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-560078

Submitted By

Vonkayala Ashwini 1DS19CS192

Pallapothu Lakshmi Sharanya 1DS19CS730

Srungarapu Sai Sri Nandan 1DS19CS172

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 - 560078

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-560078

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 **Vonkayala Ashwini**[**1DS19CS192**], **Pallapothu Lakshmi Sharanya** [**1DS19CS730**], **Srungarapu Sai Sri Nandan** [**1DS19CS172**] 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.

Vonkayala Ashwini 1DS19CS192
Pallapothu Lakshmi Sharanya 1DS19CS730
Srungarapu Sai Sri Nandan 1DS19CS172
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

1 Introduction

1.1 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. Knowledge discovery in databases (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

1.2 Real World Application

1.2.1 Personalized Learning

Educators may discover important information about unique learning preferences and behaviors by analyzing student data using data mining tools. For instance, they may monitor how students interact with learning platforms, how they move through online resources, and which resources are most useful to them. With the use of this information, educators may modify the learning environment to suit the individual requirements and preferences of each student. They can tailor the educational material to the student's level of skill, make personalised recommendations for supplemental resources, and give specific comments regarding the student's performance. This degree of personalisation improves student motivation, engagement, and overall learning results.

1.2.2 Early Intervention and Support

By analyzing student data, such as completion rates, interaction patterns, and assessment results, educators can identify students who may be at risk of falling behind or disengaging. For instance, if a student consistently struggles to complete assignments or shows a decline in participation, educational process mining can raise red flags. With this information, educators can intervene proactively, providing additional support, targeted interventions, or personalized remediation strategies. By addressing issues at an early stage, educators can prevent further academic difficulties and promote student success.

1.2.3 Curriculum Design and Improvement

Education professionals can evaluate the efficiency of the curriculum, teaching approaches, and evaluation techniques by looking at student data. They can look at how students engage with the course contents, pinpoint their strengths and weaknesses, and assess the effects of various teaching methods on students' performance. Through the use of data, educators may improve curriculum, optimize teaching materials, and create assessments that are in line with student learning objectives. As a consequence, educational institutions may improve the standard and applicability of their programs, making sure that students have the most efficient and interesting learning opportunities.

1.3 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.

2 Problem Statement and Proposed Solution

2.1 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.

2.2 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.

2.2.1 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.

Although these strategies have improved the educational system, there is still potential for improvement. Educational process mining, which combines sophisticated data mining and machine learning methods, has a lot of potential. Educational process mining can offer deeper insights into individual learning needs, optimize curriculum design, and support decision-making based on evidence by analyzing huge volumes of educational data, including student interactions, performance, and engagement. By guaranteeing that every student has a customized, efficient, and engaging learning experience, this all-encompassing and data-driven strategy has the potential to revolutionize the educational system.

2.3 Proposed Solution

Educational process mining technology and techniques offer a promising solution for improving the education system by providing valuable insights and enabling more efficient processes. By leveraging event log data, educational process mining allows for a thorough analysis of the functionality of various processes within the system. This analysis can help identify shortcomings, bottlenecks, and loops that hinder the overall effectiveness of the education system. Building upon this foundation, the next crucial step is to develop a model that generates a more efficient method for implementing these processes.

