A Study About Process Mining - Student Enrollment Event Log

Research Paper *for* University Assessment - April 2024

DOI: 10.2139/ssrn.335822

5 authors:

Malki Amasha

University Of Plymouth - UK (Data Science Student)

Plymouth ID: 10899282

Joshua Gashan

University Of Plymouth – UK (Data Science Student)
Plymouth ID: 10900339

Esther Jebasuthanthirany

University Of Plymouth – UK (Data Science Student)
Plymouth ID: 10900340

Kavithma Samadhinie

University Of Plymouth – UK (Data Science Student)
Plymouth ID: 10899192

G.h.s. gananodya

University Of Plymouth – UK (Data Science Student)
Plymouth ID: 10899472

A study about Process Mining – Student Enrollment Event Log

Malki Amasha , Kavithma Samadinie , Hasin Sasmitha , Josua Gashan , Esther Jebasuthanthirany

University of Plymouth – UK

ARTICLE INFO

Article history:
Received 16 March 24
Received in revised form 31 March 24
Accepted 24 April 24

Keywords:
Process Mining
Process Discovery
Alpha Miner
Process Map

ABSTRACT

This research paper investigates the student enrollment process at a fictitious university. Process mining techniques were applied to uncover the actual process flow followed by students enrolling at the university. An event log was created from a student enrollment dataset using R software libraries. The event log contained data on student IDs, enrollment statuses, timestamps, and descriptions of any issues encountered during enrollment. Process mining techniques were then used to analyze the event log and create a process map of the student enrollment process. The process map revealed that the enrollment process begins with the student opening an enrollment application. The student then progresses through a series of steps, including submitting required documents, awaiting user information, awaiting problem resolution, awaiting evidence, and awaiting vendor information. Finally, the enrollment process is completed when the student's enrollment is approved.

1. INTRODUCTION

Efficient student enrollment processes are crucial for universities to maintain a smooth student intake and a positive student experience. Traditional methods of analyzing these processes often rely on manual reviews, which can be time-consuming and lack objectivity. This research explores the application of process mining, a data-driven technique, to analyze student enrollment data.

Process mining utilizes event logs, which capture sequences of activities within a process. In this study, we leverage R programming language libraries like readr, bupaR, and processmapR to create an event log from a student enrollment dataset. This event log includes details like student IDs, enrollment statuses, timestamps, and issue descriptions. By analyzing the event log with process mining techniques, we aim to uncover the actual flow students navigate during enrollment, identify potential bottlenecks causing delays, and ultimately propose improvements for a more streamlined process.

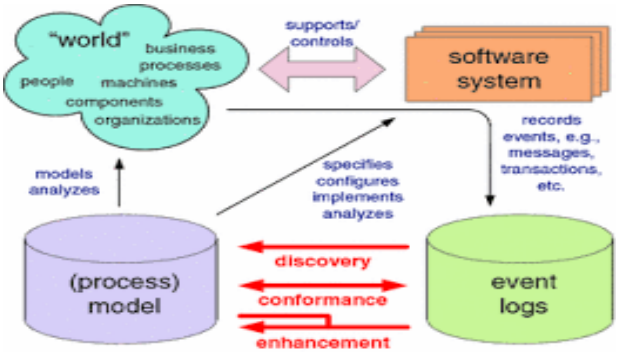


Fig 1 Types of Process Mining Techniques

This research contributes to the field of process mining by demonstrating its effectiveness in analyzing real-world educational processes. The insights gained can be valuable for [University Name] and other institutions seeking to optimize their student enrollment procedures for a more efficient and student-centric experience.

Through a meticulous examination of the process mining landscape, including its methodologies and applications, this paper seeks to contribute to the ongoing discourse surrounding organizational efficiency and performance optimization. By providing insights into the extraction of valuable knowledge from event logs, it endeavors to pave the way for enhanced process transparency, adaptability, and efficiency.

