Approaches for analyzing students scoring patterns using Process Mining

Srungarapu Sai Sri Nandan
Dept. of Computer Science
Dayananda Sagar College of Engineering
Bangalore, India
sssnandan44@gmail.com

Vonkayala Ashwini
Dept. of Computer Science
Dayananda Sagar College of Engineering
Bangalore, India
ashwinivonkayala1721@gmail.com

Vishwakumar K Hirehalli
Dept. of Computer Science
Dayananda Sagar College of Engineering
Bangalore, India
vishwahirestar@gmail.com

Pallapothu Lakshmi Sharanya
Dept. of Computer Science
Dayananda Sagar College of Engineering
Bangalore, India
pallapothusharanya7@gmail.com

Dr. Deepak G
Associate Professor
Dept. of Computer Science & Engineering
Dayananda Sagar College of Engineering
Bangalore, India
deepak.dsce@gmail.com

Dr. Harish Kumar N
Assistant Professor
Dept. of Computer Science & Engineering
Dayananda Sagar College of Engineering
Bangalore, India
nhk015@gmail.com

Abstract

Process mining is an innovative technique for analyzing business processes. It uses software agents to monitor and analyze the underlying process of the process for a given time, collecting metrics on business process performance, key metrics, transaction-level information, and event-level information. Process mining enables an organization to automate tedious and time-consuming tasks, identify systematic issues, and identify opportunities for improvement in its processes. This paper is about how a student's scoring patterns can be identified by using process mining. The researchers have done in-depth analyses of different algorithms that are used to calculate test scores for students. They found out that most students score high on tests when they use easy algorithms, while they score low when they need complex algorithms to complete the task. This means that students perform better at tasks that require simple calculations than those that require complex calculations because these complex calculations are not as easy as they seem.

Keywords: Process mining, Ict, lms, prom, xes, data mining, learning patterns, event logs, student's performance, process model

Introduction

Process mining is a method for uncovering information about processes and creating process models, such as Petri nets and data flow diagrams, using data mining techniques. A recent research approach in data mining that focuses on understanding educational processes rather than just generating predictions from large amounts of educational data is Educational Process Mining (EPM).

The goal is to create a clear process model from event logs and combine traditional process analysis methods based on simulated models with data-oriented analysis techniques. Additionally, EPM uses data from event logs recorded by different information and communication technology (ICT) tools, few of those are MOOCs and online learning management systems, rather than identifying local patterns.

Educational data mining has been a subject of study for many researchers, but few have considered the full scope of the process and its various possibilities when analyzing student learning. To address this gap, educational process mining (EPM) has been developed as a way to use process mining techniques to gain insights into online learning by analyzing event logs, activity data. It involves recording behavior of students in their learning online, recording events in an organized manner that includes a process instance, activity, timestamp, and originator for initiating activities.

Approaches on analyzing Scoring pattern

The methodology for the application depends on the specific data sources and process mining tools available. However, some general steps intent to use are:

Data collection: Initialy in the process mining process, we collect data from the organization's digital systems that are relevant to student performance. This could include data on student grades, assessments, attendance, and other factors that may impact student performance.

Data preparation: After the data has been gathered, it may require cleaning and transforming before it can be utilized for process mining. This could involve correcting inaccuracies, completing missing information, and organizing the data in a format that works with the chosen process mining tool.

Process discovery: After the data has been prepared, we discover and visualize the underlying processes that take place within the organization usingf the process mining algorithms. This may involve using techniques such as event logs, process mining algorithms, and process models to uncover patterns in the data.

Analysis and interpretation: Once the process has been discovered, the researchers can analyze and understand in order to discern patterns and tendencies in the outcomes of student performance. This may involve comparing the performance of different groups of students, identifying bottlenecks in the learning process, and identifying areas of strength and weakness for individual students.

Recommendations and action planning: Based on the findings of the analysis, can make recommendations for improving student performance and optimizing the learning process. This may involve suggesting changes to teaching strategies or interventions, or identifying opportunities to streamline the learning process.

Traditional EDM[11] techniques cannot be used to implement any of the three approaches mentioned above. As a result, new tools and methods which use event logs to extract process- related data.

An LMS is a tool that creates and stores records of events that take place within an educational system. These events may include activity logs in virtual classrooms, learner registration processes, access to educational resources, and the use of online courses and tutoring systems. Process mining techniques can leverage these event logs by linking them to process models, offering valuable insights about process efficiency, potential issues, and error prediction.

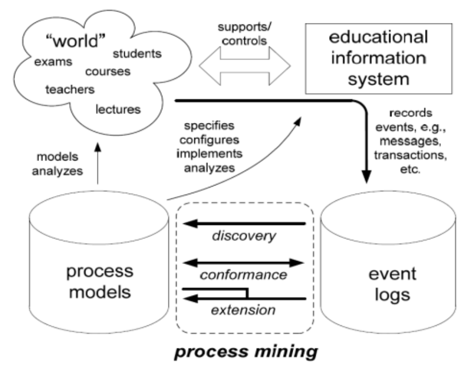


Fig 1: Generic Architecture

By examining the educational setting and its relationship with the social context, it is possible to understand the learning practices employed. This understanding can then be used to evaluate the strategies implemented for a compelling and effective learning experience using Information and Communication Technology (ICT), while considering the unique qualities of the students. This fosters a more innovative and imaginative mindset in students, promoting problem-solving and knowledge acquisition through creative thinking rather than mere rote memorization.