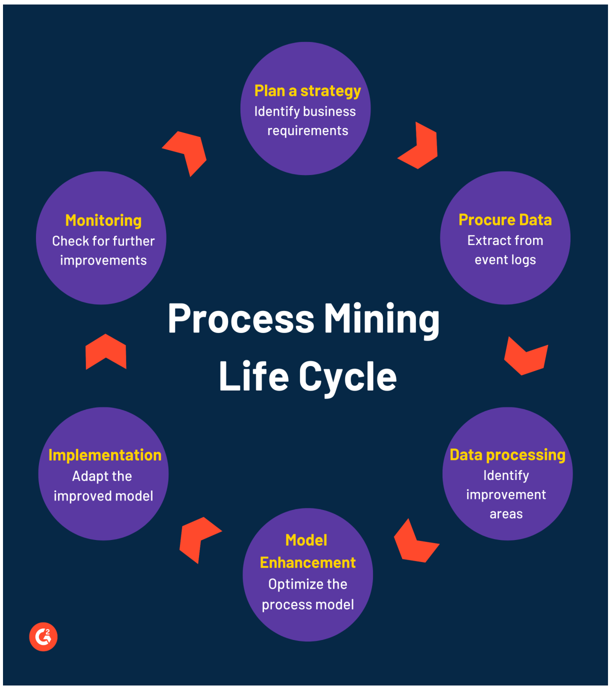


Figure 2: Process mining - Life cycle

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

- Data Collection and Integration:Collect pertinent information from the educational system's many different sources, including student records, learning management systems, and test results. Information on interactions, participation, performance, and other pertinent factors involving student interactions should be included in this data. For a thorough analysis, this different dataset must be integrated.

- Process Analysis and Visualization: Apply educational process mining techniques to analyze the collected data and gain insights into the existing processes within the education system.

Visualize these processes using process flow diagrams or process maps to understand the sequence of activities, decision points, and potential bottlenecks.

- Identify Improvement Opportunities: Analyze the process flow diagrams and identify areas where improvements can be made. This may involve pinpointing bottlenecks that slow down the system, detecting loops or redundancies that hinder progress, or identifying gaps in the learning journey. These insights help prioritize areas for improvement and guide the development of more efficient processes.

- Model Design and Simulation: Create a new model using the analysis's learnings to put improved procedures into practice. The efficiency and effectiveness of the educational system should be increased using this model, which should make use of technology, automation, and personalized learning strategies. Before usage, the model can be tested and improved using simulation approaches.

- Implementation and Monitoring: Implement the new model in a controlled setting while carefully observing the effects it has on student performance, engagement, and system performance as a whole. Continually gather data to gauge the success of the new procedures and make required corrections.

- Continuous Improvement: The analysis and improvement of the implemented processes should be a regular practice of educational process mining. Educators and policymakers can spot developing problems, new trends, and additional areas for development by gathering data and periodically analyzing it.

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.

3 Literature Survey

3.1 Use of tools

3.1.1 A Digital Twin Framework For Analysing Students' Behaviours Using Educational Process Mining

Authors: Ambrose Azeta, Frank Agono, Adesola Falade, ea Azeta, Vivian Nwaocha

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.

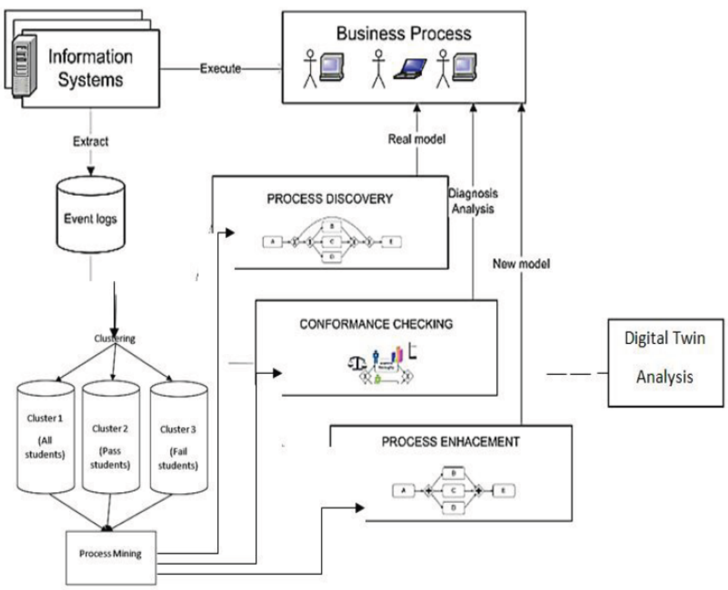


Figure 3: Overview of proposed Digital Twin Framework

3.1.2 Data mining in course management systems: Moodle case study and tutorial

Authors: Cristóbal Romero, Sebastián Ventura, Enrique García

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.

3.1.3 Using process mining to analyze students' quiz-taking behavior patterns in a learning management system

Authors: Juhaňák, Libor, Zounek, Jiří, Rohlíková, Lucie.

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.