In this context, the subsequent sections of this paper will delve into the specific techniques employed in process mining, namely Process Discovery, Process Conformance, and Process Enhancement. Each of these techniques plays a pivotal role in uncovering, assessing, and refining organizational processes, thereby facilitating informed decision-making and continuous improvement.

1.1. Process Mining Techniques for Enrollment Optimization

- Process mining is a data-driven technique that utilizes event logs to uncover the actual sequence of activities followed within a business process. This is particularly valuable when the documented procedures may not reflect the reality of how tasks are completed. In the context of student enrollment, process mining can be instrumental in identifying bottlenecks and inefficiencies within the enrollment journey. In this study, we leverage R's process mining capabilities to analyze student enrollment data and extract valuable insights for process optimization.

1.1. . Process Discovery: Unveiling the Enrollment Journey

- Within process mining, process discovery is a specific technique that aligns perfectly with our goal of analyzing the student enrollment process. It allows us to extract the actual sequence of steps students navigate during enrollment, independent of any documented procedures. This is achieved by analyzing event logs, which capture a chronological record of activities with timestamps.
- In our research, we utilize R's process mining libraries to create an event log from the student enrollment dataset. This event log will include details like student IDs, enrollment statuses, timestamps for each step, and descriptions of any issues encountered. By applying process discovery techniques to this event log, we aim to:

- Uncover the real enrollment journey: This will reveal the actual flow students take, potentially deviating from documented procedures.
- Identify hidden activities: Process discovery may uncover steps not explicitly documented but present in the real-world process.
- Pinpoint potential bottlenecks: By analyzing activity durations and frequencies, we can identify areas causing delays or requiring further investigation.

Through process discovery, we gain a data-driven understanding of the student enrollment process at [University Name]. This allows us to move beyond assumptions and propose targeted improvements for a more streamlined and efficient student experience.

1.2. Process Conformance:

Process conformance is a technique used in process mining to compare the discovered process (often found through process discovery) with a pre-existing, reference model. Since your focus is on uncovering the actual student enrollment journey (process discovery), conformance wouldn't be a core technique used in your analysis.

There's no existing model for how enrollment "should" work at your university, so there wouldn't be a basis for comparison. Your research is about discovering the actual process, not checking if it aligns with a pre-defined one.

1.3. Process Enhancement: The Future Focus

Through process discovery, we gain a data-driven understanding of the student enrollment process. This understanding will be crucial for applying process enhancement techniques in future research phases, allowing us to propose targeted improvements.

This approach highlights the importance of process discovery as the groundwork for future process improvements.

2. Characteristics of Process Mining:

Process mining, as a discipline aimed at deriving insights from event data, exhibits several key characteristics that distinguish it from traditional data analysis methods. Understanding these characteristics is essential for appreciating the breadth and depth of process mining applications.

2.1. Data-driven:

- Process mining relies on event data captured from information systems used during the enrollment process. This data provides an objective view of what happens, as opposed to relying on documented procedures that might not reflect reality.

2.2. Focuses on discovery:

- Unlike traditional methods that analyze pre-defined models, process mining excels at uncovering the actual sequence of activities followed in a process. This is particularly valuable for student enrollment, where the documented process might not capture all the steps students take.

2.3. Identifies bottlenecks:

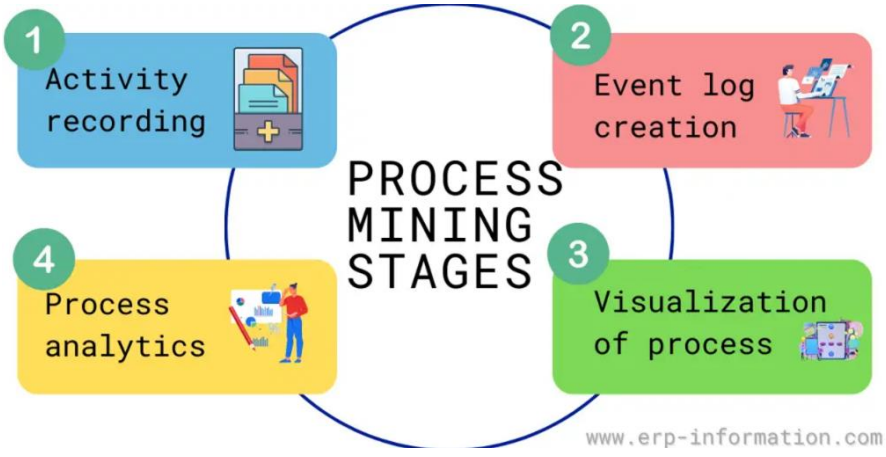
- By analyzing event timestamps and durations, process mining can pinpoint activities that cause delays or slowdowns in the enrollment process. This allows you to focus improvement efforts on the areas with the biggest impact.