An XES (extensible event stream) event log is created from the information gathered on a student's behavior. This event log can then be utilized by process mining tools and techniques to: uncover process models, display process models, verify process models, expand process models, improve process models.

A Petri net, a causal net, a business process tree, a business process model and notation (BPMN), or an activity diagram in the Unified Modeling Language (UML) can all be used to represent the final process model. The instructor can review the results to pinpoint any hindrances or discrepancies in the student's learning management process. This information can then be used to continuously enhance the educational system.

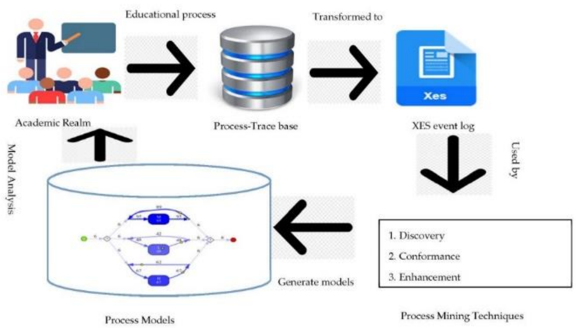


Fig 2: extensible event stream

Scenario:

The combination of technologies in education has been a key development in recent years, particularly with the rise of web-based education. It offers flexibility in location and hardware, and can be personalized using adaptive and intelligent systems that learn your preferences, objectives and expertise. Knowledge discovery in databases (KDD) and data mining [1] can be used to improve and evaluate e-learning systems by extracting useful patterns from large data collections.

Educators need to find alternative ways, such as analyzing data from web servers, to get feedback on students' learning experiences in electronic learning environments. Data mining techniques are becoming increasingly popular in web-based learning environments for analyzing learner interaction data. Although data mining has proven successful in e-commerce, there are certain particular difficulties with its use in e-learning due to variations in the domain, data, aims, and methodology..

Solution:

[2] Assessment through online means has become a vital aspect of modern education, being used both in digital-learning and blended learning formats. It serves as a tool for both self-evaluation and official exams, and can often enhance or even replace traditional evaluation methods. Intelligent data analysis may provide a greater knowledge of student performance, exam quality, and particular questions. Despite its benefits, there are ongoing challenges in terms of organizing and creating different assessment procedures. For example, in MCQ testing, the method of navigation between questions and whether the order is fixed or not can have a significant impact. Recently, researchers have been exploring the navigation of students in digital-learning systems. The goal is to identify individual navigational styles, reduce cognitive load, improve the usability and learning efficiency of digital-learning systems and support personalization of navigation.

Conventional data mining techniques are applied to various types of educational data, such as test data, to find correlations, subgroups, and emerging patterns. But these techniques lack process perspective and say little about the overall assessment process. The authors of this paper aim to familiarize the EDM community with process mining and its tool, ProM. They showcase the application of some ProM plug-ins to analyze assessment data from two separate studies. In the first study, students had to answer questions in a strict sequence and could request immediate feedback. In the second study, students were allowed to answer questions in a flexible order, revisit previous answers and make revisions.

[3] Information and Communication Technology (ICT) and latest digital media have transformed education and teaching, but all information in this system has been very static and not designed for specific learners.

Personal Learning Environments (PLEs) are the solution to this problem of not having personalised information for every student’s learning style. The internet is a huge place with a lot of content so it is difficult for students to get resources that suit the students’ preferences.

APELS is an attempt to address these problems by analysing a student's learning style, adapting to it and providing resources to assist them in their learning process.

Process mining:

[4] Learning environments that are technology-enabled are able to record learner actions, activities, and events at various degrees of detail, from keystrokes at the low level to learner activities at the high level.

These systems monitor and record a wide range of temporal data, including traffic, chat logs, modification histories, motion tracking, logs of the use of learning resources, and different interaction logs.

Process mining makes use of data, often referred to as logs, audit trails, log files, or traces, for uncovering, monitoring, and improving educational processes. Process mining serves as a connection between data mining and process modelling and analysis. Process mining, a discipline within data mining, applies a process-oriented viewpoint to data mining as a whole.

The conventional methods of Educational Data Mining (EDM) concentrate on discovering correlations or basic patterns in educational data. However, they do not provide a visual representation of the entire learning process. Data mining has been successful in uncovering useful patterns in data collected in educational settings.

Educational Process Mining (EPM) specifically applies PM to raw educational data. EPM is process-oriented and makes previously unknown or partially known processes clear by utilizing event data. Process Mining places emphasis on the full process rather than isolated patterns, recognizing the importance of comprehensive process models and overlapping activities in PM. Traditional EDM procedures do not include other suitable methods like process discovery, compliance testing, and bottleneck analysis.

Data collection:

[7] learning management systems (LMS) captures every activity happened in the platform, and methods to utilize the data to accurately monitor, and analyse to develop student online learning trends. Inductive and fuzzy mining techniques are employed in this framework to produce a process model which depicts learning patterns using a virtual model.

