Project Setup: Implementation of Process Mining Visualizations for Conformance Checking

Group 2

November 2024

Contents

1 Introduction

In recent years, process mining has gained significantly more popularity as companies across industries pursue data-driven efficiency. Process mining tools, such as PM4Py, allow companies to apply various process mining techniques to their data, revealing insights and optimizing workflows. To enhance PM4Py's accessibility, we aim to implement new visualizations that simplify complex process data, designed to be interpretable even for users with limited process mining experience.

Chapter 2 provides valuable insights into the clients we aim to reach, highlighting their needs and challenges. It includes two use cases demonstrating how our product can be effectively applied across two distinct industries, showcasing all the benefits our product offers.

Chapter 3 will examine the theoretical and technical feasibility of our project, followed by detailed explanations of our chosen programming language and libraries for this project.

In Chapter 4 a Gantt-chart detailing the project timeline and milestones will be provided as a roadmap of our project, along with thorough explanation of our team's workflow and time management.

Chapter 5 provides an overview of the various tools which will be used throughout our project and explains how each tool will be utilized.

In Chapter 6 the team members are introduced, covering their academic background, experience, expectation from this lab and specific roles within the team.

2 Business Case

2.1 Client

BlaBla Solutions, a consulting firm specializing in process mining for industries like finance and healthcare, leverages PM4Py to help clients optimize workflows and enhance compliance. To expand its offerings, BlaBla Solutions is introducing advanced conformance visualizations that provide clear insights into process timing, adherence trends, and activity flows, helping clients quickly identify and address compliance deviations. These intuitive visual tools make it easier for clients to monitor and manage compliance, providing real-time insights that support proactive decision-making. By enhancing its service line with these visualizations, BlaBla Solutions strengthens its position as a leader in compliance analytics, driving growth and attracting clients from regulated industries.

2.2 Example 1: Grocery Discount Franchise

Y GmbH, a major grocery discount franchise, aims to minimize costs to retain loyal customers and attract new ones. They want to use conformance checking at each of their franchises to reduce costs in their supply chains. However, each franchise has their separate providers, so the managers need the flexibility to make decisions on the go about adjusting product prices. OUR PRODUCT allows them to quickly and with little technical know how make informed decisions. Important criteria of Y GmbH for applying OUR PRODUCT are:

- Reduce time to process information.
- Enable franchise managers without technical expertise to make well-informed decisions.

2.3 Example 2: Customer Service Workflow Management

For service-oriented industries, efficient handling of customer service processes is crucial for customer satisfaction and retention. Consider a large customer service center aiming to minimize response times and improve resolution accuracy. Through the static visualizations generated by PM4Py, managers can pinpoint areas where deviations occur frequently, such as delayed responses or misdirected inquiries. These visual insights allow service managers to quickly address inefficiencies in the workflow, resulting in streamlined operations, quicker response times, and overall improved customer experience.

2.4 Benefits

Implementing Process Mining Visualizations for Conformance Checking into a user-friendly app would benefit all the companies mentioned. The software would be accessible to both IT specialists and non-technical staff, providing valuable insights to improve workflows, enhance employee efficiency, reduce costs, and boost the working environment. Additionally, departments like Marketing, Sales, and Finance could use the tool to increase revenue and contribute to overall company success.