3.2 Custom built software

3.2.1 LADA: A learning analytics dashboard for academic advising

Authors: Francisco Gutierrez, Karsten Seipp, Xavier Ochoa, Katherine Chiluiza, Tinne De Laet, Katrien Verbert

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 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.

3.2.2 An Integrated Framework for Course Adapted Student Learning Analytics Dashboard

Authors: Naif Radi Aljohani, Ali Daud, Rabeeh Ayaz Abbasi, Jalal S. Alowibdi, Mohammad Basheri, Muhammad Ahtisham Aslam

In terms of student involvement with the course, the study introduces the 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.

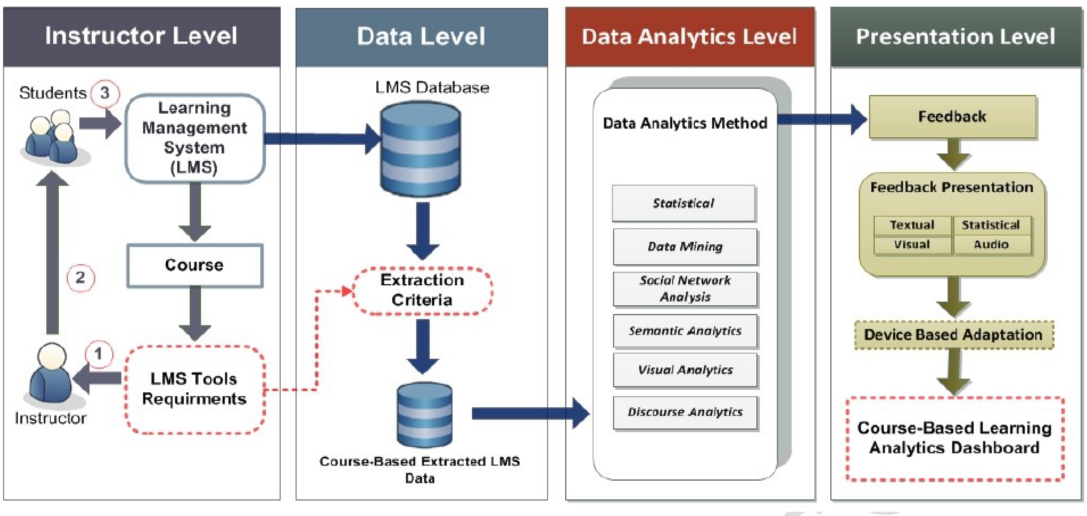


Figure 4: Course Adapted Student Learning Analytics Framework

4 Architecture and System Design

4.1 Architecture

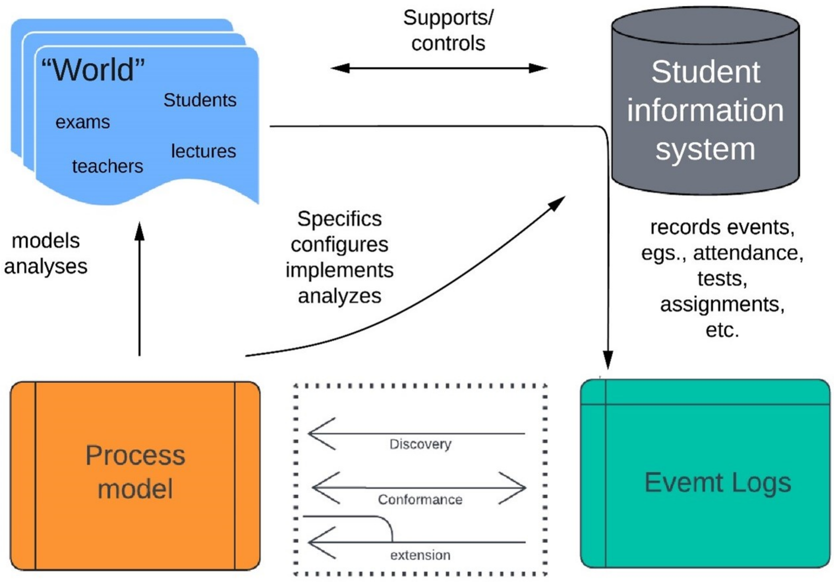


Figure 5: System Architecture

- Data collection: Initially, we gather data from the organization's digital systems that are related to student performance throughout the process mining process. Data about student grades, tests, attendance, and other elements that may affect student success might be included in this.