The goal is to give higher institutions with the ability to extract insights from their data to improve their educational processes. The resulting improved model can be used to create recommendations and enhance student-instructor interaction with educational IT systems. The findings of the study have implications for research in open and distance learning.

[8] Data mining is a topic that seeks to improve ways for analysing certain sorts of data from the educational setting. The article cites a hypothetical lesson based on the Moodle platform and investigates how data mining is implemented in learning management systems.

The whole mining procedure for e-learning data is covered by the writers, along with tips on how to apply different data mining techniques. When utilising course management systems, instructors can use data mining to learn more about online students. This can be done by applying visualisation techniques to acquire a broad overview of student usage data or by using clustering or other techniques to build a classifier. Currently, the authors are working on a Moodle data mining tool specifically designed for online instructors to use within the Moodle platform.

[9] analysis of 3458 learners' interaction sequences in three Massive Open Online Courses (MOOCs) was carried out using process mining. By comparing these patterns to theoretical self-regulated learning (SRL) strategies, we distinguished three groups of learners. Learners who demonstrated more inconsistent and less goal-directed behaviour reported lower SRL and performed poorly in comparison to the other two groups.

The goal of this study was to examine the connection between behaviour in a digital learning environment and theoretical learning processes, with a focus on SRL strategies in MOOCs. Both self-reported data and process mining of behavioural learner data were utilized. For instance, one interaction sequence consisted of a learner finishing several video lectures and then taking an assessment, which could be related to the Reviewing Records SRL strategy. Another pattern showed a learner attempting to take an assessment after finishing a video lecture, which could be tied to the Task Exploration SRL strategy.

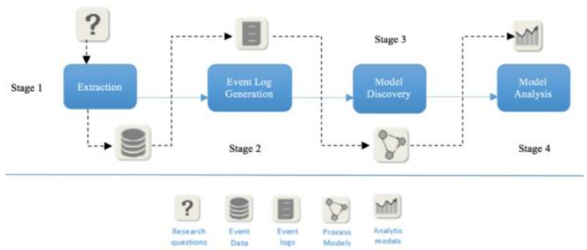
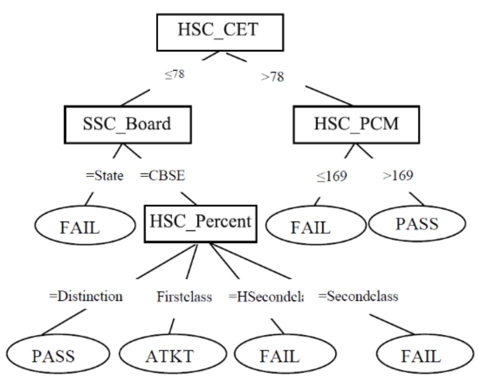


Fig 3: Task Exploration SRL strategy

Process Discovery:

[11] Decision tree is an algorithm used to classify according to continuous or discrete attributes. It is represented visually in a tree like structure with internal containing the tests for the data to evaluate, outcomes of the test are represented by the arcs and leaf nodes representing the labels.

Using various attributes related to the students we can use decision trees to analyse and predict the performance of students. Prior academics, history, previous boards, and other attributes can be utilised to forecast student success and assist students at risk in improving their learning and results. We can also check for the course of actions or attributes of good scoring students for analysis or improvement.



Decision tree for three class predictions.

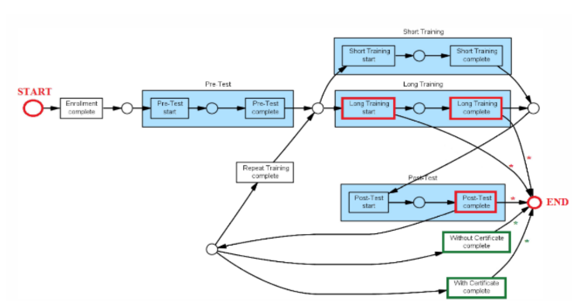
Fig 4: Decision tree

The rules generated from this tree are

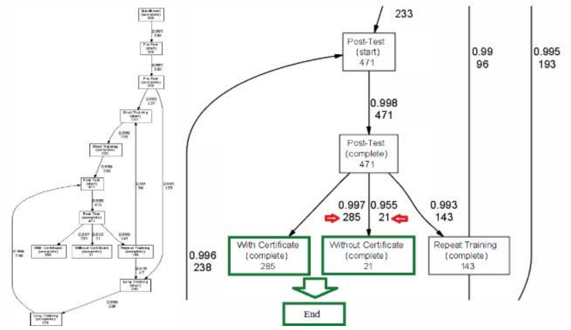
1. If HSC_CET <=78 and SSC_Board = State then FE_result = FAIL
2. If HSC_CET <=78 and SSC_Board =CBSE and HSC_Percent = Distinction then FE_result = PASS
3. If HSC_CET <=78 and SSC_Board =CBSE and HSC_Percent = FirstClass then FE_result = ATKT
4. If HSC_CET <=78 and SSC_Board =CBSE and HSC_Percent = SecondClass then FE_result = FAIL
5. If HSC_CET <=78 and SSC_Board =CBSE and HSC_Percent = Distinction then FE_result = FAIL
6. If HSC_CET > 78 and HSC_PCM <=169 then FE_result = FAIL
7. If HSC_CET > 78 and HSC_PCM > 169 then FE_result = PASS