3 Feasibility Study

3.1 Theoretical and Technical Feasibility

Conformance checking in process mining involves analyzing whether the actual execution of business processes aligns with a predefined model, usually represented by a Petri net [4] or similar formal structure. This technique is a core part of process mining and is useful for identifying deviations, inefficiencies, or bottlenecks within workflows. Conformance checking compares real event logs with the expected process model to reveal discrepancies that might indicate issues such as skipped steps, unnecessary rework, or deviations that impact performance. In our project we will try to implement different ways of visualization to provide meaningful insights for analysts. The end product will include the following visualization techniques: Temporal Behavior Patterns (*TBP*) Chart, Conformance Heatmap with Concept Drift Over Time (*CHCD Chart*), Activity Interaction Network (*AIN*), Semantic Conformance Word Clouds, Sequence Embedding Conformance (*SEC*) Visualization. The assessment of the end product should cover these topics:
Performance: The issue of speed is a vital factor for possible use in real-world implementation. At this stage of developing we do not foresee problems regarding performance. However, there are some plans for solving this issue. For instance, parallelization of processes and using other programming languages that are faster than Python (*C, C++*).
Implementation: We plan to use established process mining tools to ensure the quality and reliability of the end results. Our primary language will be Python, with libraries such as NumPy for numerical computations, PM4Py for process mining tasks, Seaborn for visualizations, and Pandas for data manipulation and analysis. For further details, refer to *Section 3.2*.
Quality: The quality of our visualizations will be ensured through rigorous model fitness assessments, focusing on conformance checking techniques that compare actual process behaviors with expected models [3]. We will employ quantitative metrics such as fitness and precision, which are essential for evaluating the accuracy of our representations [1]. User feedback will be incorporated to enhance the clarity and usability of the visualizations [4]. An iterative testing and refinement process allows us to continuously improve based on performance metrics and user engagement [2]. By developing specific quality metrics, we can systematically evaluate aspects such as interpretability and user satisfaction. This comprehensive approach will ensure that our visualizations are both informative and reliable.

3.2 Use Cases

3.2.1 Python Programming Language

Python serves as the backbone of this project due to its straightforward syntax, ease of use, and wide range of scientific and data-focused libraries. The language's simplicity, which mimics mathematical notation, makes it approachable for domain experts with limited computer science backgrounds, enabling them to translate complex, domain-specific tasks into code. Python's versatility and intuitive syntax accelerate the prototyping and implementation of various ideas, allowing for rapid development and iteration within process mining projects. Furthermore, Python's large support community and extensive documentation mean that solutions are often readily available, reducing the learning curve and encouraging broader adoption across disciplines. Python's rich ecosystem, including libraries for scientific computation, data processing, and visualization, aligns naturally with the requirements of process mining, making it an indispensable tool for managing and analyzing large datasets and developing meaningful visualizations.

3.2.2 Process Mining Library (such as PM4Py)

A specialized process mining library, such as PM4Py (Process Mining for Python), is essential for addressing the unique needs of process mining projects. Libraries like PM4Py provide a comprehensive set of algorithms and tools specifically designed to analyze event logs, discover process models, and check conformance between real-world data and theoretical models. These built-in functions establish a reliable framework for creating and enhancing process visualizations, as they are tailored to the particular challenges of process mining, such as visualizing complex workflows and identifying process inefficiencies. By leveraging PM4Py's optimized algorithms for handling event log data, we can efficiently explore critical process insights, including workflow patterns, bottlenecks, and compliance deviations. Such capabilities make PM4Py an ideal choice for implementing and extending visualizations in the process mining field.

3.2.3 Visualization Libraries: Matplotlib, Seaborn, and Plotly

Python's robust set of visualization libraries, including Matplotlib, Seaborn, and Plotly, is essential for transforming data into comprehensible and insightful visual representations. Matplotlib and Seaborn are particularly effective for creating static visualizations, which are valuable in detailed analyses and formal reporting. Both libraries allow for extensive customization and produce publication-quality figures, ensuring that static graphs and plots convey critical insights clearly and professionally. Plotly, on the other hand, is indispensable for interactive visualization, offering dynamic plotting capabilities that empower users to explore data more intuitively. For process mining, interactive visualizations are especially valuable as they allow analysts to navigate through complex process flows, trace paths, and identify patterns or anomalies interactively. Together, these libraries provide a wide range of visualization styles that support rapid prototyping as well as fine-tuned customization, which is particularly advantageous in a field like process mining where both clarity and detail are crucial.