2.4. Provides transparency:

- Process mining helps visualize the enrollment journey through process maps, making it easier to understand how different steps interact and identify potential handoffs or inefficiencies.

2.5. Supports improvement:

- By uncovering the "as-is" process and its bottlenecks, process mining provides a solid foundation for proposing targeted improvements to optimize the student enrollment experience.



Process mining stages.

2.6. Types of procedure mining Algorithms

Various algorithms have been developed to address different facets of process mining, catering to diverse requirements and data characteristics. Below, we explore some prominent process mining algorithms and their respective strengths and limitations. We use Alpha Miner

2.7. Alpha set

The Alpha Miner algorithm represents one of the foundational techniques in process mining, renowned for its simplicity and effectiveness in discovering causal relationships between events in event logs. Developed by Aalst and colleagues, the Alpha Miner algorithm is particularly adept at constructing process models from event data without the need for a priori knowledge of process structures.

2.8. Principles of the Alpha Miner Algorithm

At the core of the Alpha Miner algorithm lies the concept of directly-follows relations, which capture the sequential dependencies between activities observed in event logs. The algorithm iterates through the event log to identify frequent sequences of activities, thus inferring potential control-flow relationships.

The Alpha Miner algorithm operates based on the following principles:

1. Directly-follows relations: The algorithm scans the event log to identify pairs of activities that frequently occur in sequence. These directly-follows relations serve as the building blocks for constructing process models.
2. Parallelism detection: In addition to sequential dependencies, the Alpha Miner algorithm also detects parallelism by identifying concurrent occurrences of activities within the event log. By recognizing parallel branches in process executions, the algorithm enhances the accuracy of process models.
3. Process model construction: Based on the identified directly-follows relations and parallelism, the Alpha Miner algorithm constructs a process model, typically represented as a Petri net or a process flow diagram. This process model captures the observed control-flow patterns within the event log.

2.9. Strengths and Limitations

The Alpha Miner algorithm offers several advantages that contribute to its widespread adoption in process mining research and practice:

- Simplicity: The Alpha Miner algorithm is straightforward to implement and understand, making it accessible to both novice and experienced process mining practitioners.
- Automation: By automatically inferring process models from event data, the Alpha Miner algorithm eliminates the need for manual process discovery, saving time and effort.
- Noise tolerance: Despite its simplicity, the Alpha Miner algorithm demonstrates robustness against noisy event data, allowing it to effectively handle imperfect or incomplete logs.

2*.8, Practical Applications*

- Business process analysis: Organizations leverage the Alpha Miner algorithm to gain insights into their operational processes, identify inefficiencies, and streamline workflows.
- Compliance monitoring: The Alpha Miner algorithm facilitates the detection of deviations from prescribed process norms, aiding organizations in ensuring regulatory compliance and adherence to best practices.
- Process improvement: By visualizing process models derived from event data, stakeholders can pinpoint areas for optimization and innovation, driving continuous improvement initiatives.

