A Project Report On

Analysis of Student's Scoring Patterns using Process Mining techniques

Submitted in partial fulfillment of the requirement for the 8^{th} semester

Bachelor of Engineering

in

Computer Science and Engineering

DAYANANDA SAGAR COLLEGE OF ENGINEERING

(An Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE & ISO 9001:2008 Certified)

Accredited by National Assessment & Accreditation Council (NAAC) with 'A' grade

Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-560111

Submitted By

Srungarapu Sai Sri Nandan 1DS19CS172

Vonkayala Ashwini 1DS19CS192

Pallapothu Lakshmi Sharanya 1DS19CS730

Vishwakumar K Hirehalli 1DS19CS191

Under the guidance of

Dr. Deepak. G
Associate Professor, CSE, DSCE

Rohit Ranjan
Alumni, CSE , DSCE

2022 - 2023

Department of Computer Science and Engineering

DAYANANDA SAGAR COLLEGE OF ENGINEERING

Bangalore - 560111

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Dayananda Sagar College of Engineering

(An Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE & ISO 9001:2008 Certified)

Accredited by National Assessment & Accreditation Council (NAAC) with 'A' grade

Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-560111

Department of Computer Science & Engineering

CERTIFICATE

This is to certify that the project entitled **Analysis of Student's Scoring Patterns using Process Mining techniques** is a bonafide work carried out by **Srungarapu Sai Sri Nandan [1DS19CS172]**, **Pallapothu Lakshmi Sharanya [1DS19CS730], Vonkayala Ashwini[1DS19CS192]** and **Vishwakumar K Hirehalli [1DS19CS191]** in partial fulfillment of 8th semester, Bachelor of Engineering in Computer Science and Engineering under Visvesvaraya Technological University, Belgaum during the year 2022-23.

Dr. Deepak. G	**Dr. Ramesh Babu D R**	**Dr. B G Prasad**
(Internal Guide)	Vice Principal & HOD	Principal
Associate Prof. CSE, DSCE	CSE, DSCE	DSCE
Signature:...........	Signature:...........	Signature:...........

Name of the Examiners: Signature with date:

1.......................... ....

2............................ ....

Acknowledgement

We are pleased to have successfully completed the project Analysis of Student's Scoring Patterns using Process Mining techniques. We thoroughly enjoyed the process of working on this project and gained a lot of knowledge doing so.

We would like to take this opportunity to express our gratitude to **Dr. B G Prasad**, Principal of DSCE, for permitting us to utilize all the necessary facilities of the institution.

We also thank our respected Vice Principal, HOD of Computer Science & Engineering, DSCE, Bangalore, **Dr. Ramesh Babu D R**, for his support and encouragement throughout the process.

We are immensely grateful to our respected and learned guide, **Dr. Deepak G**, Associate Professor CSE, DSCE and our co-guide **Rohit Ranjan**, Alumni, CSE, DSCE for their valuable help and guidance. We are indebted to them for their invaluable guidance throughout the process and their useful inputs at all stages of the process.

We also thank all the faculty and support staff of Department of Computer Science, DSCE. Without their support over the years, this work would not have been possible.

Lastly, we would like to express our deep appreciation towards our classmates and our family for providing us with constant moral support and encouragement. They have stood by us in the most difficult of times.

Srungarapu Sai Sri Nandan 1DS19CS172

Vonkayala Ashwini 1DS19CS192

Pallapothu Lakshmi Sharanya 1DS19CS730

Vishwakumar K Hirehalli 1DS19CS191

Abstract

Process mining is an emerging and revolutionary technique utilized to examine and optimize various business processes. By employing software agents, it actively monitors and evaluates the underlying activities within a process over a specific timeframe. This comprehensive analysis entails capturing essential metrics related to business process performance, crucial indicators, as well as transaction and event-level data. The utilization of process mining empowers organizations to automate laborious and time-consuming tasks, thus uncovering systemic concerns and identifying opportunities for process enhancement. The focus of this project revolves around the utilization of process mining techniques to discern a student's scoring patterns. In particular, the study delves into the thorough examination and evaluation of diverse algorithms employed in calculating students' test scores. By employing process mining, we aim to unravel valuable insights into students' performance, ultimately aiding educators and institutions in understanding the underlying patterns and trends related to student achievement. By leveraging the power of process mining, this project aims to contribute to the field of education by shedding light on the intricate relationship between students' academic performance and the underlying processes used to calculate their scores. The findings of this research endeavor have the potential to revolutionize educational practices, enabling educators to identify key factors influencing student success and formulate tailored strategies to enhance learning outcomes..

Contents

List of Figures

List of Tables

List of Abbreviations

KDD Knowledge Discovery in Databases

EDM Educational Data Mining

ICT Information and Communication Technology

PLE Personalized Learning Environments

APELS A Personalised and Adaptable E-Learning System

EPM Educational Process Mining

LMS Learning Management Systems

MOOCs Massive Open Online Courses

SRL Self-Regulated Learning

DFBPM Directly Follows-Based Process Mining

PAIS Process-Aware IT systems

BPPM Backpack Process Model

ETM Evolutionary Tree Miner

EPC Event-Driven Process Chain

BPM Business Process Management

BPMN Business Process Modeling Notation

LADA Learning Analytics Dashboard for Academic Advising

AMBA Analyse My Blackboard Activities

API Application Programming Interface

XES eXtensible Event Stream

PNML Petri Net Markup Language

CSV Comma-Separated Values

Introduction

The Problem

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. KDD and data mining 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.

Figure 1: Problems with Old Education System

The combination of technologies in education, particularly web-based education, has seen significant development in recent years. It offers flexibility in location and hardware, along with personalized learning using adaptive and intelligent systems that understand preferences, objectives, and expertise. Data mining plays a crucial role in improving and evaluating e-learning systems by extracting useful patterns from large data collections, specifically through KDD.

However, applying data mining to e-learning environments poses certain challenges. The diverse nature of the educational domain, with its varied subjects and disciplines, necessitates the development of domain-specific models and algorithms. This ensures accurate and relevant results tailored to the specific requirements of each educational field. Moreover, educational data itself is complex, heterogeneous, and dynamic, requiring sophisticated preprocessing and fusion techniques to integrate and analyze diverse information effectively.

The aims of data mining in e-learning can vary widely, ranging from identifying at-risk students to personalizing learning materials and optimizing instructional strategies. Each objective requires different techniques, algorithms, and evaluation metrics. Thus, aligning the data mining approach with the specific aims of the e-learning system is crucial.

The methodology also plays a significant role in data mining for e-learning. From data collection to evaluation and interpretation, each stage requires careful consideration of the e-learning context and data limitations. Privacy and ethical concerns related to sensitive learner data must also be addressed, ensuring compliance with regulations to maintain learner trust and confidentiality.

Despite these difficulties, data mining presents significant opportunities for improving e-learning platforms. Educators can learn important information about the behavior, preferences, and learning trajectories of students by identifying patterns in learner interaction data. These discoveries assist in the creation of adaptive and intelligent systems that facilitate individualized learning. In addition, data mining enables the development of recommender systems that make pertinent learning resource recommendations based on the profiles and interactions of learners.

In conclusion, by identifying significant patterns in huge datasets, data mining has the potential to significantly enhance e-learning systems. But there are issues that must be resolved, including the diverse educational domain, complicated data, disparate goals, and methodological considerations. By doing this, educators can fully utilize data mining to improve web-based learning and increase its effectiveness and efficiency.

Real World Application

Personalized Learning

Data mining technologies can be used by educators to analyse student data and uncover distinctive learning preferences and behaviours. Teachers can modify the learning environment to suit individual needs by observing how students utilise learning platforms and identifying helpful resources. This enables targeted modifications to respond to each student's ability level, such as personalizing instructional materials and offering suggestions for supplementary resources based on their unique needs. Data mining also helps teachers to provide students with targeted feedback, highlighting their strengths and areas for development.

Tailoring education to individual student's preferences and abilities increases their motivation and involvement. When students receive customized learning experiences that align with their personal preferences and skill levels, they are more likely to remain dedicated to their studies, leading to better learning results. By utilizing data mining, educators can design an educational journey that is more effective and student-centric, fostering a positive atmosphere for learning.

Early Intervention and Support

By analyzing student data, such as completion rates, interaction patterns, and assessment results, educators have a powerful tool at their disposal to identify students who may be at risk of falling behind or becoming disengaged from their studies. This process, known as educational process mining, allows educators to raise red flags when certain patterns emerge. For example, if a student consistently struggles to complete assignments or shows a decline in participation, these signs can indicate potential issues. Armed with this information, educators can take proactive measures to support these students.

They can extend their support, carry out focused interventions, or create specialised remediation plans that are adapted to each person's need. This early intervention strategy is essential because it helps students succeed by preventing subsequent academic problems. Teachers can guarantee that pupils have the help and direction they need to overcome difficulties and succeed academically by addressing problems early on.

Curriculum Design and Improvement

Education professionals have a valuable resource in student data, enabling them to evaluate the efficiency of the curriculum, teaching approaches, and evaluation techniques. By examining student data, educators can gain insights into how students engage with course content, identify their strengths and weaknesses, and evaluate the impact of different teaching methods on their performance. This data-driven approach empowers educators to make informed decisions to enhance the curriculum.

They can optimize teaching materials, refine instructional strategies, and develop assessments that align with student learning objectives. As a result, educational institutions can improve the quality and relevance of their programs, ensuring that students have access to the most effective and engaging learning opportunities. By leveraging student data, educators can continuously enhance their teaching practices, fostering a dynamic and impactful educational environment.

Organisation of Project Report

The project report is organized as follows:

In Chapter (2) we discuss the problem statement and the proposed solution. We also take a look at the systems that exist today and the drawbacks they face.

Chapter (3) takes a more in-depth look at various ready-made tools and custom-built software that exist, with a survey of existing literature available.

Chapter (4) looks at the architecture of the proposed solution with an overview of the system design, utilizing system block diagrams and data flow diagrams.

Chapter (5) dives into the Implementation of the solution, by describing the hardware and software requirements, along with dataset descriptions and implementation details.

Chapter (6) describes our testing process, while

Chapter (7) looks at our experimentation process and the obtained results.

Chapter (8) summarizes our findings and concludes the paper.

Problem Statement and Proposed Solution

Problem statement

In the current education system, there are a lot of activities with an unstructured way of functioning, as a result of which the quality of student learning processes is decreasing.

