Utrecht University

Master Thesis

Using the Opportunities and Challenges of Process Mining to Improve Understandability in its Application Domains

Author:
M. (Max) Gompel

Supervisor:
Dr. I.M (Iris) Beerepoot

Second Examiner:
Dr. ir. J.M.E.M. (Jan Martijn) van der Werf

Department of Information and Computer Sciences

October 12, 2023

Abstract

Department of Information and Computer Sciences

Process mining is a technique that uses the event logs of a system to discover, monitor, and improve process models. It can be applied in many domains, but the knowledge of a process mining expert is often needed. This research sets out to make process mining more available to new users by making them aware of the opportunities that process mining can bring and the challenges they can encounter during the application. It does so by performing a systematic literature review that analyzes the types, perspectives, opportunities, and challenges of process mining. It looks at these four aspects from six different views, being the main domains in which process mining is applied: Healthcare, ICT, Manufacturing, Education, Finance, and Logistics. This domain perspective is chosen so that domain experts without process mining experience can learn the opportunities and challenges that are specific to their domain. The results show little difference between the domains regarding the types and perspectives of process mining. There are differences between domains in opportunities and challenges. Some are shared by multiple domains whilst others are domain-specific as they originate from the characteristics of that domain. These findings are combined in six infographics that make them accessible to new process miners. It also includes recommendations on how to deal with the challenges.

Keywords- *process mining, domains, opportunities, challenges, types, perspectives*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Process Mining

Process mining is a technique that uses the event logs of a system to discover, monitor, and improve process models. Analyzing these models can provide insights into a system's behavior. This can be used to identify bottlenecks and deviations, diagnose performance and compliance problems, and detect repetitive tasks that could be automated or removed. Because these models are derived from the event logs, they represent the actual process as-is. This can be compared to the assumed process to show the difference between desired and actual system behavior [127]. This comparison is called conformance checking and is one of the three types of process mining. The other types are discovery, where process mining is used to construct a model based on the event logs, and enhancement, where data from the event log is used to improve a process [129]. Many organizations have processes that could be optimized. Given that most digital systems produce event logs, process mining can be applied in many types of organizations. The main domains that currently apply process mining are healthcare, ICT/information technology, manufacturing, education, finance, and logistics [26].

1.2 Problem statement

The people who work in these domains are often focused on domain-specific tasks but have little to no experience with process mining applications. They might have heard about process mining and its possible benefits, but when they want to apply it, they need a process mining expert to help them with the implementation. Improving the understandability of process mining to new users is one of the eleven challenges that process mining faces, according to the process mining manifesto [129]. Given the limited experience they have with process mining applications, it could be unclear what opportunities process mining can bring within their domain, given that this can depend on the situation and the available data. There are also some challenges that can occur during the application of process mining. Knowing these opportunities and challenges can be helpful when investing in a process mining application. Martin et al. [81] used a Delphi study to investigate what opportunities and challenges are most relevant. For this study, they used experts from both academia and industry, with varying backgrounds. Having industry experts with different backgrounds combined gives a clear overview of the challenges and opportunities that process mining had in general, but misses out on the possible differences between the application domains. For example, one of the challenges of process mining is that employees might feel watched and are reluctant to cooperate with a process mining project, fearing that their work will be automated, making them obsolete [81]. Whilst this might be true for workers in IT, financial, and logistical domains, this could be less likely for healthcare workers as they experience high workloads that could be reduced by automating processes [33]. For them, the privacy of the patients and the anonymization of their data can be a much more relevant challenge [1]. This thesis aims to find the differences in the opportunities and challenges between these domains. By knowing these differences we can create infographics that help inexperienced process miners to identify the opportunities and challenges that can arise when they apply process mining in their domains. The infographics also recommend process mining techniques that are best suited given the opportunities and challenges of that domain.

1.3 Research Objectives

This research sets out to make new users aware of the opportunities and challenges that can arise when applying process mining, by listing the relevant opportunities and challenges that are specific to the six main application domains. We want to achieve this by analyzing opportunities that process mining can bring, giving new users a clearer view of the benefits they can expect. We also look at the challenges that can arise when applying process mining. Making new users aware of these challenges helps them to overcome them and prevents unexpected setbacks. For each domain, the number of times different types and perspectives of process mining are being applied is also analyzed. This can show the reason for implementing process mining in each domain. These results can be used to create infographics that educate inexperienced process miners on the opportunities and challenges that can arise when they apply process mining, thus improving their understandability.