[12] A common process mining tool for determining causality from a collection of event sequences is the a-algorithm. Initial presentations of it were made in terms of Petri Nets by a professor at Technische Universiteit Eindhoven. In this study, the a-algorithm, a basic process mining technique, was employed to identify routing patterns in an event log obtained from a private institution in Bangkok, Thailand. The investigation focused in "started" and "finished" stages of process instances, neglecting "in-progress" actions and activities. The final Alpha model, presented as a Petri net, clearly illustrates the control-flow of the training activities for the course as well as their interdependence. However, the model was discovered to conclude with 5 activities/tasks rather than the predicted 2, suggesting 3 undesirable activities/tasks in the training process. The frequently used Alpha algorithm in process mining is not a reliable method for processing logs containing concurrent actions and loop data. In order to create less susceptible to noise models and incompleteness in the logs, the Heuristic Miner method was employed in this work as a process discovery tool. To assess the relative significance of behaviour and the connections between occurrences, the Heuristic Miner algorithm took into account two essential metrics, Significance and Correlation. The resultant Heuristic model showed an atypical result, demonstrating that fewer candidates with and without credentials applied overall than there were applicants for training course in academic writing.

A representation of the Alpha model/graph resulting from the event log collected during an Academic Writing (English) training session at a private



institution in Bangkok, Thailand.



Two representations of the Heuristic miner model/graph resulting from the event log collected during an Academic Writing (English) training session at a private institution in Bangkok, Thailand.

[14] Directly Follows-Based Process Mining considers the “directly follows: relationship, that is if event B follows event A in a process, then event B is likely a direct result of event A. Process models are developed using this relationship from the event logs and this algorithm is generally used alongside other algorithms for complete understanding and improvement of the processes. The capacity to record the relationships between events is the key advantage of utilising this technique. The authors then go into a case study in which they use DFBPM to analyse an actual event log that was obtained from a manufacturing firm. The study's objectives were to identify the company's production process and evaluate the effectiveness of DFBPM in contrast to other process mining approaches like the alpha algorithm and heuristic miner. According to the study's findings, DFBPM was able to detect the company's production process and determine the relationships between operations. The study also found that the DFBPM performed better than the other methods in terms of F1 score, accuracy, and recall.

Directly Follows Models (DFM): An activity, start, or end label is placed at each node of a directed graph known as a directly follows model.

Syntax: Let Σ be an alphabet. While $start \notin \Sigma\, and\ end \notin \Sigma$. Let a directed graph be a directly follows model (M, F). Where M is a set of nodes, $N : \Sigma \cup \{start, end\}$ and F is a set of edges, $F: M \times M$.

Semantics: The language of a DFM- A , $\mathcal{L}(A)$ is $\{\langle x_{1s}, x_{1c}, \dots \dots x_{ns}, x_{nc}\rangle |\ x_1 \dots\ x_n \in M \wedge (start, x_1) \in F \wedge \forall\, 1 \le j < n(x_j, x_{i+1}) \in F \wedge (x_n, end) \cup F\} \cup \{\epsilon\ |(start, end) \in F\}$, where $_s$-events are not mandatory.

Soundness: Let (M, F) be a DFM. If path from $start$ to end includes every node $\in M$ the sound is

$$\forall_{y \in M} \exists_{x_1 \dots x_n \in M} x_1 = start \wedge x_n = end \wedge \exists a_k = y \wedge \forall_{1 \le j < n} (x_j, x_{j+1}) \in F$$

Analysis:

[16] To manage their organisational processes, several firms have built process-aware IT systems (PAIS). These procedures often document incidents associated with the implementation of actual organizational processes, but clear process models that outline the proper execution of business processes are also commonly accessible.

This study proposes a step-by-step method for assessing a process model's and an event log's compatibility, which is crucial in various scenarios. The approach introduces four metrics for evaluating conformance, including two for structural suitability (a_S and a'_S) and two for behavioural suitability. These metrics allow for the quantification of compatibility and the introduction of two new metrics, a0S and a0B, which can assess one aspect of suitability separately from the other. These metrics indicate when the optimal solution has been achieved and is ready for analysis. The presence of appropriate log data is essential for systems that facilitates business processes.

Metric 1 (Log coverage): let logs entries be B, activities be A, and tags be T, let $t_B \in$ B $\rightarrow$ T, $t_A \in$ A $\nrightarrow$ T, A_V= dom(t_A), T_A={ t_A (a)|a $\in A_V$}, and T_B= { t_B (b)|b $\in$ B}. The following metrics are for log coverage c_B and c_{TB}:

$$c_B = \frac{|\{b \in B | t_B(b) \in T_A\}|}{|B|} \qquad \textbf{(1)}$$

$$c_{TB} = \frac{|T_B \cap T_A|}{|T_B|} \qquad \textbf{(2)}$$

Metric 2 (Model coverage): let log entries be B, activities be A, and Tags be T, let $t_B \in$ B $\rightarrow$ T, $t_A \in$ A $\nrightarrow$ T, A_V= dom(t_A), T_A={ t_A (a)|a $\in A_V$}, and T_B= { t_B