Existing Systems

The existing education system has long been a cornerstone of society, providing a structured framework for knowledge dissemination and skill development. However, like any complex system, it has its share of challenges and areas that can be improved. One aspect that can benefit from advancements in data analysis and insights is the understanding of individual student needs and learning patterns. Traditional approaches often rely on standardized assessments and limited feedback mechanisms, which may overlook the unique strengths, weaknesses, and preferences of each student.

In order to better address the different requirements of students, educators may customize instructional tactics and resources by acquiring greater insights regarding student engagement, progress, and performance. A more effective and interesting learning environment for all students may be achieved by utilizing educational process mining to discover parts of the curriculum that need to be improved or changed. The incorporation of data-driven insights has the potential to alter the educational system by facilitating individualized instruction, early intervention, and decision-making based on the best available data.

Techniques used

The efficacy of the educational system has been examined and improved using a variety of ways. These methods are intended to meet the various requirements of students and provide them with the greatest learning experience possible. Analyzing student performance via standardized tests is one widely utilized strategy. These tests offer useful information on students' academic progress, enabling teachers to spot areas where pupils might need further assistance or enrichment. However, this method frequently fails to adequately capture the subtleties of unique learning preferences, styles, and development across time.

In addition to formative evaluations, which offer continuous feedback during the learning process, instructors have used them to get around this constraint. Teachers can get a more comprehensive picture of students' learning and adjust their instruction by utilizing strategies like quizzes, assignments, and classroom observations. These techniques, meanwhile, can take a lot of time, and they might not give a whole picture of student involvement and development.

Utilizing learning analytics is a different strategy that has gained popularity. Educators may examine student interactions, monitor progress, and spot trends in learning behavior by utilizing the data produced by digital learning platforms. With this strategy, it is possible to identify difficult pupils and offer them tailored help. Additionally, learning analytics can offer perceptions into the efficiency of teaching resources, assisting educators in enhancing and improving curriculum design.

In addition, qualitative techniques like questionnaires, interviews, and focus groups are employed to get teacher, parent, and student input. These techniques offer insightful information on the educational process, teaching methods, and opportunities for development. Education systems may better comprehend the various needs and expectations of their constituents and make wise decisions by embracing stakeholder viewpoints.

There is still room for improvement even if these tactics have enhanced the educational system. There is a lot of promise for educational process mining, which combines advanced data mining and machine learning techniques. By examining vast amounts of educational data, such as student interactions, performance, and engagement, educational process mining may provide deeper insights into individual learning requirements, optimise curriculum design, and promote evidence-based decision-making. This comprehensive and data-driven approach has the ability to transform the educational system by ensuring that every student gets a personalised, effective, and engaging learning experience.

Proposed Solution

By offering useful insights and allowing more effective processes, educational process mining technologies and approaches present a possible answer for enhancing the educational system. Educational process mining enables a complete examination of the operation of various system processes by utilising event log data. This study can assist in locating flaws, bottlenecks, and loopholes that reduce the efficacy of the educational system as a whole. The key next step is to create a model that results in a more effective way to carry out these procedures, building on this basis.

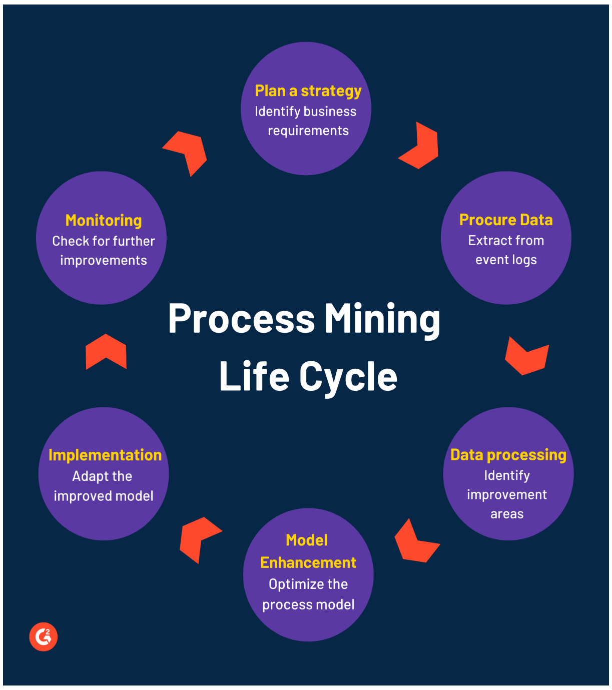

Figure 2: Process mining - Life cycle

To improve the education system using educational process mining, the following steps can be taken:

- Data Collection and Integration: Data gathering and integration are the first steps in educational process mining. A variety of sources, such as student records, learning management systems, and test results, are used by educators to compile pertinent data. This information includes interactions, involvement, performance, and other elements related to student engagement. The acquired datasets must be combined in order to perform an exhaustive analysis.

  Data integration unifies the disparate datasets by combining and merging them. Teachers can find patterns, trends, and correlations by combining data that might be difficult to identify from separate datasets alone. It makes it easier to comprehend the complex interactions between many factors that affect student performance. Additionally, data integration aids in the creation of forecasting models and algorithmic decision-making.

  Identification of at-risk kids and the execution of preventative interventions are made possible by integrated data. In order to cater to individual requirements, educators can customise teaching methods and remedial techniques. Educators foster student progress and avert academic issues by utilising integrated data.

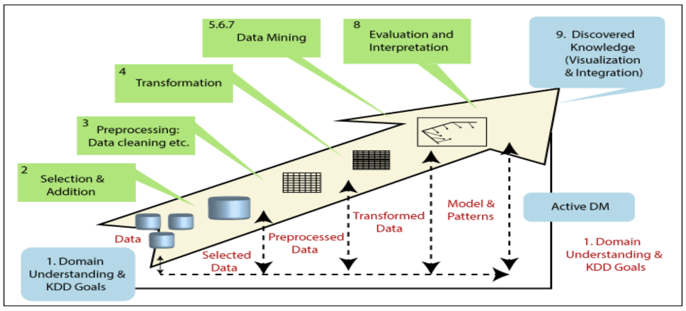

Figure 3: KDD Process in Data Mining

- Process Analysis and Visualization: Once data has been collected and integrated in the educational process mining framework, the next step is process analysis and visualization. Educational process mining techniques are applied to analyze the collected data and extract valuable insights into the existing processes within the education system.

  Educators may better comprehend the many relationships, activities, and events that take place during the learning process by using process analysis. They can find relationships, trends, and patterns in the data that affect student results. By identifying the advantages and disadvantages of existing procedures, this study aids educators in deciding what needs to be improved.

  Process visualisation is used to increase knowledge and effectively convey the results. To show the order of events, decision points, and potential bottlenecks within the educational processes, educators might build process flow diagrams or process maps. The numerous processes and interactions involved in the learning experience are clearly outlined by these graphic representations.

  Process visualization enables educators to identify areas of inefficiency or areas where interventions can be implemented to enhance student engagement and learning outcomes. It helps educators pinpoint potential bottlenecks or areas where students may face challenges or disengagement. By visualizing the processes, educators can effectively communicate their findings and collaborate with stakeholders to implement targeted interventions and improvements.

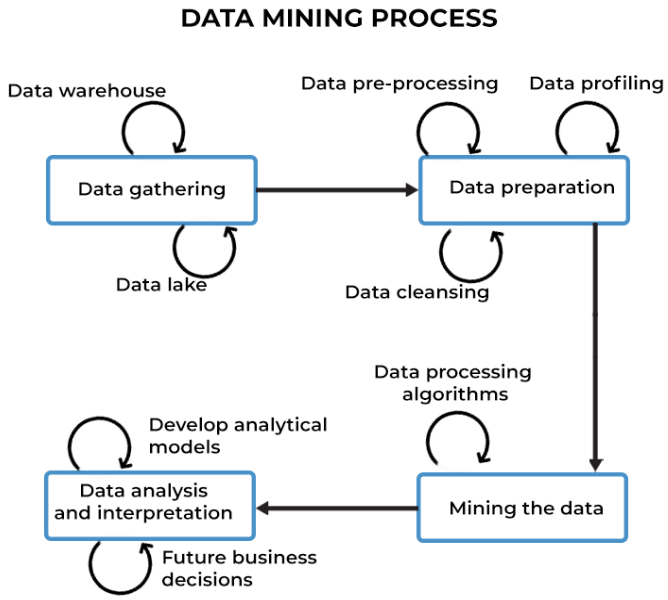

Figure 4: Data Analysis Process

- Identify Improvement Opportunities: In educational process mining, the identification of improvement opportunities is crucial. By analyzing process flow diagrams, educators can pinpoint areas for enhancement. This includes identifying bottlenecks that slow down the system, detecting loops or redundancies that hinder progress, and uncovering gaps in the learning journey.

  Bottlenecks represent points where the flow of activities is impeded, and addressing them improves overall system efficiency. Detecting loops or redundancies allows for streamlining and eliminating unnecessary steps. Identifying gaps highlights areas where additional activities or resources are needed to support students' progress.

  Teachers prioritize areas for development using the knowledge they obtain from examining process flow diagrams. They can put into place specific solutions to deal with gaps, loops, and bottlenecks. Teachers may create a more efficient learning environment that promotes student achievement and engagement by continually recognizing and addressing improvement possibilities.

  Mining educational processes help the educational system's constant efforts to improve. Education professionals may optimize the learning journey and guarantee a more effective and efficient educational experience for students by improving processes based on recognized improvement possibilities.

- Model Design and Simulation: Model creation and simulation are essential elements of educational process mining. Educators develop a new model based on the analysis findings in order to put in place better practices and increase the efficacy and efficiency of the educational system.

  The new approach optimizes the learning process by using technology, automation, and personalized learning tactics. By utilizing these components, instructors want to improve curriculum delivery, simplify administrative processes, and give students individualized help.

  The model is tested and improved through simulation prior to deployment. This enables instructors to mimic scenarios from real-world settings, monitor the model's performance, spot any problems, and make the required adjustments. The iterative method makes the model is strong and able to handle the areas that need development.

  Teachers may change the educational system by integrating educational process mining, model design, and simulation. The analytical results may be used by educators to design a new model that provides a direction for enhanced processes. They improve the learning process by using technology, automation, and personalized learning. The model's preparedness for use in the actual world is ensured by educators through simulation.

  By using this all-encompassing strategy, educators can build a more effective and personalized learning environment that encourages student achievement.

- Implementation and Monitoring: The implementation and monitoring phase is essential in educational process mining. Teachers carefully introduce the new model in a controlled environment while keeping an eye on how it affects student performance, engagement, and system performance as a whole.