3.2.4 NumPy and Pandas

Efficient data manipulation and analysis are at the core of process mining, which is why libraries like NumPy and Pandas are central to this project. NumPy is widely used for numerical operations, enabling fast and efficient handling of large numerical datasets. Pandas, with its high-level data structures like DataFrames, is ideal for handling and manipulating event log data. These structures allow us to organize, filter, transform, and aggregate data with ease, which is critical in the process mining workflow where large logs must be parsed, refined, and formatted for analysis. The performance and flexibility of these libraries make them invaluable for handling extensive data analysis tasks. The combination of NumPy and Pandas ensures that we can manage even large-scale logs and perform complex data transformations smoothly, forming a strong foundation for meaningful and insightful data analysis in process mining projects.

3.3 Conformance Checking and Process Mining Visualization Using Python and PM4Py

In the field of process mining, effective visualization is crucial for extracting actionable insights from complex data. For example, consider a project where a large organization wants to analyze whether its operational processes align with established protocols and where inefficiencies or deviations might exist. This is a typical use case for conformance checking, where recorded event logs are compared against a predefined standard process model. To achieve this analysis effectively, we need powerful tools not only to run conformance checks but also to visualize the results in a way that makes patterns, bottlenecks, and deviations clear to non-technical stakeholders.

PM4Py is instrumental for analyzing event logs by discovering process models and comparing actual recorded behavior against the standard model. With PM4Py's optimized algorithms, we can handle large event logs efficiently, making it feasible to analyze complex workflows with thousands of cases and events. By creating and extending visualizations based on this data, PM4Py provides a foundation for illustrating where process deviations occur, how processes unfold over time, and which segments of the workflow may be causing inefficiencies.

In conjunction with PM4Py, Python's visualization libraries—such as Matplotlib, Seaborn, and Plotly—are used to translate analytical results into clear visual formats. Matplotlib and Seaborn allow us to create static visualizations ideal for reporting and documentation, showing detailed views of process flows and frequency distributions. Plotly, with its interactive visualization capabilities, enhances user engagement