- Data preparation: After the data has been collected, it might need to be cleaned and transformed before it can be used for process mining. In order to do this, the data may need to be organized in a way that is compatible with the process mining tool of choice and may also need to have errors fixed and missing data filled in.

- Process discovery: Utilising process mining techniques, we find and visualise the fundamental activities that occur inside the organisation once the data has been produced. To find trends in the data, it may be necessary to employ strategies like event record investigation, process mining methods, and process models.

- Analysis and interpretation: Once the process has been identified, researchers may analyze and comprehend it to identify patterns and tendencies in academic results. This may entail contrasting the performance of various student groups, locating learning process bottlenecks, and pinpointing the areas of particular students' strengths and weaknesses.

- Recommendations and action planning: Based on the findings of the study, suggestions for enhancing student performance and optimizing the learning process may be made. This might include making suggestions for modifications to teaching tactics or interventions, as well as recognizing possibilities to expedite the learning process.

4.2 System design

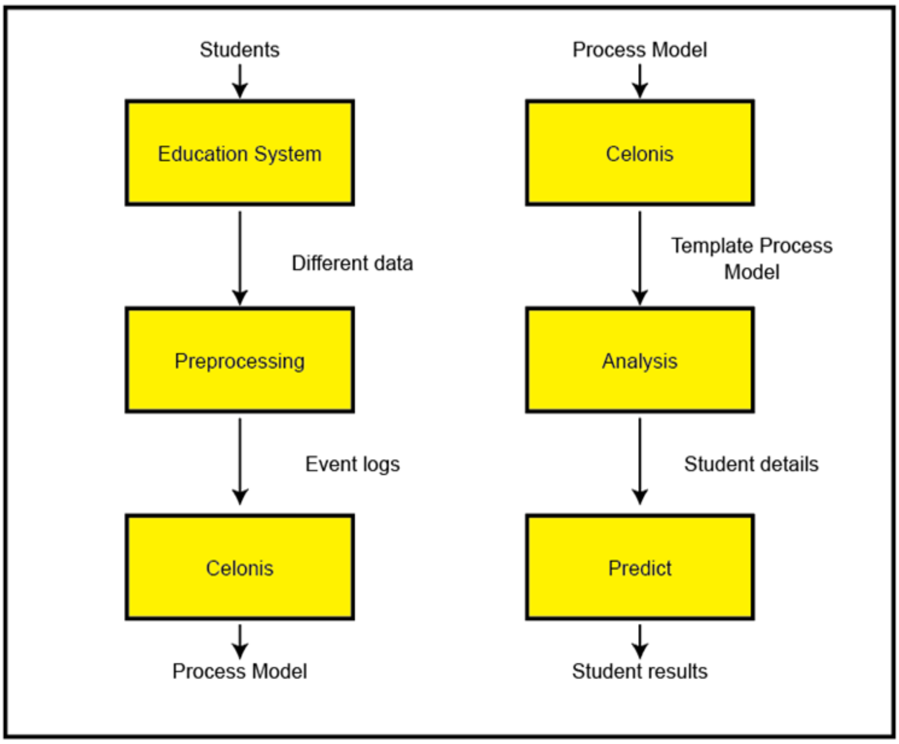


Figure 6: System Design

- 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 raw data collected from the educational system must be cleaned and prepared for further analysis by the preprocessing module. For the purpose of ensuring that the data is appropriate for process mining, this module often involves data cleansing, data transformation, and data integration. Eliminating mistakes, duplicates, or incomplete data is known as data cleaning. While data integration combines data from several sources to produce a single dataset, data transformation entails changing data from one format to another.

- 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 → y iff x > y and not y > x. In our example, we can consider that A → 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.