  To guarantee a seamless transition during implementation, educators offer the required training and assistance. They simultaneously gather and analyze data to carefully track the consequences of the new model. These statistics cover student achievement, levels of involvement, and the efficiency of the educational system.

  Continual observation enables instructors to pinpoint areas that need correction or enhancement. They hone the model to fix any flaws by gathering information and making wise selections. By using an iterative process, the educational system is made to adapt to students' changing demands.

The stage of implementation and monitoring involves an ongoing cycle of development. By watching the new model's effects, collecting data, and making the required adjustments, educators may maximise it. They improve learning, support student achievement, and provide an adaptable learning environment by successfully implementing and overseeing the model.

- Continuous Improvement: Continuous improvement is a key practice in educational process mining. It involves regularly analyzing implemented processes to identify developing problems, emerging trends, and areas for further development. By gathering and analyzing data, educators and policymakers can stay proactive in addressing challenges and making informed decisions.

  Through continuous data collection and analysis, educators can detect patterns and trends that impact the effectiveness of educational processes. They can identify areas where students face difficulties, engagement levels may decrease, or current procedures may not align with desired outcomes. This enables early intervention and targeted improvements to enhance the educational experience.

  Policymakers also benefit from data analysis, using insights to inform decision-making. By identifying systemic issues and allocating resources effectively, they can develop policies that support positive educational outcomes.

  Continuous improvement through educational process mining fosters adaptability and innovation. It allows educators and policymakers to respond to changing needs, evolving teaching methods, and technological advancements. By embracing this approach, educational institutions can create a dynamic system that promotes student success and lifelong learning.

The educational system may transition to a more effective and efficient approach by utilizing educational process mining technologies and practices. This data-driven, iterative strategy improves learning for all students by identifying and addressing systemic difficulties, ensuring that instruction is individualized for each student, and ensuring that it is matched to their requirements.

Literature Survey

Scenario

The amalgamation of technologies in the realm of education has emerged as a pivotal advancement in recent times, especially in light of the ascendance of web-based instruction. This paradigm offers a malleable approach in terms of physical location and hardware, and possesses the capacity for personalization through the utilization of adaptive and intelligent systems that discern your inclinations, objectives, and proficiencies. KDD and data mining [1] can be harnessed to refine and appraise e-learning systems by extracting invaluable patterns from extensive data repositories.

Educators are compelled to explore alternative avenues, such as scrutinizing data derived from web servers, to acquire feedback regarding students' learning encounters within electronic learning environments. Data mining techniques are progressively gaining popularity in web-based educational settings, as they enable the analysis of learner interaction data. Although data mining has evinced success in the realm of e-commerce, it encounters specific challenges when applied to e-learning due to variances in domain, data, objectives, and methodology.

Solutions

[2] The assessment conducted through online means has emerged as a pivotal facet of contemporary education, finding utility in both digital learning and blended learning formats. It serves as a valuable instrument for self-evaluation as well as official examinations, often augmenting or even supplanting traditional evaluation methods. The astute analysis of data can provide deeper insights into student performance, examination quality, and specific questions. Despite its advantages, there persist challenges in terms of organizing and establishing diverse assessment procedures.

For instance, in the realm of multiple-choice question testing, the method of navigating between questions and the fixed or variable order thereof can wield a significant impact. Recently, researchers have delved into exploring student navigation in digital learning systems with the objective of identifying individual navigational styles, reducing cognitive load, enhancing the usability and efficacy of digital learning systems, and facilitating personalized navigation.

Conventional data mining techniques are utilized to examine various types of educational data, including test data, in order to uncover correlations, subgroups, and emerging patterns. However, these techniques lack a comprehensive process perspective and provide limited insight into the overall assessment process. The authors of this paper seek to acquaint the

Educational Data Mining (EDM) community with process mining and its associated tool, ProM. They showcase the application of several ProM plug-ins to analyze assessment data from two distinct studies. In the first study, students were required to answer questions in a predetermined sequence and had the option to request immediate feedback. In the second study, students were granted the flexibility to answer questions in any order, revisit previous responses, and make revisions.

[3] Information and Communication Technology (ICT) and most recently digital media have revolutionized pedagogy and education, but the data inside this system as a whole has remained static and has not been customized for individual students.

The problem of providing personalized information that is catered to each student's particular learning style is effectively addressed by Personalized Learning Environments (PLE). Due to the abundance of content available on the global internet, it can be challenging for students to find materials that suit their own interests.

An effort to address these challenges is represented by A Personalised and Adaptable E-Learning System (APELS), which analyses students' learning preferences, then adapts and provides resources to help them on their educational path.

Process Mining Techniques

[4]Learning environments that have embraced technological advancements possess the remarkable ability to meticulously record and document every facet of learner engagement, from the minutiae of keystrokes to the profound depths of their educational undertakings. These innovative systems vigilantly monitor and capture an extensive spectrum of temporal data, encompassing not only the ebb and flow of virtual traffic but also the chronicles of dialogues, the evolutionary timelines of modifications, and even the subtle trails left by learners through motion tracking.

Moreover, they diligently log the utilization of diverse learning resources, ensuring a comprehensive archive of interactions and engagements that shape the educational journey in profound and meaningful ways.

[5]Process mining utilizes data, commonly known as logs, audit trails, log files, or traces, to unveil, monitor, and enhance educational processes. Process mining serves as a nexus between data mining and process modeling and analysis. Process mining, a discipline encompassed within data mining, employs a process-centric perspective in its entirety.

The traditional methodologies of EDM focus on unearthing correlations or fundamental patterns in educational data. Nonetheless, they fail to present a comprehensive visual depiction of the complete learning process. Data mining has achieved a triumph in exposing valuable patterns within data amassed in educational environments.

[6]Educational Process Mining (EPM) is an application of Process Mining techniques that is specifically designed to analyze raw educational data. EPM adopts a process-oriented approach, aiming to bring clarity to previously unknown or partially known educational processes by leveraging event data. Unlike focusing solely on isolated patterns, Process Mining emphasizes the entire process, recognizing the significance of comprehensive process models and the interconnections among activities.

In contrast, traditional EDM procedures lack the inclusion of essential methods like process discovery, compliance testing, and bottleneck analysis. By incorporating these additional techniques, EPM offers a more holistic and comprehensive understanding of educational processes, contributing to improved insights and decision-making in the educational domain.

Data Collection

[7]Learning Management Systems (LMS) record every activity occurring on the platform, employing methods to effectively monitor and analyze the data in order to develop trends in student online learning. This framework utilizes inductive and fuzzy mining techniques to generate a process model that illustrates learning patterns through a virtual representation.

The objective is to provide higher institutions with the capacity to extract profound insights from their data, thereby enhancing their educational processes. The resultant refined model can be utilized to formulate recommendations and foster improved student-instructor interaction within educational IT systems. The study's findings have significant implications for research in open and distance learning.

[8] Data mining is a field that aims to improve methods of analyzing specific types of data in an educational setting. The article discusses a fictional lesson on the Moodle platform and examines how data mining can be implemented in learning management systems.

The authors provide detailed information on the entire process of extracting e-learning data, as well as guidance on using various data mining techniques. When educators use course management systems, they can utilize data mining to gain deeper insights into online learners. This can be achieved by employing visualization techniques to get a comprehensive overview of student usage data or by using clustering or other methods to create a classifier. Currently, the authors are developing a specialized Moodle data mining tool for online instructors to use within the Moodle platform.

[9] An analysis of the interaction sequences of 3458 learners in three Massive Open Online Courses (MOOCs) was conducted employing process mining techniques. By juxtaposing these patterns against theoretical frameworks of SRL, we identified three distinct cohorts of learners. Those who exhibited inconsistent and aimless behavior manifested lower levels of SRL and performed inadequately in comparison to the remaining two groups.

The aim of this study was to investigate the correlation between behavior within a digital learning environment and the underlying theoretical processes of learning, specifically focusing on SRL strategies within MOOCs. Both self-reported data and process mining of learner behavior were employed. For instance, one interaction sequence entailed a learner completing multiple video lectures prior to engaging in an assessment, which could be associated with the Reviewing Records SRL strategy. Another pattern demonstrated a learner’s attempt to undertake an assessment immediately after concluding a video lecture, which could be linked to the Task Exploration SRL strategy.

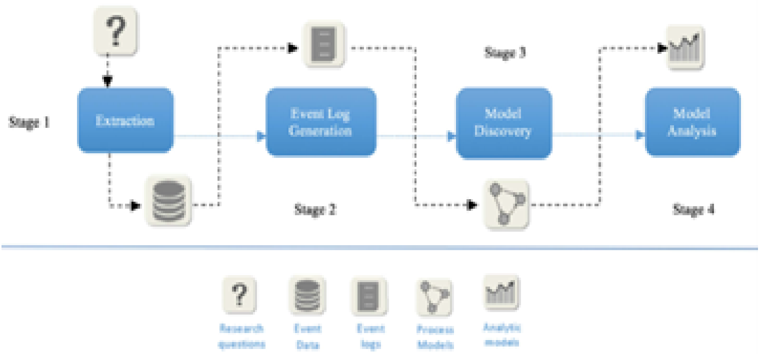

Figure 5: Task Exploration SRL strategy

Process Discovery

[11] The decision tree algorithm is employed for the purpose of classifying data based on attributes that are either continuous or discrete in nature. It takes on a visual representation in the form of a tree-like structure, wherein internal nodes contain tests to evaluate the data, and the outcomes of these tests are depicted by arcs. The leaf nodes, on the other hand, represent the corresponding labels.

By utilizing a variety of attributes pertaining to the students, we can employ decision trees to scrutinize and prognosticate their academic performance. Factors such as past scholastic achievements, educational background, previous institutions, and other relevant attributes can be employed to anticipate student success and aid those who are at risk in enhancing their learning and achieving better outcomes. Additionally, we can examine the actions or attributes exhibited by high-scoring students in order to conduct further analysis or seek avenues for improvement.

[12]The alpha algorithm is a process mining tool used to discern routing patterns within an event log procured from a private institution situated in Bangkok, Thailand. The focus of the inquiry primarily revolved around the ”commencement” and ”completion” stages of the process instances, while inadvertently disregarding the actions and activities transpiring during the ”in-progress” phase. The ultimate Alpha model, presented as a captivating Petri net, vividly illustrates the control flow pertaining to the training activities for the course, as well as their intricate interdependence.