1.4 Research Questions

Main Research Question

To achieve the research objective, the following main research question is formulated.

What are the differences between the applications of Process Mining in its six main application domains?

Sub Research Questions

The following sub-questions are formulated to answer the main research question.

RQ1: What types of process mining are being applied in the different domains?
When process mining is applied, information is extracted from event logs. This information can be used to achieve various goals. These goals can be categorized into three types of process mining. Discovery is a type of process mining where a process model is constructed based on the event data. Conformance is when an existing model, that shows desired behavior, is compared to the event

logs of the same process, or the actual observed behavior. Enhancement is a type of process mining that aims to extend or improve an existing process model by using information from the event log.

RQ2: What process mining perspectives are used in the different domains?
When process mining techniques are applied, it can look at the event logs from different perspectives. There are four main perspectives defined [127]. Control-flow is a perspective that looks at the order in which activities take place. The organizational perspective looks at the resources in the log, such as actors, and how they interact. The case perspective concerns the properties of a case, and the values that those properties have. The fourth perspective is time and looks at the timing of the events.

RQ3: What opportunities present themselves when applying process mining in each domain?
To find out what benefits process mining can bring to each domain, research has to be done on the opportunities of process mining in each domain. In this research, the following definition of opportunities by Martin et al. is used [81]. *"An opportunity is a favorable circumstance or an expected benefit for an individual, a team, or an organization enabled by the use of PM in organizations."*

RQ4: What challenges arise when applying process mining in each domain?
The same research has to be done for the challenges to find out what can go wrong when applying process mining. The definition of challenges by Martin et al. is used in this research [81]. *" A challenge is a difficulty or an obstacle that arises when using (or intending to use) PM in organizations, and that requires a lot of energy and determination from an individual, a team, or an organization to overcome."*

RQ5: Considering the major opportunities and challenges in each domain, what should be taken into account when applying process mining?
When the opportunities and challenges of each domain are known, this information can be used to inform inexperienced process miners about the specific opportunities that process mining can bring within their domain. It can help them to be aware of what process mining is and is not capable of. It can also inform them of the challenges that they might encounter during the application. Recommendations can be made to help mitigate those challenges. The information will be presented in an infographic per domain that aims to help non-experts with their process mining application.

Chapter 2

Process Mining

This chapter provides background information on subjects that are used in this research. The types and perspectives of process mining are explained, followed by an overview of the six main application domains and the opportunities and challenges.

2.1 Types of Process Mining

Process mining can be divided into three types: discovery, conformance, and enhancement. These three types all extract information from event logs but with varying goals and outcomes. The three types are explained below.

Process discovery is a type of process mining that aims to construct a process model based on an event log. The constructed model is an abstract representation of that log and presents insights into its captured behavior [131]. The model is created without a-priori information, meaning that it can be made without additional information about the process and is based only on the event log [127].

Conformance checking is a type of process mining where an existing process model is compared with the event log of the same process [127]. This comparison can show the differences between the expected or desired behavior (the process model) and the actually observed behavior (the event logs).

Enhancement is a type of process mining that aims to extend or improve an existing process model by using information from an event log. Whilst conformance checking looks at the differences between a model and reality, enhancement aims to correct these differences by changing the model [127]. A process model can be enhanced by repairing or extending it [142]. Repairing a model is when it is changed to reflect reality better. Extending is when a new perspective is added to the model. An example would be adding time stamps.

2.2 Process Mining Perspectives

Besides these types, process mining can also be categorized by the perspective it takes. Perspectives look at the aspects that process mining aims to analyze. The process mining book defines four perspectives, though these are non-exhaustive. Multiple perspectives can be used in a process mining application. The four main perspectives are explained below[127].

The control-flow perspective, also known as the process perspective, looks at the control-flow of a process [52]. It identifies the activities performed and the order of their execution [45]. This is often displayed in a process modeling language, such as Petri net, BPMN, or UML [103][28][13]. This perspective answers the 'How' questions of process mining [130].

The organizational perspective concerns the 'who' and looks at the resources in a log, like the actors and how they are involved with activities and how they interact [114]. Its goal can be to show the structure of an organization or its social structure. Tasks that belong to this perspective are identifying organizational roles, work distribution among resources, and activities that can be executed by a particular resource [45].

The case perspective looks at the 'what' and concerns the properties of cases. The values of those properties can be useful when a case stands out. For example, when you look at a case with an unpaid order, it can be helpful to know what customer is linked to that case, in order to send a reminder.