(b)|b ∈ B}. The following metrics are for model coverage c_A and c_{TA}:

$$c_A = \frac{\left|\{a \in A_V | t_A(a) \in T_B\}\right|}{|A_V|} \ldots \quad \textbf{(3)}$$

$$c_{TA} = \frac{|T_A \cap T_B|}{|T_A|} \quad \textbf{(4)}$$

Metric 3 (Fitness): in the aggregated log, let number of distinct traces be k. During log replay of each log trace j ($1 \leq j \leq$k), process instances combined into the current trace be n_j, missing tokens be m_j ,remaining tokens be r_j ,consumed tokens be c_j, and produced tokens be p_j. The following is the definition of the token-based fitness metric f:

$$f = \frac{1}{2}\left(1 - \frac{\sum_{j=1}^{k} n_j m_j}{\sum_{j=1}^{k} n_j c_j}\right) + \frac{1}{2}\left(1 - \frac{\sum_{j=1}^{k} n_j r_j}{\sum_{j=1}^{k} n_j p_j}\right) \textbf{(5)}$$

Metric 4 (Simple behavioural appropriateness): in the aggregated log, let number of distinct traces be k. During log replay of each log trace j ($1 \leq j \leq$k), process instances integrated into the current trace be n_j, and average number of enabled transitions throughout the current trace's log replay be x_j (note that invisible tasks may enable succeeding labelled tasks but they are not counted themselves). Furthermore, in the Petri net paradigm, set of visible activities be A_V. The definition to the basic behavioural appropriateness metric a_B :

$$a_B = \frac{\sum_{i=1}^{k} n_i(|A_V| - x_i}{(|A_V| - 1)\sum_{i=1}^{k} n_i} \quad \textbf{(6)}$$

Metric 5 (Advanced behavioural appropriateness):in a process model, S_F relation is represented by S_F^m and S_p relation is represented by S_p^m, let S_F relation represented by S_F^t and S_p relation represented by S_p^a from event logs. d_B is an advanced behavioural appropriateness metric defined as follows:

$$a'_B = \left(\frac{S_F^t \cap S_F^m}{2.|S_F^m|}\right) + \left(\frac{|S_F^m \cap S_P^m|}{2.|S_P^m|}\right) \textbf{(7)}$$

Metric 6 (Simple structural appropriateness): let Petri net nodes set (i.e., locations and transitions) be M, and the mappings formed between tasks and events from logs be the collection of tags T. The following is the definition of the basic structural appropriateness metric a_S:

$$a_S = \frac{|T| + 2}{|M|} \quad \textbf{(8)}$$

Metric 7 (Advanced structural appropriateness): Let the Petri net model's collection of exchanges be E, set of alternative duplicate tasks be E_{DA}, and set of redundant invisible jobs be E_{IR}. a'_S is an advanced structural appropriateness metric defined as follows:

$$a'_S = \frac{|E| - (|E_{DA}| + |E_{IR}|)}{|E|} \quad \textbf{(9)}$$

[17] Current process discovery algorithms often prioritise just two of the four primary quality dimensions: replay fitness, accuracy, generalisation, and simplicity. The ETM algorithm enables you control over the discovery process based on these four choices. All dimensions are demonstrated to be relevant for process discovery, but only if the replay fitness is sufficient.

With relation to any of the four dimensions, the ETM algorithm, a genetic algorithm, optimises the process discovery outcome. This approach uses process trees to guarantee that the final model adequately characterises the observed log and also meets the criteria for being the best in terms of a weighted average of following metrics:

-Replay Fitness metric evaluates how much of the traces can be recreated from the recorded logs. To calculate the process fitness, a tree-based fitness calculation using alignment is employed. Essentially, as many events from the trace as feasible are aligned in this approach, leading to an alignment. Any mandatory events that are skipped or activities that are added without a corresponding entry in the log are subject to consequences. The total replay fitness score is obtained through the computation of the overall score.

$$Q_{rf} = 1 - \frac{\text{cost for aligning model and event log}}{\substack{\textit{Minimal cost to align arbitrary event } \text{log} \\ \textit{on model and vice versa}}} \quad \textbf{(10)}$$

-Simplicity is a metric for determining the model's complexity. It is a comparison of event log activity to tree size. Because the size of the tree greatly increases complexity and introduces flaws into process models, Furthermore, because we are employing binary trees, the number of operational nodes is impacted. Thus, if the actions are only shown once on the tree, it is deemed simple. The total simplicity score is calculated by multiplying:

$$Q_s = 1 - \frac{\#\text{duplicate activities} + \#\text{missing activities}}{\#\text{nodes in process tree} + \#\text{event classes in event log}} \quad \textbf{(11)}$$

- When replaying the log, precision compares the tree execution state space. This measurement is based on escape edges, or choices that are conceivable in the model but never made in the log. If no edges can escape, the accuracy is perfect. Information from this replay fitness is utilised to determine the portion of the state space that is used, and we discard events that are in the log but do not fit the alignment of an activity with an event. In summary, we compute precision as follows:

$$Q_p = 1 - \frac{\sum_{\substack{visited \\ markings}} \#visits * \frac{\#outgoing\ edges - \#used\ edges}{\#outgoing\ edges}}{\#total\ marking\ visits\ over\ all\ markings} \quad \textbf{(12)}$$