However, it was discovered that the model concluded with five activities/tasks, as opposed to the originally predicted two, suggesting the existence of three undesirable activities/tasks within the training process. To create models that are less susceptible to noise and incompleteness in the logs, the Heuristic Miner method was employed as a process discovery tool.

[13]To evaluate the relative significance of behaviors and the interconnections between occurrences, the Heuristic Miner algorithm duly considered two indispensable metrics: Significance and Correlation.

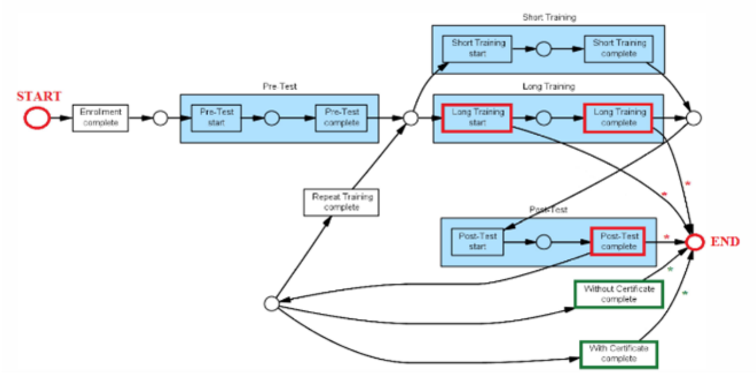

Figure 6: The resulting Heuristic model

[14] Directly Follows-Based Process Mining (DFBPM) encompasses the examination of the ’directly follows’ relationship, wherein if event B follows event A within a process, it suggests that event B is a probable direct outcome of event A. Process models are constructed utilizing this correlation extracted from the event logs, and this algorithm is commonly employed in conjunction with other algorithms to achieve a comprehensive comprehension and enhancement of

the processes. The ability to capture the interconnections between events stands as the pivotal advantage of employing this technique.

Subsequently, the authors delve into a case study wherein they employ DFBPM to scrutinize an actual event log acquired from a manufacturing firm. The objectives of the study were to ascertain the company's production process and assess the efficacy of DFBPM in contrast to other process mining approaches such as the alpha algorithm and heuristic miner. As per the findings of the study, DFBPM successfully identified the company's production process and determined the relationships between operations. Moreover, the study revealed that DFBPM outperformed the other methods in terms of F1 score, accuracy, and recall.

Analysis

[16]In order to manage their organizational processes, numerous companies have constructed Process-Aware IT systems (PAIS). These procedures frequently record incidents related to the implementation of actual organizational processes, but they also commonly provide clear process models that outline the proper execution of business processes.

This study presents a methodical approach for evaluating the compatibility between a process model and an event log, a vital consideration in various scenarios. The proposed approach introduces four metrics for assessing conformance, encompassing two metrics for structural appropriateness (denoted as a_S and a'_S) and two metrics for behavioral appropriateness. These metrics facilitate the quantification of compatibility, along with the introduction of two novel metrics, namely a0S and a0B, which enable the assessment of suitability in isolation. These metrics indicate the achievement of an optimal solution that is prepared for analysis. The availability of relevant log data is indispensable for systems that facilitate business processes.

[17]Replay fitness, accuracy, generalization, and simplicity are the four main quality factors that are frequently prioritized by current process discovery methods. Based on these four options, the Evolutionary Tree Miner (ETM) method gives you control over the finding process. However, only if the replay fitness passes the necessary threshold, have all dimensions been shown to be significant for process discovery.

The ETM algorithm, a genetic algorithm, optimizes the result of process discovery with regard to any of the four dimensions. Using a weighted average of the following metrics, this method uses process trees to make sure that the final model accurately characterizes the observed log and meets the requirements for being the best model.

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

- Simplicity serves as a metric for determining the complexity of the model. It compares the activity in the event log with the size of the tree. As the tree's size increases, complexity grows, leading to potential flaws in process models. Additionally, the number of operational nodes is affected when binary trees are employed. Therefore, if actions are only represented once on the tree, it is deemed simple.

- Precision, during log replay, compares the tree's execution state space. This measurement is based on escape edges, which are conceivable choices within the model but never taken in the log. If no edges can escape, accuracy is perfect. Information derived from replay fitness is utilized to determine the portion of the state space that is utilized, and events that do not fit the alignment of an activity with an event are discarded.

- Generalization calculates the frequency at which all tree nodes must be traversed in order to generate the provided log. For this purpose, we leverage the alignment produced by the replay fitness calculation. The more frequently a node is visited, the higher the confidence in its correctness or incorrectness. Conversely, poor generalization occurs when certain parts of the tree are only visited sporadically.

The aforementioned metrics are assessed on a scale of 0 to 1, with 1 denoting the optimal outcome. Replay fitness, simplicity, and accuracy can all achieve a score of one, representing an ideal state. However, generalization can only approach 1 as the limit, implying that the closer the score is to 1, the more nodes are visited. By utilizing a genetic algorithm to maximize a weighted total of these four metrics, the challenge of creating a process model that balances these metrics can be effectively overcome.

[18] Curricular analytics constitutes a division of learning analytics that delves into the correlation between curricular components and academic performance. This investigation holds substantial value for institutions of higher education seeking to evaluate the merits and drawbacks of their curriculum. Backpack Process Model (BPPM), which employs the backpack analogy to scrutinize the trajectory of the curriculum, can offer profound insights into how students navigate failed courses, the courses that necessitate retaking, and can unveil behavioral

patterns among akin students within comparable contexts.

This data can be utilized to formulate and dispense timely recommendations that foster student achievement. Furthermore, BPPM can aid counseling services in identifying students who are at risk and serve as an auxiliary instrument in this endeavor.

Execution

[19] Data plays a pivotal role in process mining, for even a minute alteration or erroneous information possesses the potential to reshape the entirety of a process captured by the algorithm. To ensure the preservation of data quality, the most optimal approach entails the utilization of the L* life-cycle model, encompassing a phase recognized as data quality assessment.

Within this L* life cycle model, we incorporate a supplementary phase of database reconstruction to evaluate the integrity of the data. This particular step follows the stages of design and justification, preceding the subsequent stages of extraction, control flow model construction, integrated process model creation, data quality evaluation, and operational support.

[20] Process mining entails gathering data from event logs, which proves particularly valuable in situations where events are documented but there is no prescribed system dictating how employees ought to fulfill their responsibilities. This article focuses on the exploration of non-restrictive choice scenarios or instances where synchronization and choice coexist.

This study elucidates a technique capable of identifying distinct types of implicit interdependencies, such as non-restrictive choice Petri nets. Process mining represents a pioneering and effective approach to extracting crucial insights from event logs. Furthermore, the discourse encompasses the ProM framework, a publicly accessible toolkit for developing diverse process mining algorithms.

[21]Process mining's primary functions entail extracting data from event logs, discerning processes, and constructing process models. Therefore, all process-related data is gathered from event logs, necessitating the files to adhere to a suitable and compatible format for the algorithm's data extraction. The CaseID serves as a distinctive field within event logs, enabling us to trace events throughout the system. Additional fields contribute to our comprehension of the event flow within a specific case.

Prior to inputting the logs into the algorithm for accurate capture and construction of process models, it is imperative to format the various occurring events appropriately. For instance, in the analysis of a quiz_assessment, the student's/case's attributes can be designated as quiz_viewed, quiz_try_viewed, quiz_try_reviewed, quiz_try_started, try_summary_viewed, quiz_try_submitted, and so forth. These activities facilitate our understanding of the quiz's lifecycle as experienced by the student.

[22] Process mining facilitates the exploration of trends or processes that have hitherto remained unidentified or overlooked due to the limited number of instances associated with them. Moreover, it enables us to comprehend the impact of these processes on overall performance. Additionally, process mining allows us to discern the deviations that occur in real time by juxtaposing the designed process model with real-time data, thereby aiding in the mitigation of these inefficiencies.

Each process necessitates a standardized flow or process model, which can serve as a benchmark to compare the resultant activities, thereby enhancing the achievement of the process objectives. For instance, in the context of course analysis, it becomes imperative to establish a set of guidelines that students should adhere to or that the system can utilize to analyze and predict students' performance, subsequently taking appropriate measures.

Tools

[23] Both the academic and business sectors have devised a multitude of techniques for extracting information from event logs. These sophisticated instruments peruse and archive log files in diverse formats and present their discoveries through diverse means. Incorporating some of these principles into other tools can pose a challenge, yet they are indeed present in certain technological advancements.

The ProM framework, which accommodates multiple languages such as Petri nets, Event-Driven Process Chain (EPC)s, formats, and social networks, was developed to facilitate process mining as an adaptable environment. The plug-ins offered by the ProM framework can be harnessed in a variety of ways, seamlessly integrated, and employed in real-world scenarios. The addition of a new plug-in to the ProM framework is a straightforward process that necessitates no alterations to the existing code.

[24]Business Process Management (BPM) has greatly benefited from the application of Process Mining, revolutionizing the manner in which processes are conceptualized, analyzed, redesigned, implemented, and monitored. A significant drawback of traditional process modeling techniques lies in their time-consuming and costly nature, involving extensive interviews and workshops, often resulting in a conceptual model that inadequately represents the actual process in operation (referred to as the "As-Is" process).

To address this challenge, Celonis has developed a suite of sophisticated toolsets known as the Process Repository. These toolsets enable comprehensive process documentation and seamless integration with Process Mining applications, thus forging a stronger connection between BPM and Process Mining. Consequently, this approach facilitates a more streamlined and impactful method for uncovering and managing processes.

[25]ProM offers an extensive array of filtering alternatives in contrast to other process mining tools. It possesses the ability to execute fundamental tasks such as process discovery, and it presents the discovered model in various formats, catering to user preferences, including BPMN diagrams, Petri nets, and fuzzy models. Disco also features a transparent filtering mechanism, albeit with a limitation to fuzzy mining models. ProM supports all core operations, whereas Disco and Celonis only encompass a subset. Disco is acknowledged for its user-friendliness and efficiency in handling event logs, rendering it suitable for both neophytes and experts alike. Each tool can provide a visual representation of a process, although the notation employed may diverge, as indicated in Table 1 of the report, which compares the fundamental attributes of open-source and proprietary process mining tools. The choice of tool may be contingent upon its compatibility with the organization's existing process modeling notation.