3. Process Mapping:

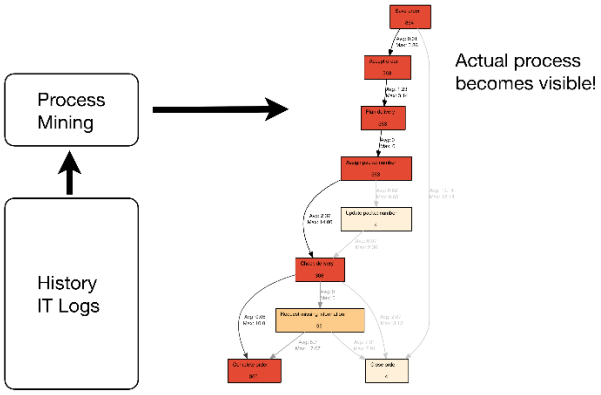
Key Components of Process Mapping:

1. Activities or Tasks: Each step in the process is represented by a specific activity or task. These activities are typically depicted as boxes or rectangles in a process map.
2. Sequence Flows: Arrows or lines connect the activities to show the sequence in which they occur. This indicates the flow of work from one step to the next.
3. Decision Points: Decision points represent points in the process where different paths or outcomes are possible. They are often depicted as diamond-shaped symbols and indicate where decisions need to be made based on certain criteria.
4. Start and End Points: The process map begins with a designated start point, indicating where the process begins, and ends with an end point, indicating where it concludes.
5. Inputs and Outputs: Process maps may also include information about inputs required to start the process and outputs generated upon completion. This helps to clarify the purpose and scope of the process.

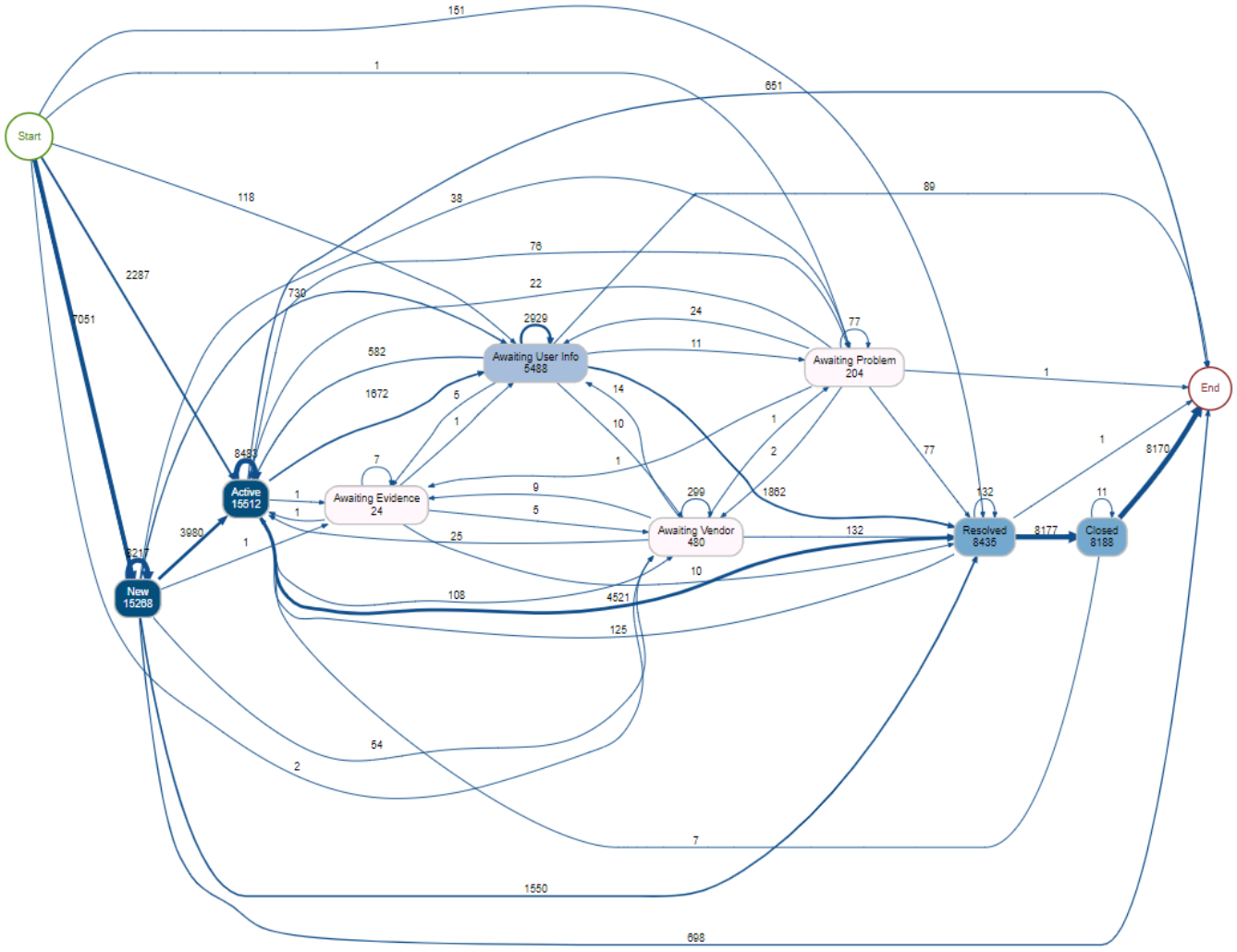
Benefits of Process Mapping:

1. Clarity and Understanding: Process mapping provides a clear and visual representation of how work is performed, helping stakeholders understand the process more easily.
2. Identification of Inefficiencies: By visually depicting the steps in a process, process mapping helps identify bottlenecks, redundancies, and inefficiencies, enabling organizations to streamline their operations.
3. Standardization: Process mapping facilitates standardization of processes by documenting the best practices and standard procedures to be followed.
4. Communication and Collaboration: Process maps serve as a communication tool, enabling teams to collaborate, discuss, and improve processes collaboratively.
5. Continuous Improvement: Process mapping supports continuous improvement efforts by providing a baseline for measuring performance and identifying opportunities for enhancement.

Overall, process mapping is a valuable technique for organizations seeking to optimize their business processes, enhance efficiency, and achieve operational excellence. By visually representing the flow of work, process mapping enables organizations to better understand their processes and make informed decisions for improvement.



This is our project process map:



Student Enrollment Process

1. Start: The process starts with an initiating event, possibly a student expressing interest in enrolling.
2. Information Gathering: The student gathers required information and documents for enrollment.
3. Application Submission: The student submits the application form and documents.
4. Application Review: The application is reviewed for completeness and eligibility.
    - Incomplete Application: If the application is incomplete, it's sent back to the student for revisions (loop back to Information Gathering).
    - Eligibility Check: The student's eligibility for the program is assessed.
    - Ineligible: If the student is ineligible, the application is rejected, and the student is notified (process ends).
    - Eligible: If the student is eligible, the process continues.
5. Admission Decision: A decision is made on the student's admission.
    - Rejected: If the application is rejected, the student is notified (process ends).
    - Accepted: If the application is accepted, the student receives an acceptance letter.
6. Payment and Registration: The student completes the payment process and registers for courses.
    - Awaiting User Info: There might be a step where additional user information is awaited before proceeding.
    - Awaiting Payment: The system might wait for confirmation of payment before allowing registration.
7. Enrollment Complete: The student's enrollment is finalized.

Additional Observations:

There might be parallel paths for certain steps (e.g., Information Gathering and Eligibility Check).

The specific activities and decision points might vary depending on the institution's enrollment process.