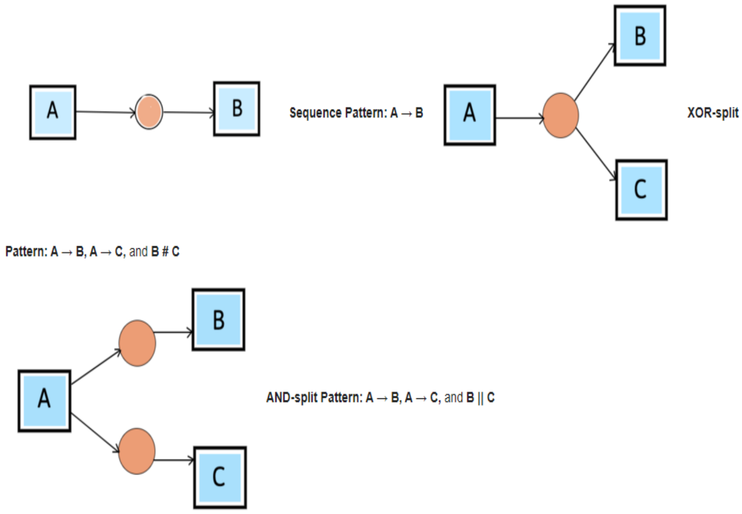


Figure 7: Patterns

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.

Footprint matrix for the log L1

	a	b	c	d	e
a	#	→	→	#	→
b	←	#	||	→	#
c	←	||	#	→	#
d	#	←	←	#	←
e	←	#	#	→	#

Figure 8: 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.

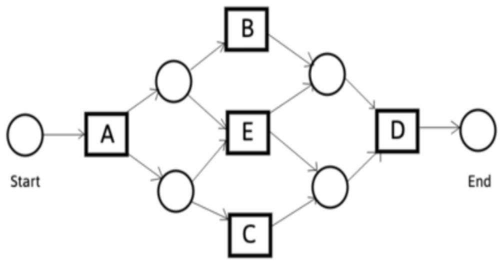


Figure 9: Petri Net

- 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.

5 Implementation

5.1 Implementation Platform

5.1.1 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

5.1.2 Software

1. Operating System: Windows 11 64bit
2. Software Used: Jupyter, celonis
3. Programming Languages: Python 3, PQL
4. Storage platform: Excel, CSV, XES, PNML

5.2 Implementation details

5.2.1 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/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 (eXtensible Event Stream) or PNML (Petri Net Markup Language).

- 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.

5.2.2 Implementation workflow

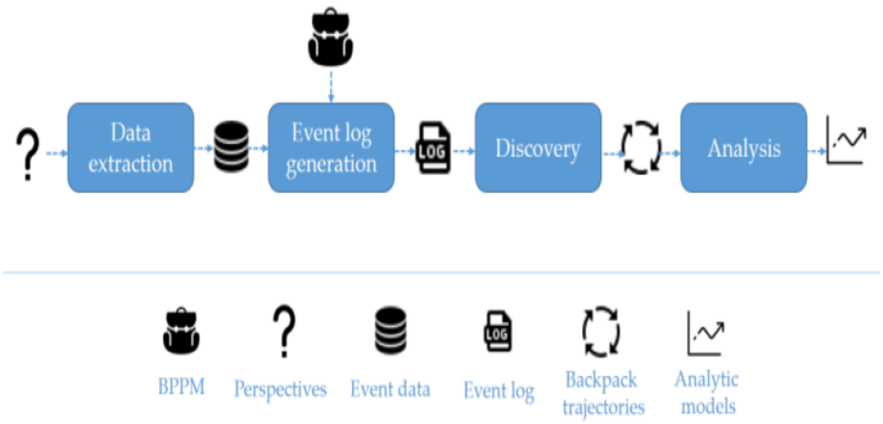


Figure 10: Data flow Overview

- 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 (eXtensible Event Stream) or PNML (Petri Net Markup Language) 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.

5.3 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.

6 Experiments and results

6.1 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.