- Generalization calculates how many times all tree nodes must be read in order to create the provided log. Because of this, we make advantage of the alignment that the replay fitness calculation produced. The frequency with which a node is visited offers higher confidence in its correctness or incorrectness. In contrast, the generalisation is considered poor if some tree parts are only sometimes visited. As a result, generalisation is computed as follows:

$$Q_g = 1 - \frac{\sum_{nodes}\left(\sqrt{\#executions}\right)^{-1}}{\#nodes\ in\ tree} \quad \textbf{(13)}$$

The four mentioned metrics are rated on a scale of 0 to 1, with 1 being the best outcome. Replay fitness, simplicity, and accuracy can all acquire a score of one, which is regarded optimum. Generalization, on the other hand, can only approach 1 in the limit, which means that the more nodes visited, the closer the score goes to 1. By using a genetic algorithm to maximise a weighted total of the four metrics, the problem of creating a process model that balances these metrics may be effectively overcome.

[18] Curricular analytics is a branch of learning analytics that investigates the connection between curricular elements and academic performance. This can be valuable for higher education looking to assess the strengths and weaknesses of their curriculum. BPPM, which utilizes the backpack metaphor to analyse curricular trajectory, can provide insights into how failed courses are handled by students, courses that students have to retake and can identify patterns of behaviour among similar students in similar contexts.

This data can be used to develop and give timely suggestions to support student success. BPPM can also assist counselling services in identifying at-risk students and serve as a complementary tool in this process.

<u>Execution:</u>

[19] The data plays a very significant part in process mining, as even a small change or incorrect information can change the whole flow of a process that is being captured by the algorithm. The best way to make sure that data quality is being preserved by utilizing L* life-cycle model which includes a stage known as data quality assessment.

In this L* life cycle model we include an additional stage of database reconstruction to assess the data quality. This step comes after the design and justify stages, before moving on to extract, control flow model construction, integrated process model creation, data quality evaluation, and operational support.

[20] Process mining is collecting data from event logs, which is especially useful when events are recorded but no system dictates how employees should complete their duties. This article is concerned with mining non-free choice constructions or circumstances in which synchronization and choice coexist.

This study describes a method that can recognise specific kinds of implicit dependencies, such non-free-choice Petri nets. Process mining is an innovative and efficient way to extract important information from event logs. Additionally discussed is the ProM framework, an open-source toolkit for creating different process mining algorithms.

[21] Process mining's main functions are: take data from event logs, identify processes and build process models. So, all the data of processes is being collected from event logs. So, the files have to be in a proper/compatible format for the algorithm to extract data. The unique field in event logs used

during process mining is the CaseID, which will help us track an event throughout the system. Other fields help us understand about the flow of the events this particular case is going through.

The different events that occur also have to be formatted properly before we enter the logs into the algorithm for proper capturing and building of the `process models. For example, in a quiz assessment analysis the attributed of the student/case can be quiz_viewed, quiz_try_viewed, quiz_try_reviewed, quiz_try_started, try_summary_viewed, quiz_try_submitted, etc. These activities help us understand the life cycle of the quiz that has been taken by the student.

[22] Process mining helps us discover the trends or processes that have been unidentified or ignored due to the low quantity of cases in them, and it also helps us understand the effect of these processes on the overall performance. Process mining also helps us realise the deviation that have been taken in the real-time by comparing the designed process model with the real-time data and help us reduce these inefficiencies.

Ever process needs a standard flow/ process model that can be used to compare the resulting activities to better achieve the goal of the process. For example, in a course analysis, we need to establish a set of guild lines for students to follow or for the system to analyse the student's performance or predict the student's performance and take appropriate actions.

Tools:

[23] Both academic and business sectors have created a number of methods for extracting information from event logs. These tools read and store log files in various formats and deliver their findings in various ways. The integration of some of these principles into other tools can be challenging, yet they are present in some of these technologies.

ProM framework, which supports numerous languages including Petri nets, EPCs, formats, and social networks, was developed for process mining as a "pluggable" environment. The plug-ins of ProM framework may be utilized in various ways to be integrated and used in scenarios of real-world. Adding a new plug-in to the ProM framework is straightforward and does not require any changes to the code.

[24] Business Process Management has benefited greatly from Process Mining (BPM) by transforming the way processes are modeled, analyzed, redesigned, implemented, and monitored. A major limitation of conventional process modeling techniques is the time-intensive and expensive nature of conducting interviews and workshops, which can lead to a conceptual model that fails to accurately reflect the actual process being executed (As-Is process).

To solve this problem, Celonis created a collection of toolsets called Process Repository, which provides capabilities for process documentation and completely integrates them with Process Mining applications, bringing BPM and Process Mining closer together. This allows for a more efficient and effective approach to process discovery and management.

[25] ProM offers a wide range of filtering options in comparison to other process mining tools. It can perform common operations such as process discovery and present the discovered model in different formats based on user preference, such as BPMN diagrams, Petri nets, and fuzzy models. Disco also has a transparent filtering mechanism, but its output is limited to fuzzy mining models. ProM supports all core operations, while Disco and Celonis only cover some. Disco is considered user-friendly and efficient for processing event logs, making it suitable for both beginners and experts. A graphical depiction of a process may be provided by each tool, even though the notation employed may differ, according to Table 1 in the report, which compares the core features of open source and proprietary process mining tools. The choice of tool may be based on its compatibility with the process modelling notation already utilized by the organization.