ProM possesses the capability to execute all fundamental process mining operations, while Disco and Celonis handle only a subset of these operations. Disco is acclaimed for its user-friendliness and efficacy in processing event logs, making it a commendable choice for both novice and seasoned users. Table 1 in the paper demonstrates that all three tools can filter, discover, and visualize processes. ProM is an open-source tool, whereas Disco and Celonis are commercial tools. Process mining is a means of creating business process models through the analysis of event logs, with the objective of gaining novel insights that can enhance an organization's existing operations. By extracting vital information from event logs, process mining offers a fresh perspective on business processes.

Applications

[7]By offering invaluable insights into students' learning patterns and behaviors, A Digital Twin Framework for Analysing Students' Behaviours Using Educational Process Mining has the potential to revolutionize the educational system. To construct virtual representations of students and their educational journeys, this framework combines the capabilities of digital twinning with educational process mining tools. Educators can better understand students' cognitive processes, learning preferences, and engagement levels by examining the data collected by these digital twins. The learning experience may then be tailored, areas for improvement

can be found, and focused interventions can be created to help struggling pupils.

Furthermore, the framework enables the detection of early warning signs, such as disengagement or academic decline, allowing educators to intervene proactively and provide timely support. Ultimately, the integration of this digital twin framework into the educational system has the potential to enhance student learning outcomes, promote individualized instruction, and foster a more inclusive and effective educational environment.

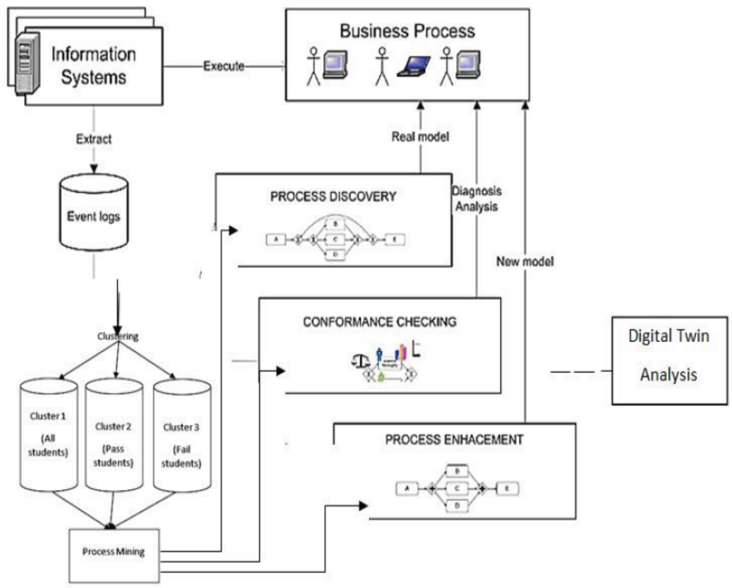

Figure 7: Overview of proposed Digital Twin Framework

[8]As demonstrated in the paper, the use of data mining techniques in course management systems offers useful advantages for online teachers and has the potential to improve the educational system. Instructors may obtain insight into student usage data, spot abnormalities, categorize new students, and spot possible problems by employing visualization, statistical analysis, clustering, and association rule mining. By seamlessly integrating data mining into the Moodle environment and enabling teachers to immediately apply mining findings to their courses, the creation of a specialized Moodle data mining tool significantly simplifies the procedure.

Overall, incorporating data mining techniques in course management systems enables educators to make data-informed decisions, personalize instruction, and improve the overall educational experience.

[21]The study valuable insights for improving the educational system. By applying process mining to event logs, the research identifies both standard and non-standard behavior, offering

a comprehensive understanding of student engagement. The findings highlight the importance of considering the specific context and instructional conditions when interpreting behavior patterns. The study emphasizes the role of instructors in shaping student behavior through quiz settings. Overall, utilizing process mining techniques in analyzing quiz-taking behavior can enhance the educational system by informing the instructional design and support strategies based on student engagement and identifying non-standard behaviors.

[26]In the study, a tool is introduced that helps advisors make decisions by conducting comparative and predictive analysis. In particular, in difficult cases, the research shows that Learning Analytics Dashboard for Academic Advising (LADA) enables advisers to evaluate more scenarios in less time, leading to better-informed decisions. The tool is thought to be useful for making accurate decisions, and further advancements are encouraged.

In order to gain knowledge for focused future development, the study recommends conducting a longitudinal study to compare academic performance between students advised using LADA and traditional methods. LADA enhances academic advising and decision-making processes, which helps to improve the educational system.

[27]In terms of student involvement with the course, the study introduces the Analyse My Blackboard Activities (AMBA) Prototype, a student-centered learning analytics dashboard, and contrasts it with a teacher-centered analytical dashboard. The outcomes reveal how the student-centered dashboard may improve student engagement as students who utilized the AMBA tool had greater levels of Blackboard system engagement, frequented the discussion board, and received higher final grades.

These results underline how important it is to provide students access to their own analytics dashboards. By encouraging students to share their findings with their classmates and facilitating fast access to information, the personalized student-centered dashboard increased engagement. The inclusion of these frameworks in the educational system has the potential to increase student engagement and performance.

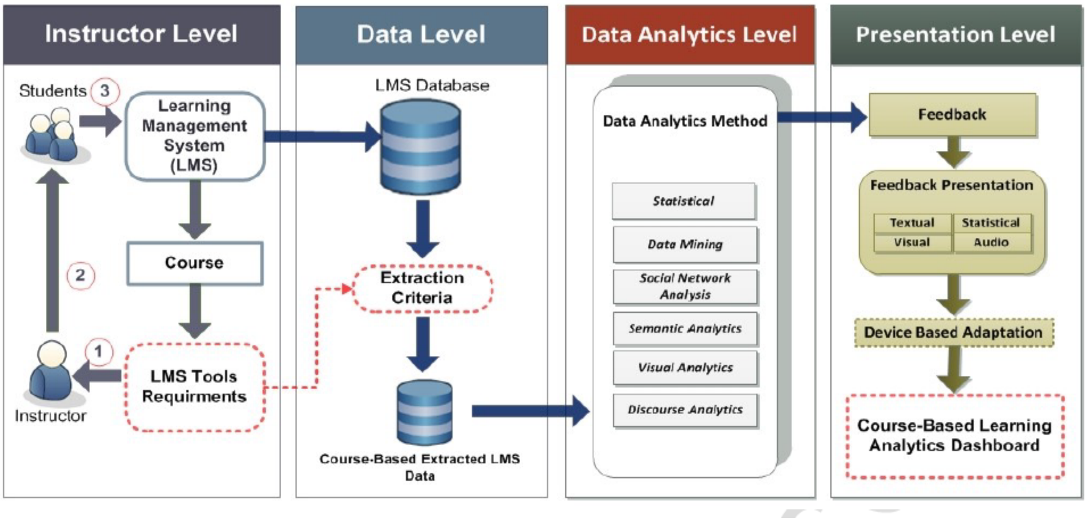

Figure 8: Course Adapted Student Learning Analytics Framework

Conclusion

The literary survey unveiled that process mining possesses the capacity to evaluate patterns of student scoring. By employing process mining, it becomes feasible to discern domains necessitating enhancement within the pedagogical endeavor. Consequently, through the utilization of process mining tools and methodologies, the extraction and visualization of data sourced from a repository of student educational activities can be accomplished. Subsequently, this data can be transformed into an event log, which, upon evaluation, allows for the identification of impediments or irregularities within the process of managing student learning. This valuable discernment empowers educators and administrators to make judicious decisions regarding the optimization of the educational process and the enrichment of student learning outcomes. In summary, the integration of these techniques within higher education holds promising prospects for amplifying the efficacy of the teaching-learning process and heightening the educational journey for students.

Architecture and System Design

Architecture

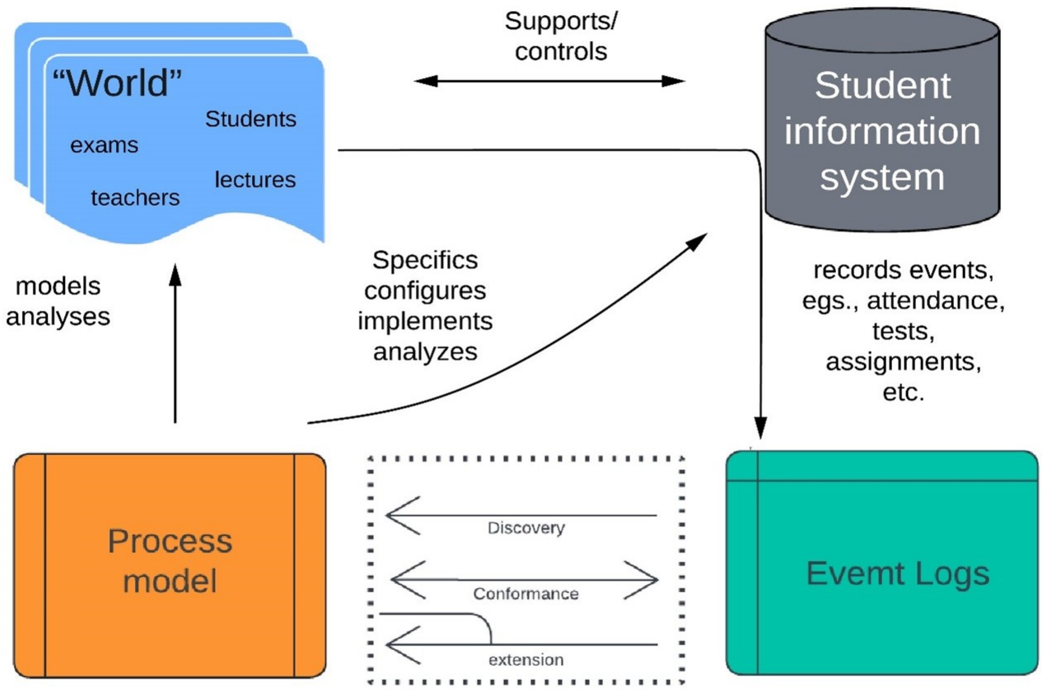

Figure 9: System Architecture

- Data collection: The system architecture for data collection in student performance during process mining involves gathering data from the organization's digital systems. This includes student grades, tests, attendance, and other elements that impact student success. The architecture comprises several components. Firstly, a secure connection is established to the digital systems via Application Programming Interface (API)s or direct integration. Data extraction processes use methods like scraping or querying to get the necessary data. Consistently formatted extracted data allows for processing and analysis. The gathered data is safely stored in an appropriate storage medium, such a database or data warehouse. By locating and fixing mistakes and inconsistencies, data validation and cleaning processes guarantee correctness.