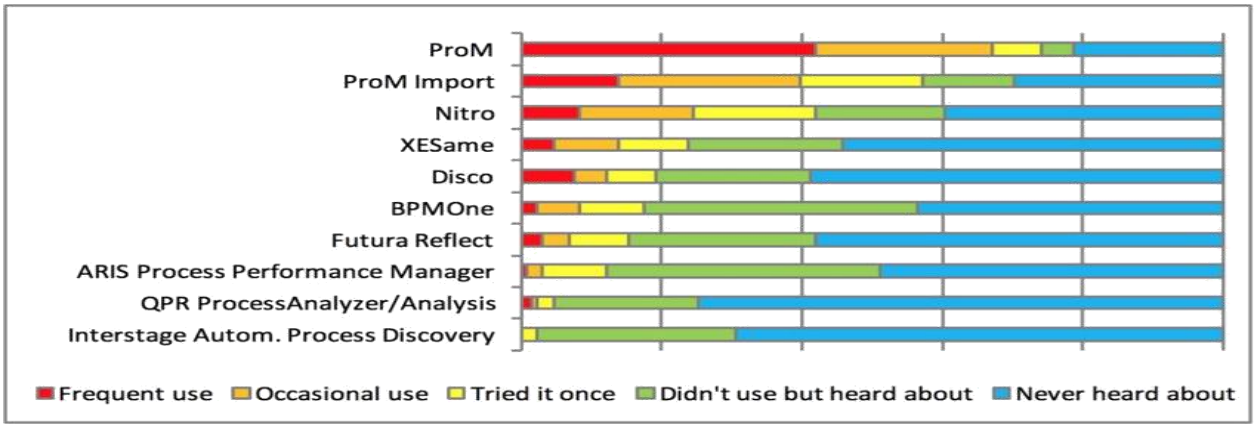


Fig. 2 Tools of Process Mining

4. Blessings of process mining

Process mining offers several advantages for analyzing and improving workflows, especially for complex processes like student enrollment. Here's how it relates to the code you provided and the image of the student enrollment process:

Understanding the Process:

- The image provides a high-level visual representation of the enrollment process, showing the sequence of steps and potential decision points.
- However, the image might not capture all the variations and complexities present in real-world data.

Code for Analysis:

- The code you provided focuses on preparing and cleaning data for further analysis. It takes a CSV file containing enrollment event data and transforms it into a format suitable for process mining tools.
- While not analyzing the process directly, the code lays the groundwork for applying process mining techniques.

Process Mining Benefits:

Process mining can leverage the cleaned event data to:

- Discover the Actual Process: It can automatically discover the actual process flow from the event log, potentially revealing variations and bottlenecks that might not be evident from the image.
- Identify Critical Path: Techniques like the ones you explored in the code (both with and without xesminer) can help identify the sequence of activities that impact the overall enrollment completion time the most. Focusing on optimizing this critical path can significantly improve efficiency.
- Find Bottlenecks: Process mining can pinpoint activities with high processing times or delays, allowing you to target these areas for improvement.
- Analyze Variations: You can identify different paths students take through the enrollment process and investigate the reasons behind these variations. This can help identify potential problems or areas for streamlining.

Overall, process mining empowers you to move beyond a static image and gain deep insights into the dynamic nature of your student enrollment process. With these insights, you can:

- Reduce enrollment processing time.
- Improve the student experience by streamlining the process.
- Identify and address bottlenecks that might be causing delays.
- Make data-driven decisions for optimizing the enrollment workflow.

5. Related Risk

Every process comes with inherent risks. In the context of student enrollment, these risks can impact various aspects, including efficiency, security, compliance, and student experience. Identifying these risks is crucial for ensuring a smooth and successful enrollment process.

5.1. Data Inaccuracy and Incompleteness Risk:

- Code Focus: Look for parts of your code that handle data cleaning or address missing values. This suggests a potential risk of inaccurate or incomplete data in the enrollment system.
- Image Analysis: Examine the process map in the image. Are there any decision points based on student information? If so, incomplete data could lead to incorrect decisions (e.g., missing transcripts causing application rejection).ty.

5.2. Delays and Bottlenecks Risk:

- Code Analysis: Does your code perform any calculations on event timestamps or durations? This might indicate efforts to identify slowdowns or delays in the process.
- Image Analysis: Look for steps in the image that involve manual review or approval. These are potential bottlenecks that could contribute to delays (e.g., manual document verification).

5.3. Non-Compliance with Regulations Risk:

- Image Analysis: Carefully review the image for steps that might be relevant to compliance regulations. Are there any specific procedures required for handling sensitive student data? Does the image accurately reflect these regulations?