ProM is capable of performing all core process mining operations, while Disco and Celonis only handle a subset of these. Disco is considered user-friendly and efficient in processing event logs, making it a good option for both novice and experienced users. Table 1 in the paper shows that all three tools can filter, discover, and visualize processes. ProM is open-source, while Disco and Celonis are commercial. Creating business process models by analysing event logs can be done by Process mining. The purpose is to gain new insights that can be used to enhance a company's existing operations. By extracting important information from the event logs, process mining provides a fresh perspective on business processes.

Applications:

[26] The primary goal is to give academic advisers with data and prediction models to help them make better decisions. The authors describe the design and implementation of Learning Analytics Dashboard for Advisors, as well as the findings of user studies assessing the dashboard's use and usability.

Through comparative and predictive analysis, LADA, a learning analytics application, aids advisors in making judgements. It enables advisors to analyse more options in less time before reaching a final choice. The academic achievement of pupils recommended using LADA and those advised using conventional techniques should be compared in a longitudinal study in the future.

[27] Handling large amounts of educational data presents a major challenge when it comes to creating individualized learning experiences and curricula. This study presents a paradigm for learning analytics designed to help with the overall management of learning data. The authors also provide the research foundation for a prototype that integrates a learning analytics dashboard with well-known learning management systems.

Results show that students who accessed the Blackboard system and discussion board more frequently had higher final grades. This highlights the significance of offering students a personalized analytical dashboard and the role that a student-centred learning analytics dashboard tailored to their needs plays in enabling them to access information in a timely manner.

Results and discussions:

This case study analyzes data collected between 2000 and 2009 about students participating in an EE program. The target dataset included 648 first-year students from VWO or polytechnical education, with the goal of predicting their success in obtaining a propedeuse within three years. Three datasets were considered: pre-university data (495 instances), university grades (516 instances), and both sets of attributes (516 students). Various classifiers, including decision tree algorithms, a Bayesian classifier, a logistic model, and a rule-based learner, were used to compare their performance in predicting student success. Attribute ranking revealed that VWO Science mean, VWO main, and VWO Math mean were the most informative in the pre-university dataset. The OneR classifier had the highest accuracy (68%), with no other algorithm significantly outperforming it. Similarly, the OneR algorithm was the best in the university grades dataset, with the CART classifier showing a statistically significant improvement in accuracy compared to the baseline. The CART classifier learned a compact tree with five leaves, using Linear Algebra as the root and CalcA, Calc1, and Project nAttempts as further discriminators. When both sets of attributes were used, the Random Forest classifier had the highest accuracy (85.4%), but the J48 algorithm was more interpretable with only five attributes. Overall, the study found that a simple algorithm like OneR was highly effective in predicting student success using pre-university or university data. The study also emphasized the importance of attribute selection and correlation analysis in developing accurate and interpretable models.

Conclusion:

The literature review found that process mining can help assess student scoring patterns. Using process mining can identify areas for improvement in the educational process. It is possible to extract and visualise data from a database of student learning activities using process mining tools and techniques. It should be converted into an event log that can be evaluated to find bottlenecks or irregularities in the learning management process. This can provide valuable insights for teachers and administrators to make informed decisions about how to optimise the educational process and enhance student learning outcomes. In conclusion, the application of these techniques in higher education holds the potential for augmenting the efficiency of the learning-teaching process and enhancing the learning experience of students.

References:

[1] C. Romero, S. Ventura, Educational data mining: A survey from 1995 to 2005, Expert Systems with Applications, Volume 33, Issue 1, 2007, Pages 135-

146, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2006.04.005

[2] Pechenizkiy, M., Trcka, N., Vasilyeva, E., Aalst, van der, W. M. P., & De Bra, P. M. E. (2009). Process mining online assessment data. In T. Barnes, M. Desmarais, C. Romero, & S. Ventura (Eds.), Educational Data Mining 2009: 2nd International Conference on Educational Data Mining: proceedings [EDM'09], Cordoba, Spain. July 13, 2009 (pp. 279-288). International Working Group on Educational Data Mining.

[3] Aeiad, E., Meziane, F. An adaptable and personalised E-learning system applied to computer science Programmes design. Educ Inf Technol 24, 1485–1509 (2019). https://doi.org/10.1007/s10639-018-9836-x

[4] Bogarín, Alejandro & Cerezo, Rebeca & Romero, Cristóbal. (2017). A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 8. 10.1002/widm.1230.

[5] Trcka, Nikola & Pechenizkiy, Mykola. (2009). From Local Patterns to Global Models: Towards Domain Driven Educational Process Mining. ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications. 1114-1119. 10.1109/ISDA.2009.159.

[6] Aalst, Wil & La Rosa, Marcello & Santoro, Flávia. (2015). Business Process Management: Don't Forget to Improve the Process! Business & Information Systems Engineering. 58. 10.1007/s12599-015-0409-x.