  Privacy and security measures safeguard student information through encryption and access controls. Scalability is considered to handle growing data volumes by using scalable storage solutions and distributed computing frameworks. In summary, the system architecture ensures secure connectivity, data extraction, reliable storage, validation, privacy, security, and scalability. This architecture enables organizations to gather valuable insights into student performance and improve educational outcomes.

- Data preparation: The system architecture for data preparation in the context of process mining involves cleaning and transforming collected data before it can be effectively used. To achieve these objectives, the system architecture should consist of several key components: a robust data preprocessing module to identify and resolve inconsistencies or inaccuracies, a data transformation component to reshape and reformatting the data to align it with the specific requirements of the process mining tool, and a missing data handling component to ensure the overall integrity and completeness of the dataset.

  The system architecture for data preparation in process mining involves a comprehensive approach that includes a data preprocessing module for cleaning and error fixing, a data transformation component for reshaping and reformatting the data, an imputation mechanism for handling missing data, scalability considerations for efficient processing, and privacy and security measures. This architecture should be designed to handle the processing requirements effectively and adhere to relevant data protection regulations and safeguards to protect sensitive information. Organizations may successfully prepare their data for process mining by building a solid architecture, which will allow them to acquire insightful information about their business processes and support reasoned decision-making.

- Process discovery: Process mining techniques are vital in system architecture for process discovery within organizations. They locate and visualize core activities, while event record research and process mining uncover trends. Process models visually illustrate activity relationships and workflow. Visualization tools help stakeholders understand processes, identify bottlenecks, and recognize the potential for change. Process monitoring and analysis approaches discover data trends and patterns.

  The design should be scalable and efficient, with the ability to handle rising data volumes through distributed computing or parallel processing. In summary, the system architecture for process discovery uncovers insights through event record examination, process mining methodologies, and visualisation tools. It contains data analysis, automated process discovery, and trend detection components. Organisations obtain important insights into operational efficiency and effectiveness by utilising this design.

- Analysis and interpretation: The system architecture for analysis and interpretation involves examining and comprehending the identified academic process to identify patterns and tendencies in academic results. This architecture comprises several key components: a data analysis module, grouping and segmentation component, learning process bottleneck identification, and personalized analysis components. The data analysis module uses statistical techniques, data mining algorithms, or machine learning models to uncover patterns, trends, and correlations within the data. The grouping and segmentation component categorizes students based on various attributes such as demographics, academic levels, or performance indicators. The learning process bottleneck identification component uses process mining techniques to uncover inefficiencies, delays, or areas where students encounter difficulties.

The personalized analysis component analyzes the academic data at an individual student level, identifying specific areas where students excel or struggle. The system architecture should also consider scalability and performance requirements.

- Recommendations and action planning: The system architecture for recommendations and action planning involves utilizing the findings of the study to make targeted suggestions for enhancing student performance and optimizing the learning process. The architecture consists of several key components, such as a recommendation engine that analyzes the study findings and generates personalized recommendations for various stakeholders involved in the educational process. Additionally, a decision support system provides educators, administrators, and other stakeholders with the necessary tools and information to implement the recommended actions. Finally, a feedback mechanism ensures continuous evaluation and refinement of the recommendations and action plans. The system architecture for recommendations and action planning involves a recommendation engine, a decision support system, a feedback mechanism, scalability considerations, and data privacy measures.

  This allows for ongoing assessment and adjustment to align with evolving needs and challenges in the educational environment. Scalability and performance considerations are essential in the system architecture to ensure efficient generation and delivery of recommendations. Data privacy and security measures are paramount in the architecture to safeguard student information. Access controls, encryption techniques, and anonymization methods can be implemented to ensure privacy and confidentiality.

System design

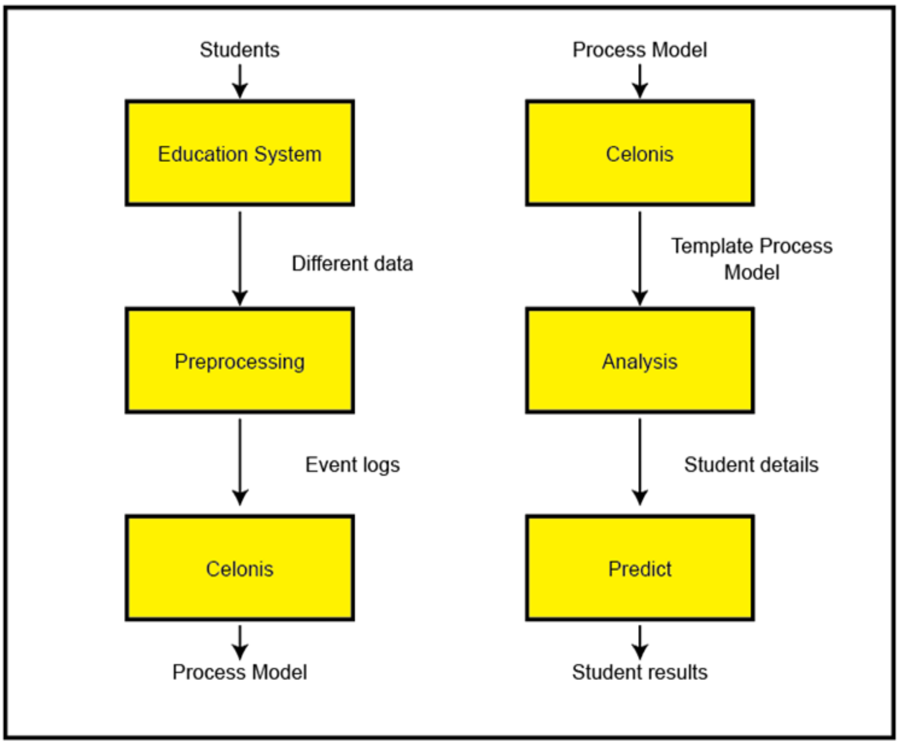

Figure 10: System Design

- Educational System: The key source of data for the process mining system is the educational system. It might be a learning management system, an online learning environment, or any other type of technological instrument used in education to get insights into how students engage with the system. The educational environment creates a vast amount of raw data that must be processed and analyzed.

- Preprocessing: The preprocessing module is responsible for cleaning and preparing the raw data gathered from the educational system for further analysis. This module frequently requires data cleansing, data transformation, and data integration to make sure the data is suitable for process mining. Data cleaning entails removing errors, duplication, or missing data. Data transformation involves converting data from one format to another, whereas data integration incorporates data from various sources to create a single dataset.

- Celonis: Celonis is a well-known process mining program that may be used to analyze and visualize large amounts of data. Celonis is used in the educational process mining system to uncover, monitor, and enhance educational processes by identifying trends and bottlenecks in data. Celonis includes a graphical interface that allows users to visualize and analyze data, as well as support for a variety of process mining techniques such as process discovery, conformance testing, and performance analysis.

To uncover the fundamental process model in Celonis, we use the Alpha Miner technique. The Alpha algorithm, also known as the Alpha-Miner, is a process mining method that is used to recover causality from a set of event sequences. It was proposed for the first time by van der Aalst, Weijters, and Mruter [1]. Alpha Miner's purpose is to turn the event log into a workflow net based on the correlations between the activities in the event log. An event log is a collection of traces, and a trace is a list of activity names.

An event log is a multiset of traces, and a trace is a sequence of activities. Thus, an event log such as the above can be represented using the following notation:

$L1 = [< A, B, C, D >, < A, C, B, D >, < A, E, D >]$

Every event log can be reduced to a collection of traces, which can then be used to dissect relationships between various activities in the process. According to the alpha miner's rules, activities belonging to different cases can have four types of relationships with one another:

– Succession: x $>$ y if and only if some relation x is directly followed by y. In our example, we can consider that A $>$ B, A $>$ E, and A $>$ C.

– Causality: x $\rightarrow$ y iff x $>$ y and not y $>$ x. In our example, we can consider that A $\rightarrow$ E.

– Parallel: x $||$ y iff x $>$ y and y $>$ x. In our example, we have B $||$ C.

– Choice: x # y iff not(x $>$ y) and not(y $>$ x). In our example, we have A # D.

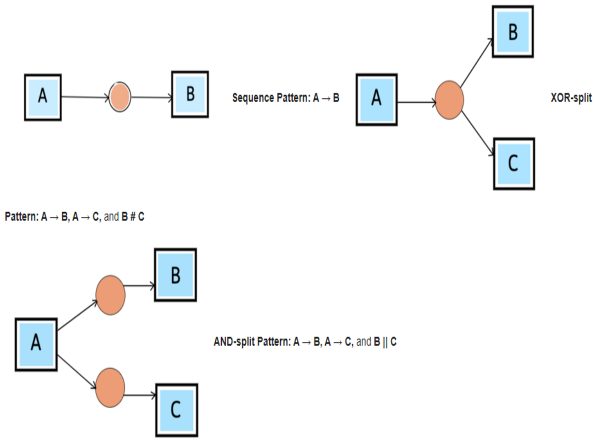

Figure 11: Patterns

A Petri net defining the process model is first constructed by the alpha miner utilizing relations such as directly follows, sequence, parallel, and choice that are extracted from an event log. The algorithm starts by creating a footprint matrix. A process model can be built using the footprint matrix and the pattern illustrated above. A footprint-based matrix is initially discovered using the four previously established relations. Places are discovered using the footprint-based matrix. To keep the number of places low, each location is recognized by a pair of sets of activities.

Matrix of log L1					
.	a	b	c	d	e
a	#	$\rightarrow$	$\rightarrow$	#	$\rightarrow$
b	$\leftarrow$	#	$\|\|$	$\rightarrow$	#
c	$\leftarrow$	$\|\|$	#	$\rightarrow$	#
d	#	$\leftarrow$	$\leftarrow$	#	$\leftarrow$
e	$\leftarrow$	#	#	$\rightarrow$	#

Table 1: Foot Print Matrix

- Y_W is the set of all pairs (A, B) of maximal sets of tasks such that
  - Neither A x A and B x B contain any members of $>$ and
  - A x B is a subset of $\rightarrow$
- P_W contains one place $p_{(A,B)}$ for every member of Y_W ,plus the input place i_W and the output place o_W

The flow relation F_W is the union of the following:

- $\{(a, p_{(A,B)}) | (A, B) \in Y_W \wedge a \in A\}$
- $\{(p_{(A,B)}, b) | (A, B) \in Y_W \wedge b \in B\}$
- $\{(i_W, t) | t \in T_I\}$
- $\{(t, i_O) | t \in T_O\}$

The result is

- a Petri net structure $\alpha(W) = (P_W, T_W, F_W)$
- with one input place i_W and one output place o_W
- because every transition of T_W is on a F_W-path from i_W to o_W, it is indeed a workflow net.