The time perspective looks at the 'When' of process mining. Some research defines only three perspectives, excluding the time perspective [130] [63]. However, they refer to [127] for these definitions, where all four perspectives are given. The time perspective uses timestamps to look at the timing and frequency of events. This makes it possible to discover bottlenecks and estimate the duration of similar events. Some refer to this perspective as the performance perspective [52].

2.3 Application Domains

Process mining is applied in a variety of domains. A mapping study [26] looked at which domains have been researched the most. The top six domains account for 79% of all papers. These six domains are elaborated upon below.

2.3.1 Healthcare

Healthcare is the domain that covers clinical paths, patient treatment, and the primary processes of a hospital. Covering 28% of papers, it is by far the most researched domain in the field of process mining [26]. Process models in this domain are characterized by high variability, complexity, security, and privacy [80][85][146]. As with other domains, the application of process mining can help with decision making and cost reduction. More specifically for the healthcare domain, it can also help with identifying flows of patients with certain diseases, estimating chances of treatment, finding correlations among treatments, and improving the quality of treatments [146][102]. An example of a process mining application in the healthcare sector is a case study in a Dutch hospital that looked at careflows of patients. They analyzed event logs containing diagnostic and treatment activities of patients. By looking at the process from the control-flow, organizational, and performance perspectives, they gained insights that could be used to improve the careflows. They identified core paths and found correlations between departments. It was also concluded that healthcare processes can be unstructured and that new process mining techniques need to be developed to handle this [78]. The healthcare domain also faces some challenges. Events are often still entered manually in medical systems, leading to missing or incorrect events [79]. Other challenges are the lack of information integration from different health systems, and the high variability of patients' event trajectories [21].

2.3.2 ICT

The Information and Communication Technology (ICT) domain covers process improvement/evaluation, agile task management, on IT systems. It accounts for 16% of research papers [26]. In this domain, process mining can be used to evaluate and analyze software processes [105]. It has also been shown capable of improving the maturity levels of software processes [71], and it is used to analyze and improve maintenance processes using bug report logs [44].

2.3.3 Manufacturing

Manufacturing takes up 13% of the research papers and covers production processes. These are processes from enterprise resource planning systems (ERP) and manufacturing execution systems (MES) [26]. Process mining can help to discover hidden relationships in production processes that can lead to higher flexibility and real-time decision support [47]. In the manufacturing domain, process mining can also be applied to industrial equipment. This introduces a challenge as these processes are less structured compared to administrative processes [104]. Another implementation of process mining is the use of a Bayesian network of predictive models to predict maintenance intervals, thus reducing downtime [108].

2.3.4 Education

Process mining is being applied in education to discover, analyze, and improve learning processes, by analyzing data from educational information systems [122]. The educational domain accounts for 10% of the research papers on process mining. It uses process mining to analyze recurring behavior in learning processes and recommend learning paths based on user groups and learning styles. Also, the relations among students and between students and instructors are analyzed [26]. Challenges in the educational domain are mostly related to the data. These are challenges like dealing with duplicate events and having small data sets [122].

2.3.5 Finance

The financial domain accounts for 6% of research papers on process mining. It focuses on financial institutions such as banks and insurance companies. Process mining is applied for risk analysis and prediction [20], insurance claim handling [65], ATM processes, loan approval, credit card checks, and fraud analysis [26]. Process models of loan applications can be complex and unstructured, making analysis difficult and time-consuming. Process mining techniques, such as the CSM miner, can help to analyze these processes [30]. The size of datasets can also be an issue in this domain, as some businesses collect much data about transactions [139].

2.3.6 Logistics

The sixth domain is logistics and covers 5% of the research papers on process mining. It looks at logistical processes about industrial shifts, warehouse layouts, traffic control, and resource allocation, among others [26]. In this domain, process mining can be applied to supply chain systems, where a lack of data aggregation can lead to difficulties when identifying process instances. This is due to the way the data is logged, as logistic systems often do not use events to log data [38]. The logistics domain is also challenged by often having multiple parties involved, resulting in duplicate data and different data formats [48].

2.4 Opportunities and Challenges

The application of process mining can bring both opportunities and challenges. Martin et al. have researched these opportunities and challenges for process mining in organizations by using a Delphi study [81]. They defined them as follows.