	A	B	C	D	E	F	G	H	I	J	K	L
1	usn	2022-11-12	2022-11-13	2022-11-14	2022-11-15	2022-11-16	2022-11-17	2022-11-18	2022-11-19	2022-11-20	2022-11-21	month1
2	1DS19CS001	1	1	1	0	0	1	1	0	1	0	60
3	1DS19CS002	0	0	0	0	0	0	1	0	0	0	10
4	1DS19CS003	1	0	0	0	1	1	0	1	0	1	50
5	1DS19CS004	1	1	1	0	1	1	0	1	0	1	70
6	1DS19CS005	0	1	0	1	1	1	1	1	0	0	60
7	1DS19CS006	1	1	0	0	1	0	1	1	1	1	70
8	1DS19CS007	0	1	1	0	1	0	1	0	0	1	50
9	1DS19CS008	1	1	1	0	0	0	0	1	1	0	50
10	1DS19CS009	0	1	1	1	1	0	0	1	0	0	50
11	1DS19CS010	0	1	1	0	1	1	0	0	1	0	50
12	1DS19CS011	0	0	1	1	1	1	1	0	0	0	50
13	1DS19CS012	0	1	1	1	1	0	0	0	0	1	50

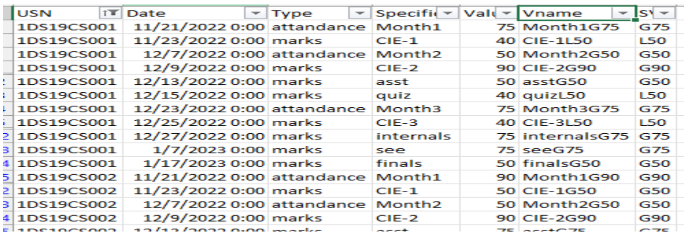
Figure 11: Attendance excel sheet

Marks Data: Marks obtained by students in various assessments are recorded and stored in a separate Excel sheet.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	USN	CIE-1 (50)	CIE-1(10)	CIE-2 (50)	CIE-2(10)	CIE-3 (50)	CIE-3(10)	CIE TOTAL	ASSIGNMENT	QUIZ	INTERNALS MARKS	SEE	FINAL MARKS
2	1DS19CS001	40	8	5	1	5	1	10	9	6	25	15	40
3	1DS19CS002	5	1	5	1	25	5	7	9	2	18	30	48
4	1DS19CS003	5	1	15	3	15	3	7	2	10	19	35	54
5	1DS19CS004	15	3	45	9	10	2	14	5	7	26	15	41
6	1DS19CS005	15	3	35	7	15	3	13	2	4	19	35	54
7	1DS19CS006	20	4	20	4	25	5	13	6	3	22	20	42
8	1DS19CS007	30	6	20	4	40	8	18	10	0	28	10	38
9	1DS19CS008	35	7	15	3	5	1	11	5	8	24	35	59
10	1DS19CS009	20	4	40	8	35	7	19	10	7	36	35	71
11	1DS19CS010	0	0	25	5	50	10	15	2	4	21	25	46
12	1DS19CS011	40	8	35	7	30	6	21	9	3	33	5	38
13	1DS19CS012	30	6	5	1	10	2	9	4	1	14	45	59
14	1DS19CS013	40	8	45	9	5	1	18	1	9	28	30	58
15	1DS19CS014	35	7	0	0	50	10	17	6	1	24	10	34

Figure 12: Marks excel sheet

- Conversion to Event Log: The attendance data and marks data are transformed into event logs, which capture the chronological order of events.



USN	Date	Type	Specifi	Valu	Vname	SY
1DS19CS001	11/21/2022 0:00	attandance	Month1	75	Month1G75	G75
1DS19CS001	11/23/2022 0:00	marks	CIE-1	40	CIE-1L50	L50
1DS19CS001	12/7/2022 0:00	attandance	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	attandance	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
1DS19CS002	11/21/2022 0:00	attandance	Month1	90	Month1G90	G90
1DS19CS002	11/23/2022 0:00	marks	CIE-1	50	CIE-1G50	G50
1DS19CS002	12/7/2022 0:00	attandance	Month2	50	Month2G50	G50
1DS19CS002	12/9/2022 0:00	marks	CIE-2	90	CIE-2G90	G90

Figure 13: Event log

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.

  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.