For the example given above, the following petri net would result from the application of an alpha miner.

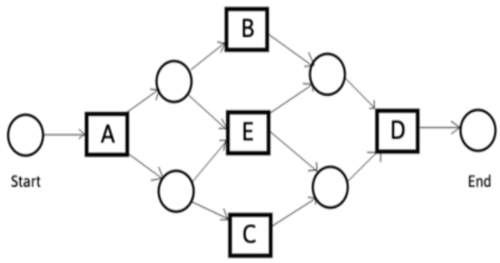

Figure 12: Petri Net

- Analysis: To find patterns and trends in the data, the analysis module employs a variety of statistical and machine-learning techniques. In this module, descriptive and exploratory data analysis methods are typically used to understand the data before more sophisticated methods like clustering, classification, and regression analysis are applied to create predictive models. The analysis module can assist in identifying elements like learning style, engagement, and motivation that influence student performance.

  We use conformance checking to see if the given detail will yield the desired results. Conformance checking is a class of process mining techniques that compares a process model to an event log of the same process.[1] It is employed to determine whether the model accurately represents how a business process will actually be carried out and vice versa as recorded in the event log.

For instance, a process model could indicate that purchase orders exceeding one million euros necessitate two inspections. The event log provides evidence regarding compliance with this rule.

Token-based replay is a method that utilizes four counters. These counters include generated tokens, consumed tokens, missing tokens, and remaining tokens. The purpose of these counters is to assess the fitness of an observation trace in relation to a specific process model represented in Petri-net notation.

These four counters monitor the token status during the replay of a Petri net trace. When a transition creates a token, the total number of tokens created grows by one. Similarly, when a token is spent to initiate a transition, the total number of consumed tokens rises by one. When a token is expected but not present to facilitate a transition, the count of missing tokens grows by one. The remaining tokens counter indicates the total amount of tokens left after replaying the trace. Only if there are no missing tokens during the replay and no lingering tokens at the end is the trace regarded to comply to the process model.

The fitness between an event log and a process model is computed as follows:

$$\frac{1}{2}\left(1-\frac{m}{c}\right)+\frac{1}{2}\left(1-\frac{r}{p}\right)$$

where m is the number of missing tokens, c is the number of consumed tokens, r is the number of remaining tokens, and p is the number of produced tokens.

- Predict: The prediction module is utilized to anticipate forthcoming events by analyzing past data. This module commonly entails constructing predictive models through the application of machine learning algorithms, aiming to identify patterns and trends within the data. These predictive models enable the forecasting of student performance, identification of students at risk, and provision of recommendations for interventions to enhance learning outcomes. Additionally, the prediction module can contribute to the optimization of educational processes by pinpointing areas in need of improvement and predicting the potential impact of proposed modifications.

Implementation

Implementation Platform

Hardware

1. Processor: Intel(R) Core(TM) i5-9300H CPU @ 2.40GHz 2.40 GHz
2. RAM: 8.00 GB DDR4 3533 MT/s
3. GPU: Nvidia GeForce GTX 1650 8GB
4. Storage platform: NVME SSD

Software

1. Operating System: Windows 11 64bit
2. Software Used: Jupyter, celonis
3. Programming Languages: Python 3, PQL
4. Storage platform: Excel, Comma-Separated Values (CSV), eXtensible Event Stream (XES), Petri Net Markup Language (PNML)

Implementation details

Organization of files

- Attendance File:
  - Student ID: Unique identifier for each student.
  - Dates: Dates of the attendance record.
  - Attendance Status: Indicates the attendance status of the student on the given date (e.g., present, absent, late).
  - Percentage: indicates the percentage of attendance the students has in that particular period
- Marks File:
  - Student ID: Unique identifier for each student.
  - Assessment Type: Type of assessment (e.g., exam, assignment, project, internal, SEE, etc.).
  - Marks: Numeric score or percentage obtained by the student for the specific assessment.
- Other Metadata File:
  - Student ID: Unique identifier for each student.
  - Additional Metadata: Any relevant information about the student that may be useful for analysis (e.g., examination dates, class periods demographic data, study habits, extracurricular activities).
- Event Logs File:
  - Case ID: Unique identifier for each student or educational process instance.
  - Activity: Description of the activity or event (e.g., attending a lecture, submitting an assignment).
  - Timestamp: Date and time when the activity occurred.
  - Additional Attributes: Any additional information related to the activity, such as subject, assessment type, or outcome.

- Process Model File (XES or PNML):
  - Process Model: Represents the discovered process model, capturing the sequence of activities and dependencies within the educational process.
  - Nodes: Describes the different activities or events in the process.
  - Transitions: Specifies the flow of the process from one activity to another.
  - Process Model Format: The file format can be XES or PNML.
- Predictions File:
  - Student ID: Unique identifier for each student.
  - Predicted Score: Estimated or predicted score for the student based on their current attendance and marks, using the discovered process model as a template.

Implementation workflow

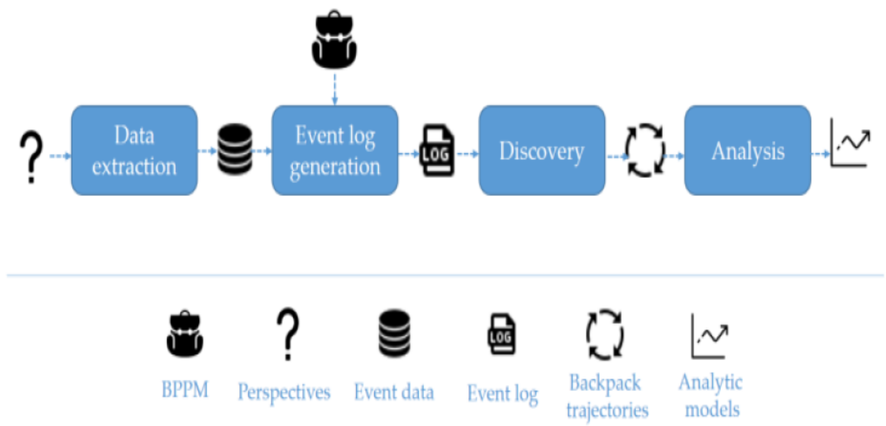

Figure 13: Data flow Overview

- Data Collection: Collect data on attendance, marks, and other relevant metadata in the form of Excel/CSV files. Ensure that the data includes unique identifiers for each student, dates for attendance records, and subject-specific information for marks.

- Preprocessing and Event Log Creation: Preprocess the collected data to clean and transform it into a suitable format for process mining analysis. Convert the preprocessed data into event logs that capture the sequence of student activities and interactions within the educational process.

- Process Discovery: Apply process discovery algorithms to the event logs to generate process models in the form of XES or PNML files. The process models represent the flow and dependencies of activities within the educational process.

- Common Trace Analysis: Analyze the discovered process models to identify the most common trace or the trace that provides the most valuable insights. By examining the most frequent path or sequence of activities, gain insights into patterns and bottlenecks that may impact student scoring.

- Score Prediction: Utilize the discovered process model as a template to predict future scores for other students based on their current attendance and marks. Apply the process model's patterns and dependencies to estimate or predict the scores of students who follow similar activity sequences.

DataSet

The dataset for analyzing students' scoring patterns using Process Mining is a meticulously curated collection of student academic records from various educational institutions. It includes anonymized data on student demographics, course details, assessment scores, grading criteria, and timestamps. The dataset covers a wide range of subjects, academic levels, and assessment types to ensure its applicability across different scenarios.

To construct the dataset, we obtained permission from educational institutions and adhered to strict ethical guidelines to protect student privacy. The data was carefully cleaned and preprocessed, eliminating incomplete or inconsistent records while addressing missing values and outliers. Event logs capturing the sequence of student activities and interactions within the learning environment were also incorporated to enable Process Mining analysis.

The dataset offers insights into diverse scoring patterns, highlighting variations in student performance across subjects, assessment types, and periods. Its comprehensive nature allows for the identification of trends and factors influencing student scores. By adhering to ethical considerations and ensuring data security, this dataset provides valuable information for educational stakeholders, supporting improved decision-making processes and educational outcomes.

Experiments and results

Experimentation

The experiment aims to utilize process mining techniques to analyze students' scoring patterns based on attendance data and marks data. By converting these datasets into event logs and leveraging the capabilities of Celonis, the most common/happy path can be discovered and visualized. This path will serve as a reference to better understand students' scoring patterns. Additionally, the fitness of an individual student's scores concerning the common/happy path will be evaluated, ultimately determining the student's future score. Experimental Procedure:

- Data Collection: Attendance Data: Attendance records of students are collected and stored in an Excel sheet.

usn	11/12/2022	11/13/2022	11/14/2022	11/15/2022	11/16/2022	....	month1
1DS19CS001	1	1	1	1	1	....	80
1DS19CS002	1	1	1	1	1	....	100
1DS19CS003	1	1	1	0	0	....	60
1DS19CS004	1	1	1	1	1	....	80
1DS19CS005	1	1	0	1	1	....	80
1DS19CS006	1	0	1	0	0	....	40
1DS19CS007	1	0	0	0	0	....	40
1DS19CS008	1	1	1	1	1	....	80
1DS19CS009	1	0	0	1	0	....	50
1DS19CS010	1	0	1	1	1	....	70
1DS19CS011	1	1	1	1	0	....	70
1DS19CS012	0	1	1	1	1	....	90

Table 2: Sample of Attendance Excel sheet

- Marks Data: Marks obtained by students in various assessments are recorded and stored in a separate Excel sheet. A sample marks sheet is shown in Table 2.

- Conversion to Event Log: The attendance data and marks data are transformed into event logs, which capture the chronological order of events.A sample event log is shown in table 3.

  Both Excel sheets are processed and converted into a unified format, such as a CSV file, suitable for further analysis.

- Process Discovery and Visualization: The CSV file containing the event log is fed into Celonis, a process mining software.