Definition: *"An opportunity is a favorable circumstance or an expected benefit for an individual, a team, or an organization enabled by the use of PM in organizations."*[81]

Definition: *"A challenge is a difficulty or an obstacle that arises when using (or intending to use) PM in organizations, and that requires a lot of energy and determination from an individual, a team, or an organization to overcome."*[81]

They proposed a set of 30 opportunities and 32 challenges that can be categorized based on the BPM core elements being strategic alignment, governance, methods/information technology, people, and culture [23]. For the opportunities, the methods/IT category is split up into the phases of the BPM lifecycle, process discovery, process analysis, process redesign and implementation, process monitoring, and controlling [28]. The research concluded, among others, that academics and practitioners generally agree with each other when it comes to the relevance of process mining opportunities. This is quite the opposite with the challenges, as only three were deemed extremely relevant by both academics and practitioners. This makes it interesting to also look at the differences in opportunities and challenges between the application domains, as the research did not make a distinction between domains.

Chapter 3

Research Method

This section proposes the research methods that will be used to answer the research questions. Two methods will be used to do this. Firstly, a systematic literature review (SLR) will be conducted to gather information about the types, perspectives, opportunities, and challenges of the six process mining domains. Secondly, interviews will be held with domain experts to verify the SLR results. The method that is used for each research question can be seen in Table 3.1

	SLR	Expert interviews
RQ1: What types of process mining are being applied in the different domains?	X	
RQ2: What process mining perspectives are used in the different domains?	X	
RQ3: What are the major opportunities when applying process mining in each domain?	X	X
RQ4: What are the major challenges when applying process mining in each domain?	X	X
RQ5: Considering the major opportunities and challenges in each domain, what should be taken into account when applying process mining?	X	X

Table 3.1: Research methods applied per research question

3.1 Systematic Literature Review

To find the current opportunities and challenges of process mining in the six main domains, a systematic literature review will be performed. The systematic literature review consists of two rounds. The first round is used to select papers about process mining and the domain. The second round is used to select papers about the opportunities and challenges of process mining. These papers will be collected using the Scopus database. This database is chosen as [40] showed that it provides a good balance between relevance and quality, but also covers the majority of process mining-related papers.

3.1.1 Queries

For each domain, a query is created that consists of the name of the domain and 'Process mining'. Sometimes two terms are used for the name of the domain to ensure more representative papers. 'Information technology' is added to the ICT domain and 'Banking' is added to the Finance domain. A filter is applied that whitelists articles and conference papers to exclude other types of publications. The full queries can be seen in table 3.2.

Domain	Articles	Query
Healthcare	339	DOCTYPE (ar OR cp) (TITLE-ABS-KEY ("process mining") AND TITLE-ABS-KEY (healthcare))
ICT	73	DOCTYPE (ar OR cp) (TITLE-ABS-KEY ("process mining") AND TITLE-ABS-KEY ("ICT" OR "Information technology"))
Manufacturing	185	DOCTYPE (ar OR cp) (TITLE-ABS-KEY ("process mining") AND TITLE-ABS-KEY ("Manufacturing"))
Education	152	DOCTYPE (ar OR cp) (TITLE-ABS-KEY ("process mining") AND TITLE-ABS-KEY ("Education"))
Finance	43	DOCTYPE (ar OR cp) (TITLE-ABS-KEY ("process mining") AND TITLE-ABS-KEY ("Finance" OR "Banking"))
Logistics	82	DOCTYPE (ar OR cp) (TITLE-ABS-KEY ("process mining") AND TITLE-ABS-KEY ("Logistics"))

Table 3.2: Queries used for the domains

3.1.2 Paper selection

The papers that resulted from the queries go through two selection rounds. In the first round, the titles and abstracts of the papers are read. Papers get selected when they present research on process mining specifically within the given domain. In the second round, the results and discussions of the selected papers are scanned to determine if they mention opportunities or challenges. It is chosen to only read the results and discussions as literature sections that mention opportunities or challenges will likely refer to another paper that is also part of the SLR. This means that the same research is not counted twice, and it is a time-saving that is needed to make the SLR feasible. In this second round, the papers are scanned until either an opportunity or challenge is found. If one is found, the paper is selected. After the second round, we are left with a set of papers on the opportunities and/or challenges of process mining. These are the papers that will be analyzed further. This will result in a set of opportunities and challenges per domain. This set will be validated through interviews with process mining experts. The number of papers per round, per domain can be seen in table 3.3.