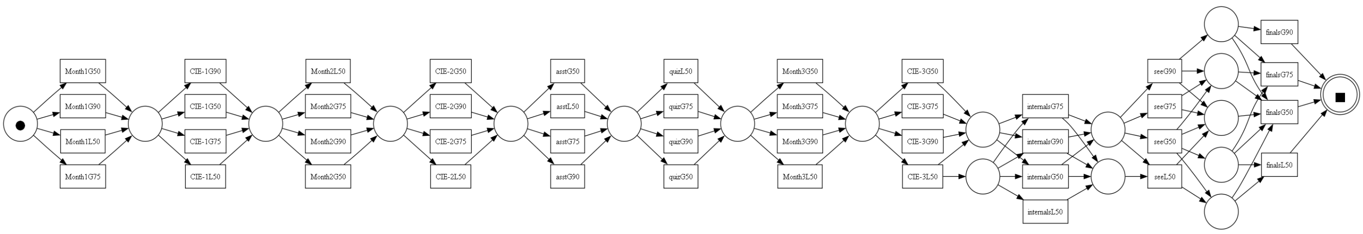


Figure 14: Whole process model

- 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.

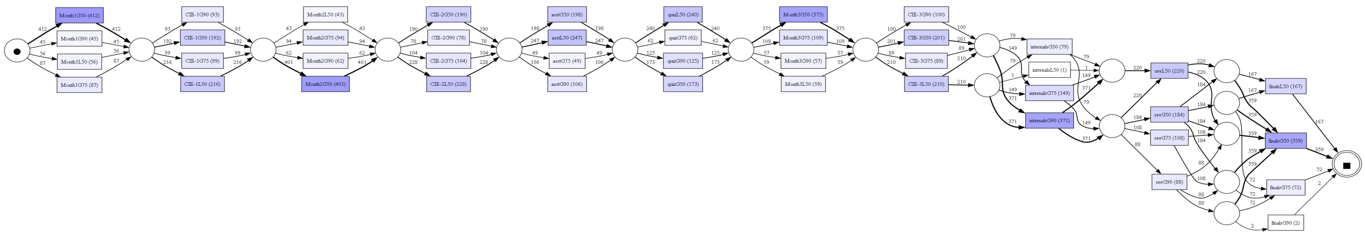


Figure 15: Common/happy path

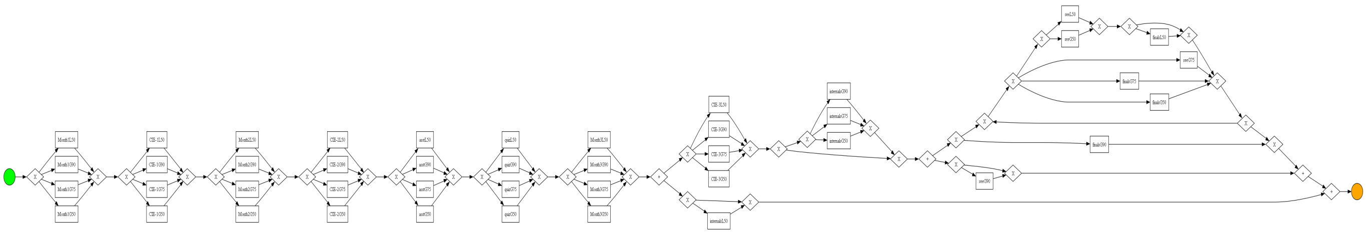


Figure 16: BPMN inductive patri-net

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

	USN	Date	Value	Vname	SV
0	1DS19CS001	11/21/2022 0:00	50	Month1G50	G50
1	1DS19CS001	11/23/2022 0:00	90	CIE-1G90	G90
2	1DS19CS001	12/7/2022 0:00	50	Month2G50	G50
3	1DS19CS001	12/9/2022 0:00	50	CIE-2G50	G50
4	1DS19CS001	12/13/2022 0:00	40	asstL50	L50
5	1DS19CS001	12/15/2022 0:00	50	quizG50	G50
6	1DS19CS001	12/23/2022 0:00	50	Month3G50	G50
7	1DS19CS001	12/25/2022 0:00	40	CIE-3L50	L50
8	1DS19CS001	12/27/2022 0:00	90	internalsG90	G90
9	1DS19CS001	1/7/2023 0:00	50	seeG50	G50
10	1DS19CS001	1/17/2023 0:00	50	finalsG50	G50

Figure 17: Individual Student Data

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.

- 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.

6.2 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.

7 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.