[7] Azeta, Ambrose & Agono, Frank & Adesola, Falade & Azeta, ea & Nwaocha, Vivian. (2020). A Digital Twin Framework for Analysing Students' Behaviours Using Educational Process Mining. 10.21203/rs.3.rs-51184/v1.

[8] Romero, Cristóbal & Ventura, Sebastian & García, Enrique. (2008). Data mining in course management systems: Moodle case study and tutorial. Computers & Education. 51. 368-384. 10.1016/j.compedu.2007.05.016.

[9] Maldonado, Jorge & Pérez-Sanagustín, Mar & Kizilcec, René & Morales, Nicolás & Munoz-Gama, Jorge. (2017). Mining Theory-Based Patterns from Big Data: Identifying Self-Regulated Learning Strategies in Massive Open Online Courses. Computers in Human Behaviour . 80. 10.1016/j.chb.2017.11.011.

[10] Günther, Christian & Aalst, Wil. (2006). A Generic Import Framework for Process Event Logs. LNCS. 4103. 81-92. 10.1007/11837862_10.

[11] R Kabra and R S Bichkar. Article: Performance Prediction of Engineering Students using Decision Trees. International Journal of Computer Applications 36(11):8-12, December 2011

[12] Porouhan, Parham & Premchaiswadi, Wichian. (2015). Process Simulation and Pattern Discovery through Alpha and Heuristic Algorithms. 10.1109/ICTKE.2015.7368472.

[13] Hameed AlQaheri, Mrutyunjaya Panda, An Education Process Mining Framework: Unveiling Meaningful Information for Understanding Students' Learning Behavior and Improving Teaching Quality, Information, 10.3390/info13010029, 13, 1, (29), (2022).

[14] Leemans, E. Poppe and M. T. Wynn, "Directly Follows-Based Process Mining: Exploration & a Case Study," 2019 International Conference on Process Mining (ICPM), 2019, pp. 25-32, Doi: 10.1109/ICPM.2019.00015.

[16] Rozinat, Anne & Aalst, Wil. (2008). Conformance checking of processes based on monitoring real behaviour . Information Systems. 33. 64-95. 10.1016/j.is.2007.07.001.

[17] Buijs, Joos & Dongen, Boudewijn & Aalst, Wil. (2012). On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery. 7565. 305-322. 10.1007/978-3-642-33606-5_19.

[18] Salazar-Fernandez, Juan & Munoz-Gama, Jorge & Maldonado, Jorge & Bustamante, Diego & Sepulveda, Marcos. (2021). Backpack Process Model (BPPM): A Process Mining Approach for Curricular Analytics. Applied Sciences. 11. 4265. 10.3390/app11094265.

[19] Kurniati AP, Rojas E, Hogg D, Hall G, Johnson OA. The assessment of data quality issues for process mining in healthcare using Medical Information Mart for Intensive Care III, a freely available e-health record database. Health Informatics J. 2019 Dec;25(4):1878-

1893. Doi: 10.1177/1460458218810760. Epub 2018 Nov 29. PMID: 30488750.

[20] Wen, L., van der Aalst, W.M.P., Wang, J. et al. Mining process models with non-free-choice constructs. Data Min Knowl Disc 15, 145–180 (2007). https://doi.org/10.1007/s10618-007-0065-y

[21] Juhaňák, Libor & Zounek, Jiří & Rohlíková, Lucie. (2017). Using process mining to analyze students' quiz-taking behaviour patterns in a learning management system. Computers in Human Behaviour . 92. 10.1016/j.chb.2017.12.015.

[22] Pablo Munguia, Amelia Brennan, Sarah Taylor, David Lee, A learning analytics journey: Bridging the gap between technology services and the academic need, The Internet and Higher Education, Volume 46, 2020, 100744, ISSN 1096-7516, https://doi.org/10.1016/j.iheduc.2020.100744.

[23] Dongen, Boudewijn & Medeiros, Ana & Verbeek, H. & Weijters, A. & Aalst, Wil. (2005). The ProM Framework: A New Era in Process Mining Tool Support. Lecture Notes in Computer Science. 3536. 444-454. 10.1007/11494744_25.

[24] Badakhshan, Peyman & Geyer-Klingeberg, Jerome. (2020). Celonis Process Repository: A Bridge between Business Process Management and Process Mining.

[25] Devi, Aruna and Dr. M V Sudhamani. “An Informative and Comparative Study of Process Mining Tools.” (2017).

[26] Gutierrez, Francisco & Seipp, Karsten & Ochoa, Xavier & Chiluiza, Katherine & De Laet, Tinne & Verbert, Katrien. (2018). LADA: A learning analytics dashboard for academic advising. Computers in Human Behaviour . 107. 10.1016/j.chb.2018.12.004.

[27] Aljohani, Naif & Daud, Ali & Abbasi, Rabeeh & Alowibdi, Jalal & Basheri, Mohammad & Aslam, Muhammad. (2018). An Integrated Framework for Course Adapted Student Learning Analytics Dashboard. Computers in Human Behaviour . 92. 10.1016/j.chb.2018.03.035.

[28] Dekker, Gerben & Pechenizkiy, Mykola & Vleeshouwers, Jan. (2009). Predicting Students Drop Out: A Case Study... Computers, Environment and Urban Systems. 41-50.