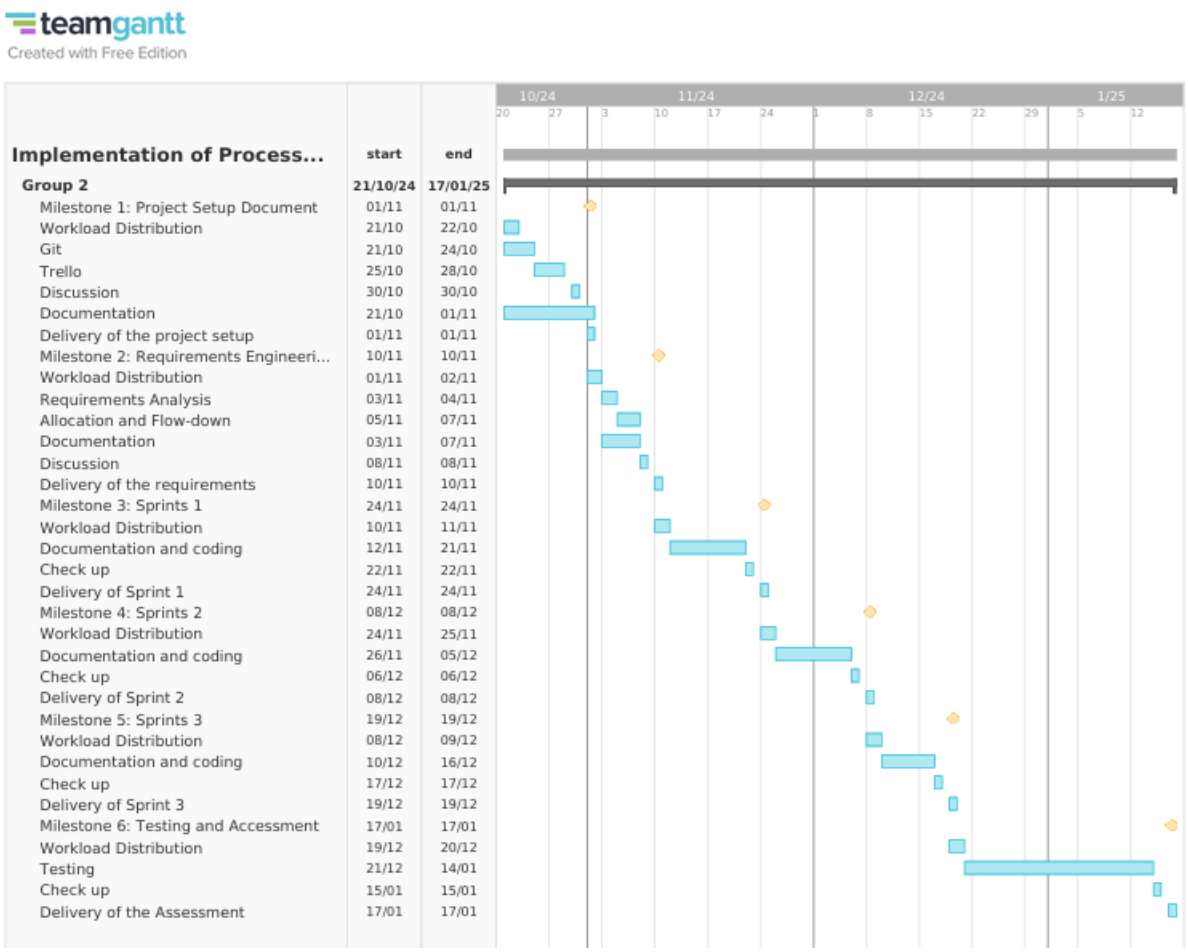
by enabling stakeholders to explore visualizations dynamically, revealing insights that static images cannot provide. This interactive aspect is especially beneficial for process mining, where users can examine process flows, zoom into specific transitions, and identify variations across different cases. Together, PM4Py and Python's visualization libraries form a cohesive toolkit for conducting detailed conformance analyses and creating insightful process visualizations. This approach allows stakeholders to view and interact with data in a meaningful way, revealing inefficiencies and potential improvements in process flow. By using Python's accessible syntax and specialized libraries, this project not only ensures the accuracy of conformance checking but also provides an intuitive interface for stakeholders to make data-driven decisions based on process mining insights.
Our project will enhance the process mining visualization library by adding tools that provide deeper insights into process conformance and behavior over time. The CHCD Chart adds a time-based dimension to track concept drift, while the Activity Interaction Network visualizes interactions and identifies frequent deviations. Semantic Conformance Word Clouds offer a textual perspective on compliance trends, and Sequence Embedding Conformance (SEC) maps similar process instances for clustering analysis. Together, these visualizations will make the library a more robust tool for identifying and understanding process variations.

1. **Event Log Extraction:** The project leverages an event log extracted from an enterprise system, containing:
    - Approximately 100,000 cases (representing individual workflow instances).
    - Around 800,000 events (actions performed within each workflow).
    - 20 distinct task types involved in each workflow instance.
2. **Key Questions Posed:**
    - Are the actual workflows in alignment with the established standard model?
    - How does conformance change over time, and can we identify any concept drift or changes in process patterns?
    - Do different types of tasks and case types exhibit unique interaction patterns or conformance levels?
3. **Results and Insights:**

- **Conformance Heatmap with Concept Drift Over Time (CHCD Chart)**: Analysis revealed that process conformance fluctuated during high-volume periods. This tool made it clear that certain tasks deviated more frequently under high load, signaling a need for resource adjustments during these times.
- **Activity Interaction Network (AIN)**: Frequent transitions between certain activities showed more deviations than others, highlighting specific handoffs that could benefit from streamlining or additional training.
- **Semantic Conformance Word Clouds**: The word clouds revealed recurring non-conformance issues in specific task descriptions, such as "follow-up" and "recheck," suggesting potential bottlenecks or areas where tasks frequently fall out of compliance.
- **Sequence Embedding Conformance (SEC) Visualization**: Analysis showed that instances from similar workflows clustered in ways that reflected distinct compliance profiles, with non-compliant instances clustering separately from compliant ones. This clustering enabled identification of consistent process issues tied to specific case types.