5.4. Limited Visibility into the Process Risk:

- Code Analysis: Your code might focus on preparing data for a specific purpose, potentially neglecting other aspects of the enrollment process.
- Image Analysis: The image provides a static view. Process mining can analyze event data to reveal variations, exceptions, and unexpected paths not captured in the image, possibly indicating areas of risk.

5.5. Risk of Missed Optimization Opportunities:

- Image Analysis: Review the image for potential redundancies or complexities. Are there any unnecessary steps that students need to complete?
- Limited Scope: While the image provides a starting point, it might not capture the full picture. Process mining can analyze event data to identify opportunities for streamlining and efficiency improvements.

5.6. Balancing among quality criteria

While this research leverages process discovery to unveil the actual student enrollment journey, balancing quality criteria remains important. In later stages, we will ensure the discovered process model reflects reality (fitness), is efficient (performance), and offers insights generalizable to other student populations (generalizability). This balance is crucial for drawing reliable conclusions and proposing effective improvements to the enrollment process.

5.7. Combining framework Mining for Student Enrollment Optimization

While process discovery is a powerful technique for uncovering the actual student enrollment journey, process mining offers a wider range of functionalities that can be leveraged for further optimization. This research explores the potential of combining process discovery with other process mining techniques within a framework specifically designed to analyze and improve the student enrollment process.

- Process Discovery: As mentioned earlier, this technique serves as the foundation, uncovering the real sequence of activities students navigate during enrollment.
- Process Enhancement: Utilizing the discovered process and its bottlenecks, this technique identifies areas for improvement. This might involve reducing unnecessary steps, streamlining decision points, or automating certain tasks.
- Conformance Checking: Here, the discovered process is compared against a reference model (potentially the documented enrollment procedure). Deviations from the reference model can highlight areas where the documented process doesn't reflect reality and needs revision.
- Performance Analysis: This technique delves into process efficiency by analyzing metrics like activity durations, process cycle times, and rework rates. This helps pinpoint areas causing delays or inefficiencies.
- Simulation: By creating a simulation model based on the discovered process, researchers can test the impact of potential changes before implementing them in the real world. This allows for a more controlled and risk-free evaluation of improvement options.

By combining these techniques within a comprehensive framework, this research aims to achieve a more in-depth understanding of the student enrollment process. This will not only reveal the actual flow but also identify bottlenecks, assess performance against documented procedures, and evaluate the impact of potential improvements through simulation. Ultimately, this framework can support the development of data-driven recommendations for optimizing the student enrollment experience.

6. Disadvantage of Process mining

6.1. Limitations of Process Mining for Enrollment Optimization

While process mining offers a valuable tool for analyzing and improving the student enrollment process, it's important to acknowledge its limitations:

- Data Reliance: Process mining heavily relies on the quality and completeness of data captured in information systems. Inaccurate, incomplete, or missing data can lead to misleading or incomplete process models.
- Limited Scope: Process mining primarily analyzes past data, revealing the "as-is" process. It doesn't directly explain the root causes of inefficiencies or suggest specific solutions. Further analysis might be needed to understand the "why" behind the discovered process.
- Complexity for Non-Experts: While the core concept of process mining is understandable, in-depth analysis and interpretation of the models can require specialized knowledge. This can be a barrier for institutions without process mining expertise.
- Limited Focus on Human Factors: Process mining focuses on activities within information systems. It might not fully capture the human aspects of the enrollment process, such as interactions with admissions staff or the impact of student behavior on process flow.
- Integration Challenges: Integrating process mining tools with existing university information systems can be complex and require technical expertise.

