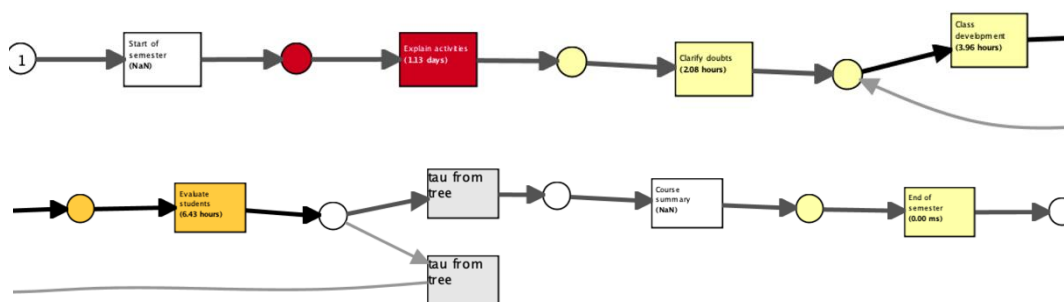
**Figure 9:** Resulting log-model alignments after applying the Replay plug to Log on Petri Net for Conformance Analysis

The results of the algorithm in numerical format show a set of metrics: *Trace Fitness*, *Move-Model Fitness* and *Move-Log Fitness*, the values of these metrics fluctuate between 0 and 1, the closer they are to 1 means that the event log fits the model. On the contrary, if the values move away from 1, it means that the event log cases differ from the model [23]. The results of our comparison show that the *Trace fitness* has a value of 0.98, which indicates that the event log and the model do not present significant differences; the activities are being executed as expected.

### 4.3. Performance

To perform a performance analysis of the process, the *Replay a Log on Petri Net for Performance/Conformance Analysis* algorithm is used, this algorithm requires the same two inputs explained in the previous section. The performance analysis returns values such as waiting time, dwell time and frequency of occurrence per activity, these values allow calculating the average duration of the process [24].

Figure 10 shows the performance results in graphical format, it is possible to appreciate bottlenecks (i.e., red circle before the *Explain activities* activity) and the frequency of execution of the activities represented by the thickness of the arrows.



**Figure 109:** Resulting after applying the Replay plug to Log on Petri Net for Performance/Conformance Analysis



The algorithm also shows global statistics, for example, minimum, maximum and average time required by the process (see Figure 11). It can be seen that the results obtained are quite similar to reality, since in a daily scenario of the teaching-learning process, the activity that demands more time is *Explain activities*, in which the teacher explains the course content and details each of the activities to be carried out during the semester. After this, doubts usually arise or questions are asked by the student about what has been explained and the teacher proceeds to resolve the concerns.



**Figure 1110:** Resulting statistics after applying the Replay plug to Log on Petri Net for Performance/Conformance Analysis

## 5. Conclusions

*Process mining* has allowed us to demonstrate that we can automatically identify the activities carried out by a teacher, as well as analyze these activities automatically; thus, we have achieved the objective set for our work. The discovery algorithms have allowed us to graphically identify the activities carried out by a teacher and the order in which they are executed. The compliance and performance algorithms have allowed us to analyze the process by identifying differences, bottlenecks, and execution times.

With the above, we are able to carry out our second objective, to investigate the relationship between the activities performed by the teacher and the level of learning achieved by the student, but we will do this in future work.

## 6. References

- [1] T. E. Webster & Paquette J., «“My other hand”: The central role of smartphones and SNSs in Korean students’ lives and studies,» *Computers in Human Behavior*, 2023.
- [2] P. C. Herrera, M. Hurtado & P. Arteaga-Juárez, «Visual Programming for Teaching Geometry in Architectural Education,» de *Lecture Notes on Data Engineering and Communications Technologies*, Lima, 2023, pp. 958-969.
- [3] J. Zhao & M. Wang, «The Internet of Things Computer Aided Technology Oriented by the English Teaching System,» de *Computer-Aided Design and Applications*, Qinhuangdao, 2023.
- [4] W. e. a. Van der Aalst, «Process Mining Manifesto,» de *International Conference on Business Process Management*, 2012.
- [5] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer Publishing Company, Incorporated, 2011.
- [6] W. v. d. Aalst, *Process Mining: Data Science in Action*, Springer Berlin, Heidelberg, 2016.
- [7] G. Calderón-Ruiz & D. Fernández, «Process Mining: The first successful Peruvian case.,» de *Proceedings of the LACCEI International Multi-conference for Engineering, Education and Technology*, 2022.