The visualization tools provided a detailed and multi-faceted view of the workflows, offering actionable insights for ongoing improvements. This analysis ultimately led to the development of an internal dashboard, enabling managers to monitor compliance in real time and make informed decisions to maintain efficiency. Hospital staff appreciated these insights and requested further exploration of high-priority workflows to enhance overall process consistency and customer satisfaction.

4 Project Plan Charter



Implementation of Process...	start	end
Group 2	21/10/24	17/01/25
Milestone 1: Project Setup Document	01/11	01/11
Workload Distribution	21/10	22/10
Git	21/10	24/10
Trello	25/10	28/10
Discussion	30/10	30/10
Documentation	21/10	01/11
Delivery of the project setup	01/11	01/11
Milestone 2: Requirements Engineeri...	10/11	10/11
Workload Distribution	01/11	02/11
Requirements Analysis	03/11	04/11
Allocation and Flow-down	05/11	07/11
Documentation	03/11	07/11
Discussion	08/11	08/11
Delivery of the requirements	10/11	10/11
Milestone 3: Sprints 1	24/11	24/11
Workload Distribution	10/11	11/11
Documentation and coding	12/11	21/11
Check up	22/11	22/11
Delivery of Sprint 1	24/11	24/11
Milestone 4: Sprints 2	08/12	08/12
Workload Distribution	24/11	25/11
Documentation and coding	26/11	05/12
Check up	06/12	06/12
Delivery of Sprint 2	08/12	08/12
Milestone 5: Sprints 3	19/12	19/12
Workload Distribution	08/12	09/12
Documentation and coding	10/12	16/12
Check up	17/12	17/12
Delivery of Sprint 3	19/12	19/12
Milestone 6: Testing and Accessment	17/01	17/01
Workload Distribution	19/12	20/12
Testing	21/12	14/01
Check up	15/01	15/01
Delivery of the Assessment	17/01	17/01

Our team's work process is structured to ensure alignment and efficiency. The initial stage of our process is a two-day planning phase. This is not only a time for task distribution, but also for ensuring that all members have a clear and comprehensive understanding of the objectives of each phase. This allows us to reach a consensus on our goals and approach. Throughout the project, we maintain open communication to prevent any team member from becoming overly focused on a single task, which allows us to progress smoothly as a team.

Following the planning phase, each member will work on their assigned tasks with the objective of completion before the three-day check-up phase. This period allows for problem-solving, handling any unforeseen issues, and final adjustments. Three days before the official milestone for discussions, final reviews, and any last-minute changes, we ensure that we have a buffer for unexpected adjustments, keeping the process adaptable but organized. This allows us to make any necessary changes without undue stress.

This initial schedule serves as a roadmap, though specific dates may adjust as the project evolves.

5 Tools

5.1 GitHub

We chose GitHub for its ability to streamline team collaboration and provide a unified space to manage both code and documents. GitHub's easy-to-use interface allows us to track changes, manage tasks, and review work collaboratively, helping everyone stay organized and aligned. With its ability to upload and manage all project files in one place, GitHub makes teamwork smoother and keeps our project progress clear and accessible. Click here to visit our Git repository.

5.2 Trello

It was decided to use Trello as the management tool for our team. Trello is a web-based application that provides an opportunity to create kanban-style boards, to-do lists and has a collaboration mode. It's a perfect solution that would help our team manage our workload in a way that is efficient and understandable. Our board is divided in such a way that the tasks are divided in three sections: To-Do, Doing, Done. So that the progress is clear for everyone. Apart from that there are sections for notes and resources. This way all information is accessible and organized. Click here to visit our Trello board.

5.3 PM4Py

Our aim is to extend PM4Py, a comprehensive Python library designed for process mining. PM4Py supports essential tools for analyzing and modeling process data, making it essential for developing our Temporal Behavior Patterns (TBP) Charts and Conformance Heatmap with Concept Drift Over Time. These visualizations will provide valuable insights into compliance issues within business processes. It is an open-source tool and intended to be used in both academia and industry projects developed by the Fraunhofer Institute for Applied Information Technology.