USN	CIE-1	CIE-2	CIE-3	CIEs	ASST	QUIZ	INTERNALS	SEE	FINAL
1DS19CS001	2.4	10	2.2	14.6	6	4	24.6	43	68
1DS19CS002	5.8	9.4	9.8	25	8	8	41	26	67
1DS19CS003	4.8	7.4	2.8	15	8	4	27	42	69
1DS19CS004	5.6	8.4	4	18	6	7	31	21	52
1DS19CS005	9.2	6	9.4	24.6	8	9	41.6	39	81
1DS19CS006	4.6	7.6	7.6	19.8	1	9	29.8	41	71
1DS19CS007	5.6	3.6	7.6	16.8	1	6	23.8	42	66
1DS19CS008	9.8	2.2	4.2	16.2	3	3	22.2	47	69
1DS19CS009	8.8	8.8	3.2	20.8	3	9	32.8	13	46
1DS19CS010	7	8.4	5.6	21	3	2	26	20	46
1DS19CS011	7.8	3.4	6	17.2	5	3	25.2	14	39
1DS19CS012	7.8	5.4	6.8	20	4	9	33	40	73

Table 3: Sample of Marks Excel sheet

USN	Date	Type	Specific	Value	Vname	SV
1DS19CS001	11/21/2022 0:00	attendance	Month1	75	Month1G75	G75
1DS19CS001	11/23/2022 0:00	marks	CIE-1	40	CIE-1L50	L50
1DS19CS001	12/7/2022 0:00	attendance	Month2	50	Month2G50	G50
1DS19CS001	12/9/2022 0:00	marks	CIE-2	90	CIE-2G90	G90
1DS19CS001	12/13/2022 0:00	marks	asst	50	asstG50	G50
1DS19CS001	12/15/2022 0:00	marks	quiz	40	quizL50	L50
1DS19CS001	12/23/2022 0:00	attendance	Month3	75	Month3G75	G75
1DS19CS001	12/25/2022 0:00	marks	CIE-3	40	CIE-3L50	L50
1DS19CS001	12/27/2022 0:00	marks	internals	75	internalsG75	G75
1DS19CS001	1/7/2023 0:00	marks	see	75	seeG75	G75
1DS19CS001	1/17/2023 0:00	marks	finals	50	finalsG50	G50

Table 4: Sample of Events Log

Celonis performs process discovery, which involves extracting process models from the event log to identify the most common/happy path.

The most common/happy path represents the sequence of events or steps that students typically follow in achieving higher scores.

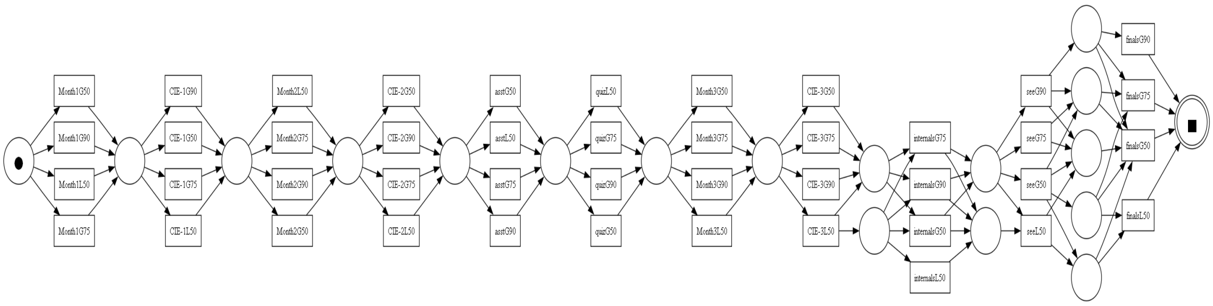

Figure 14: Whole Process Model

- Analysis of Scoring Patterns: The discovered most common/happy path is thoroughly examined to gain insights into the patterns that lead to higher scores.

  Various process mining techniques, such as process visualization and performance analysis, are applied to identify bottlenecks, deviations, and potential improvements.

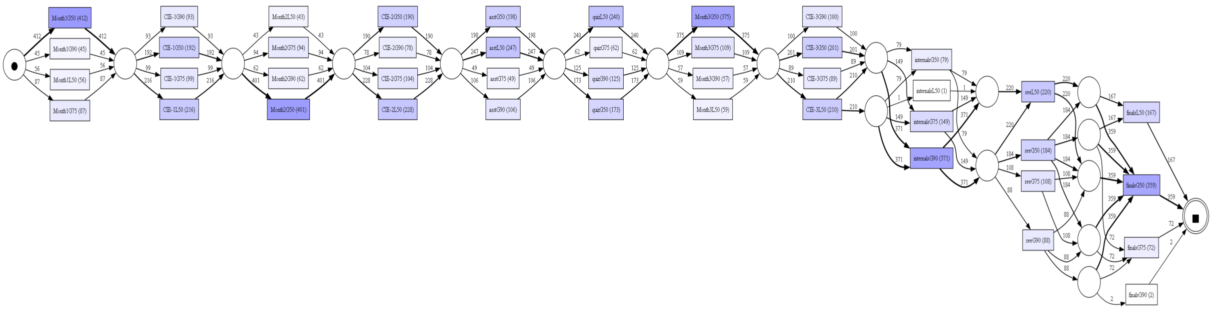

Figure 15: Common/Happy Path

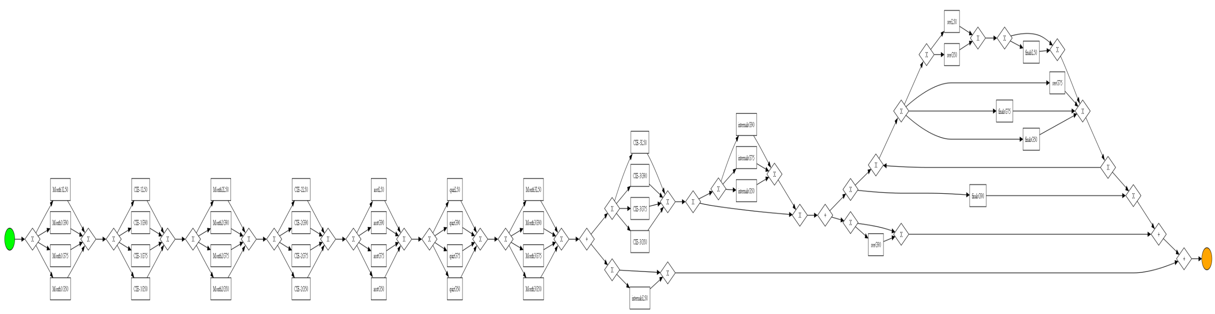

Figure 16: BPMN Inductive Patri-Net

- Fitness Evaluation: Individual student data, including attendance records and previous marks, are selected for analysis.

USN	Date	Value	Vname	SV
1DS19CS001	11/21/2022 0:00	50	Month1G50	G50
1DS19CS001	11/23/2022 0:00	90	CIE-1G90	G90
1DS19CS001	12/7/2022 0:00	50	Month2G50	G50
1DS19CS001	12/9/2022 0:00	50	CIE-2G50	G50
1DS19CS001	12/13/2022 0:00	40	asstL50	L50
1DS19CS001	12/15/2022 0:00	50	quizG50	G50
1DS19CS001	12/23/2022 0:00	50	Month3G50	G50

Table 5: Data of a Student

The fitness of a student's scores concerning the common/happy path is determined by comparing their actual scoring patterns against the established path.

fitness value is calculated based on the level of conformity to the common/happy path.

```
[{'alignment': [('Month1G50', 'Month1G50'),
   ('CIE-1G90', 'CIE-1G90'),
   ('Month2G50', 'Month2G50'),
   ('CIE-2G50', 'CIE-2G50'),
   ('asstL50', 'asstL50'),
   ('quizG50', 'quizG50'),
   ('Month3G50', 'Month3G50'),
   ('>>', None),
   ('>>', 'CIE-3L50'),
   ('>>', None),
   ('>>', None),
   ('>>', 'finalsG90'),
   ('>>', None),
   ('>>', None),
   ('>>', None)],
  'cost': 20006,
  'visited_states': 15,
  'queued_states': 71,
  'traversed_arcs': 71,
  'lp_solved': 9,
  'fitness': 0.875,
  'bwc': 160006}]
```

Figure 17: Fitness Diagnostics

- Prediction of Future Score: The fitness value obtained from the analysis is used to predict the student's future score.

  Higher fitness values indicate a stronger alignment with the common/happy path, suggesting a higher likelihood of achieving better scores in future assessments.

Results

Possible Results:

- Identification of Most Common/Happy Path:
  - The process discovery analysis using Celonis reveals a well-defined most common/happy path followed by a significant number of students.
  - The path highlights the sequence of events or steps that lead to higher scores, providing valuable insights into effective strategies for academic success.
- Visualization of Scoring Patterns:
  - The process visualization capabilities of Celonis showcase the different paths and variations taken by students in achieving their scores.
  - Patterns such as frequent attendance, specific study habits, or timely submission of assignments may emerge, indicating correlations with higher scores.
- Identification of Deviations:
  - Deviations from the most common/happy path may signify areas where students face challenges or experience lower performance.
- Fitness Evaluation:
  - The fitness evaluation of individual students' scoring patterns against the most common/happy path yields a range of fitness values.
  - Higher fitness values indicate students whose scoring patterns closely align with the established path, suggesting a higher chance of future success.
  - Lower fitness values may indicate areas where students need additional support or interventions to improve their scores.
- Prediction of Future Scores:
  - Based on the fitness values obtained, predictions for future scores of individual students can be made.
  - Students with higher fitness values are likely to maintain or improve their scores, while those with lower fitness values may require targeted interventions or academic support to enhance their performance.

- Insights for Academic Interventions:
  - The results provide valuable insights into the effectiveness of different academic interventions or strategies.
  - By analyzing the scoring patterns of students who significantly deviate from the most common/happy path, specific interventions can be identified to address their unique challenges and improve their scores.

It's important to note that the specific results of the experiment will depend on the unique characteristics of the student dataset and the analysis conducted using process mining techniques.

Conclusion

In conclusion, the experiment utilizing process mining techniques to analyze students’ scoring patterns has provided valuable insights into factors contributing to academic success. The most common/happy path, identified through Celonis, serves as a reference for effective strategies. Visualization of scoring patterns and identification of bottlenecks and deviations enhance our understanding of variations in student approaches.

Fitness evaluation enables score prediction, with higher fitness values indicating the potential for maintaining or improving scores. The results inform academic interventions and system improvements. Tailored interventions can address deviations from the path, while personalized recommendations guide students toward success.

This experiment demonstrates the effectiveness of process mining in analyzing scoring patterns and supporting educational stakeholders. By leveraging data-driven analysis, educators can understand student performance and implement targeted interventions to support academic growth.

References

[1] Romero, Cristóbal Ventura, Sebastian. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications. 33. 135-146. 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 personalized 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 behavior. 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.