- [8] J. De Weerd, M. De Backer, J. Vanthienen & B. Baesens, «A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs,» *Inf. Syst.*, p. 654–676, 2012.
- [9] R. García, J. Santos & J. Armas, «Control Metrics Evaluation Model for Business Processes using Process Mining,» de *The Tenth International Conference on Information, Process and Knowledge Management*, 2018.
- [10] C. Ortega-Ventura & H. Pilco-Naula, «Descripción, Modelamiento y Rediseño del Proceso de Convalidación de Materias - Proceso de Enseñanza y Aprendizaje utilizando el lenguaje de modelamiento BPMN,» *ESPOL*, pp. 1-10, 2015.
- [11] J. Gonzalez-Dominguez & P. Busch, «Automated Business Process Discovery and Analysis for the International Higher Education Industry,» *Knowledge Management and Acquisition for Intelligent Systems*, p. 170–183, 2018.
- [12] L. Juhaňák, J. Zounek & L. Rohlíková, «Using process mining to analyze students' quiz-taking behavior patterns in a learning management system,» *Computers in Human Behavior*, p. 496–506, 2019.
- [13] R. Cerezo, A. Bogarín, M. Esteban & C. Romero, «Process mining for self-regulated learning assessment in e-learning,» *Journal of Computing in Higher Education*, pp. 74-88, 2019.
- [14] H. Alqaheri & M. Panda, «An Education Process Mining Framework: Unveiling Meaningful Information for Understanding Students' Learning Behavior and Improving Teaching Quality,» *Information (Switzerland)*, 2022.
- [15] R. Umer, T. Susnjak, A. Mathrani & S. Suriadi, «On predicting academic performance with process mining in learning analytics,» *Journal of Research in Innovative Teaching & Learning*, p. 160–176, 2017.
- [16] J. Hilera & D. Palomar, «Modelado de procesos de enseñanza-aprendizaje reutilizables con XML, UML e IMS-LD,» *RED. Revista de Educación a Distancia*, pp. 1-11, 2020.
- [17] I. Maslov, «Towards Empirically Validated Process Modelling Education Using a BPMN Formalism,» de *Lecture Notes in Business Information Processing*, Barcelona, 2022.
- [18] G. Aagesen & J. Krogstie, «Analysis and Design of Business Processes Using BPMN,» de *Handbook on Business Process Management 1*, 2010, pp. 213 -235.
- [19] W. van der Aalst, «Getting the Data,» de *Process Mining*, Berlin, Heidelberg, 2011, p. 95–123.
- [20] T. Barboza, F. Santoro, K. Cerqueira, & R. Costa, «A Case Study of Process Mining in Auditing,» de *the XV Brazilian Symposium*, 2019.
- [21] J. Carmona, «Decomposed Process Discovery and Conformance Checking,» *Encyclopedia of Big Data Technologies*, 2018.
- [22] F. M. Santoro, K. C. Revoredo, R. M. Costa, and T. M. Barboza, «Process Mining Techniques in Internal Auditing: A Stepwise Case Study» *iSys: Revista Brasileira de Sistemas de Informação (Brazilian Journal of Information Systems)*, vol. 13, nº 4, pp. 48-76, 2020.
- [23] R. Ghawi, «Process Discovery using Inductive Miner and Decomposition,» American University of Beirut, Beirut, 2016.
- [24] R. A. García, J. J Santos, J. A. Armas, «Control Metrics Evaluation Model for Business Processes using Process Mining,» de *The Tenth International Conference on Information, Process and Knowledge Management*, 2018.