5.4 Whatsapp

We will use Whatsapp as our primary communication tool. As a widely-used messaging platform, WhatsApp supports instant messaging, voice notes, polls, and more. We've set up a group chat with all team members, making it easy to keep everyone updated and share information quickly. The poll feature is especially helpful for finding suitable dates and times for group meetings. WhatsApp is ideal for staying connected in real-time regardless of everyone's location.

5.5 Discord

We will use Discord as a secondary communication Tool for team collaboration. Discord is a versatile social platform that supports text messaging and voice calls, making it also ideal for real-time collaboration regardless of team members' location. The platform facilitates communication by allowing us to create dedicated channels for different topics, such as General, Notes-Resources,Session planning and so on. Our group meetings will take place in a voice channel, where we can discuss problems, assign tasks to each member and set milestones together.

6 Team

6.1 Personal Description

6.1.1 Lingjing Zhou

I am currently an eleventh-semester bachelor student at RWTH-Aachen University, majoring in Computer Science and Mathematics, with Business Administration as my application subject. I have foundational knowledge in Java, C, and Python and Software Engineering, with a deep understanding of the Software Development Life Cycle. Furthermore, I am eager to gain more practical experience and am highly motivated to learn and complete the online course to expand my knowledge in Process Mining. My enthusiasm for learning new concepts drives me to continuously improve and apply my skills in meaningful ways.

6.1.2 Hanna Yashchuk

I am a Computer Science student in my 7th semester at RWTH Aachen University. Throughout my studies, I have gained experience across various areas of computer science, with a particular passion for Data Science and Machine Learning. My interest in these fields has motivated me to join this project to further expand my knowledge. I possess both theoretical knowledge in Data Science, especially in Process Mining, and practical experience as a Student Software Developer. My core technical skills include Python, SQL, Statistics, and Big Data. Through this project, I aim to gain valuable experience and insights that will support my professional growth and career development.

6.1.3 M. Ali Agharazi

I am a Computer Science student in 6th semester at RWTH Aachen University with psychology as a minor. I am experienced in writing C and Java codes, with additional experience in Python in statistics. My knowledge in Elements of Machine Learning and Data Science and my previous experience visualizing data in Processing and p5js as a hobby would be beneficial to this project. I am looking forward to learning to use new techniques to visualize data.

6.1.4 Uy Sa Huynh

I am a Computer Science student in 7th semester at RWTH Aachen University with business administration as a secondary subject. I have basic knowledge in Java, Python and C. I have a high interest in Data Science and am looking forward to applying my theoretical knowledge from Business Process Intelligence and Elements of Machine Learning and Data science in a practical setting. I expect to learn more about process mining and enhance my programming skills.

6.1.5 Maulik Jain

I am a 6th-semester Computer Science student at RWTH Aachen University. I possess foundational knowledge in programming languages such as Java, C, Python, Swift, Node.js, JavaScript, and MySQL. My coursework includes Computer Vision, Elements of Machine Learning and Data Science (EMLDS), and Data Science, equipping me with a robust understanding of data analysis and process optimization. I am eager to gain practical experience in the field of process mining and compliance analytics, leveraging my academic background to contribute effectively to the project and enhance my technical skills.

6.2 Roles

It was decided to use a team approach for each role. The reason for that is that sharing responsibilities could ensure that goal would be accomplished in the most effective and comfortable way possible. Additionally, every team member would be engaged according to their skills and intentions.

Scrum Master: The Scrum Master role is handled by Maulik Jain and Hanna Yashchuk. Scrum Master task is to organize workflow in the most efficient, effective and collaborative way possible. They also map out the plan to ensure that all the goals are achieved in a timely manner and the end product is high-quality. We have two Scrum Masters in our team. The tasks will be divided as follows. Hanna is managing the Trello board and divides the tasks. Maulik is responsible for the GitHub repository and moderation of team meetings. Both Scrum Masters examine every task.

Product Owner: Product owner is responsible for establishing clear communication between the team and teaching assistants. These communication tasks include, but not limited to, handing in the final product for each deadline, arranging meetings and being in touch with the teaching assistants. The Product Owner role will be performed by Lingjing Zhou and Ali Agharazi will act as the backup option in some cases where the lead can not perform their assigned tasks.