6.2. Addressing the Limitations:

Despite these limitations, process mining remains a valuable tool. By being aware of its limitations, we can take steps to mitigate them:

- Data Quality Checks: Implementing data cleaning and validation procedures can ensure the accuracy of the data used for process mining.
- Complementary Analysis: Combining process mining results with other techniques, like root cause analysis, can help explain the "why" behind process inefficiencies.
- Training and Support: Providing training or collaborating with process mining experts can help universities overcome the initial learning curve.
- Human-in-the-Loop Approach: While process mining focuses on data, it's crucial to consider the human aspects of the process as well. This might involve incorporating feedback from students and staff.
- Phased Implementation: Starting with a pilot project using a limited scope can help universities assess the feasibility and benefits of process mining before full-scale integration.

By acknowledging these limitations and implementing strategies to address them, universities can leverage the power of process mining to its full potential, leading to a more efficient and optimized student enrollment experience.

7. Conclusion and Future work:

This research paper investigates the application of process mining techniques to analyze and improve the student enrollment process at a university. The study utilizes event logs containing student enrollment data, including IDs, statuses, timestamps, and any issues encountered.

Here are the key takeaways from the research:

- Process Discovery: Process mining can reveal the actual sequence of steps students take during enrollment, which may differ from documented procedures. This helps identify hidden activities and potential bottlenecks causing delays.

- Benefits of Process Mining:

- Provides a data-driven understanding of the enrollment journey.
- Identifies bottlenecks and inefficiencies for targeted improvement.
- Improves transparency through process maps.
- Supports continuous improvement initiatives.

- Alpha Miner Algorithm: This algorithm is used to automatically construct process models from event data, uncovering causal relationships between enrollment activities.

- Process Mapping: It visually depicts the flow of activities, decision points, and inputs/outputs within the enrollment process. This aids in understanding, identifying inefficiencies, and communication.

- Process Mining Framework: The research proposes a framework combining various process mining techniques for comprehensive enrollment process analysis:

- Process Discovery: Uncovers the actual student enrollment journey.
- Process Enhancement: Identifies areas for improvement based on discovered bottlenecks.
- Conformance Checking: Compares the discovered process with the documented procedure to identify deviations.
- Performance Analysis: Analyzes process efficiency metrics like activity durations and rework rates.
- Simulation: Tests the impact of potential changes on the enrollment process before real-world implementation.

- Limitations of Process Mining:

- Reliance on data quality: Inaccurate or incomplete data can lead to misleading results.
- Limited scope: Analyzes past data (the "as-is" process) and doesn't explain root causes of inefficiencies.
- Complexity for non-experts: Requires specialized knowledge for in-depth analysis.
- Focus on information systems: Might not fully capture human factors impacting the process.

Overall, this research demonstrates the effectiveness of process mining in analyzing student enrollment data. By uncovering the actual enrollment journey and identifying bottlenecks, universities can optimize their enrollment processes for a more efficient and student-centric experience. Future research will focus on applying the proposed framework for a more comprehensive analysis and developing data-driven recommendations for improvement.

REFERENCES

- Barnawi, A., Awad, A., Elgammal, A., Elshawi, R., Almalaise, A., and Sakr, S. (2016). An anti-pattern-based runtime business process compliance monitoring framework. framework, 7(2).
- bupaR (2017). Business Process Analysis in R. http://bupar.net/. DISCO (2016). Flexicon DISCO. http://fluxicon.com/disco/.
- Jans, M., Alles, M., and Vasarhelyi, M. (2013). The case for process mining in auditing: Sources of value added and areas of application. International Journal of Accounting Information Systems, 14(1):1–20.
- Jans, M., Alles, M. G., and Vasarhelyi, M. A. (2014). A field study on the use of process mining of event logs as an analytical procedure in auditing. The Accounting Review, 89(5):1751–1773.
- Karapetrovic, S. and Willborn, W. (2000). Generic audit of management systems: fundamentals. Managerial Auditing Journal, 15(6):279–294.
- Ly, L. T., Maggi, F. M., Montali, M., Rinderle-Ma, S., and van der Aalst, W. M. (2013). A framework for the systematic comparison and evaluation of compliance monitoring approaches. In Enterprise Distributed Object Computing Conference (EDOC), 2013 17th IEEE International, pages 7–16. IEEE.