Theory Expert: Theory Expert will handle the theoretical part of the project. They should have thorough knowledge of the theoretical background of used algorithms or topics, perform necessary literature research on the topic, and also be able to explain and simplify difficult concepts to the rest of the team. These duties will be performed by Maulik Jain, Uy Sa Huynh and Hanna Yashchuk depending on their area of expertise.

Tools Manager: Lingjung Zhou and Uy Sa Huynh will contribute as Tools Managers, bringing expertise in tool selection and integration. They will provide the necessary tools for the project, ensuring that we have access to the right resources for each phase. Their insight into various tools will enable seamless integration to support our team's work efficiently.

Software Expert: Hanna Yashchuk, Maulik Jain, and Ali Ahgharazi are experts in software development. They will apply their skills in Python, PM4PY, and other libraries to help our team transform theoretical algorithms into efficient, practical programs.

Test Manager: Tests and Quality Assurance will be handled by every member of the team. Designing and performing tests will ensure the best possible outcome of the project while minimizing risks for potential clients.

7 Phase review

7.1 Lingjing Zhou

I am very pleased with the communication and cooperation that has developed among our team members. Everyone has been highly engaged and proactive, which has allowed us to establish a productive workflow using WhatsApp for instant communication and Discord for meetings and focused work sessions. Together, we have gained a foundational understanding of process mining and have started learning about the functionalities of PM4Py.
The use of GitHub and Trello has streamlined our project management and task allocation, which will be invaluable for our collaborative efforts in the upcoming phases. Continuously learning new skills and knowledge has been greatly rewarding, and I look forward to the upcoming tasks and further collaboration with our enthusiastic team.

7.2 Hanna Yashchuk

I'm thoroughly enjoying the experience of working on this project, especially as it allows me to deepen my knowledge of process mining. Every team member is fully committed to achieving the best possible outcome, and our communication has been exceptional. This open and efficient interaction keeps us aligned and has been key to our smooth progress. I truly appreciate the chance to collaborate with such dedicated colleagues, and I'm excited about the learning and growth this project brings. We're well on track to meet our goals, and I look forward to what lies ahead in the next phases.

7.3 Maulik Jain

I am very excited to gain my first practical experience participating in a software project with my great team. We successfully planned our project, ensuring smooth collaboration from the start. Our teamwork has been efficient, thanks to our use of Trello for task assignment and tracking completed tasks. We maintained open communication through Discord and Whatsapp and Git, consistently providing each other with constructive feedback. I am very satisfied with the collaboration and support within our team, as everyone is committed to being present and assisting one another. This proactive approach has enhanced our overall performance and morale. We are on track with completing the first Milestone and are confident in our ability to tackle the upcoming phases of this project effectively.

7.4 Uy Sa Huynh

I am exited to learn more about process mining throughout this whole project with my great team. I am really satisfied how we are acting as a team, we listen to each other and make sure that we have a good working environment. We made sure everyone is satisfied with their roles and a even distribution of tasks on this document. Everyone worked really hard on this document. Overall the first phase went really well and I look forward on finishing this lab successfully with my team.

7.5 M. Ali Agharazi

I am very excited to work on this project with a great team. We have successfully planned our project. Our team has been dedicated and efficient. Everyone has taken part in their respective roles and have helped others when needed. We are all motivated to work on this project, and we are also communicating well, given the tools we use. I appreciate the chance to learn new things on this topic with this group.

References

[1] Jorge Munoz-Gama. *Conformance Checking and Diagnosis in Process Mining: Comparing Observed and Modeled Processes*. Lecture Notes in Business Information Processing. Springer Cham, 1 edition, 2016.

[2] Hind R'bigui and Chiwoon Cho. The state-of-the-art of business process mining challenges. *International Journal of Business Process Integration and Management*, 8:285, 01 2017.

[3] A. Rozinat and W.M.P. van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 2008.

[4] Wil van der Aalst. *Data Science in Action*, pages 3–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.