Inclusion criteria query:
-The paper contains the words "process mining" and "(domain name)"
-The paper is a journal article/paper or a conference paper

Inclusion criteria first round:
-The paper is about process mining within the specified domain.
-The paper is written in English
-We can access the full paper for free.

Inclusion criteria second round:
-The paper mentions process mining opportunities or challenges in its results or discussion.

Domain	Initial search	1st Round	2nd Round
Heathcare	339	227	8
ICT	73	23	17
Manufacturing	185	100	30
Education	152	82	27
Finance	43	17	16
Logistics	82	31	20
Total	874	480	117

Table 3.3: SLR papers

3.1.3 Paper analysis

When a paper is selected for the first round, the titles and abstracts are used to determine the process mining types and perspectives that have been used in that paper. If this cannot be determined from the title and abstract, more contents of the paper are read until the types and perspectives used are known. The types and perspectives are logged in a spreadsheet. Each domain has its own spreadsheet that contains the title of the paper, the SLR stage the paper is in, the storage location of a digital copy of the paper, the types used, and the perspectives used. This makes it easy to add up the types and perspectives that are used in each domain.

The 117 papers that are selected for the second round are analyzed on their opportunities and challenges. The results and discussions of each paper are read. The opportunities and challenges that are found are written down in the domain spreadsheet.

The opportunities and challenges are generalized to make comparisons between domains possible. For example, lowering a student's chances of failing a course is an opportunity in education but will not occur in other domains. On the other hand, improving the customer experience by shortening the reply time of customer service can appear in finance, but is not likely in education. By generalizing these into *Optimize the process to improve the quality of the result*, domains can be compared on their use of process mining to improve the quality of the result of a process. It was chosen not to reuse the opportunities and challenges by Martin et al., as they defined 30 opportunities and 32 challenges [81]. To make the comparison between domains, a smaller set is required. The chosen set of opportunities and challenges is explained below.

Opportunities

- **Optimize process (Cost and time)**. Process mining is being used to improve a process by shortening its duration or lowering the cost.
- **Optimize process (Quality of result)**. Process mining is being used to improve the result of a process. This can be a higher quality product or a happier customer.
- **Check conformance with regulations**. Process mining is used to look at the current process, and determine if it is performed according to the rules and regulations. These regulations can come from the organization itself, but also from a governmental level.
- **Analyze communication**. Process mining is used to look at the interaction between two entities. In most cases, this is done to optimize the process. This opportunity is for the discovery of communications that are unknown, but do not aim to optimize.

- **Track performance**. This opportunity is similar to *Optimize process (Cost and time)*. The difference is that process mining is used here to monitor the process. Once something is detected that is not desired, the process can be optimized to reduce cost and time. To be categorized under this opportunity, the process mining application cannot directly look for optimizations but must monitor until deviations are found.

- **Identify and predict fraud cases**. This opportunity is specific to the financial domain. It gets its own category as it does not fit others. It does not fit under *Check conformance with regulations*, as this category looks at the conformance of the process itself, not of a client. Whilst it is intended to save money, it does not categorize as *Optimize process (Cost and time)*, as this aims to save costs by changing the process, not by identifying cases that should not be accepted.

Challenges

- **Handling data from heterogeneous sources**. This challenge is on the difficulties that arise when data comes from different systems and in different formats.

- **Low data quality**. This challenge occurs when the data is too low in quality to get a reliable result. Data is considered low quality when it contains incorrect data, missing data, inconsistent data, or too little data.

- **Handling variable data**. This challenge occurs when the data is variable in nature. This occurs when each case is slightly different, resulting in many unique traces.

- **Handling private data**. This challenge is about handling private data from individuals like customers and employees.

- **Lacking visualizations**. This challenge is encountered when the results of a process mining project cannot be visualized in a way that makes it understandable for people without a process mining background.

- **Large computational costs**. This challenge occurs when the analysis of a dataset takes too long, or too much computational power to make it feasible.

- **Reluctant employees**. This challenge occurs when employees are not willing to cooperate with a process mining project.

- **Complex process**. This challenge occurs when the process is so complex that its analysis becomes a challenge.

- **Low granularity of timestamps**. This challenge occurs when the granularity of timestamps is so low that it becomes a challenge to look at a process from the time perspective.

3.2 Interview

The systematic literature review gives a scientific view of the challenges and opportunities of process mining. These findings are validated through interviews. They will be conducted with domain experts to find out what challenges and opportunities they experienced when applying process mining. A semi-structured interview is chosen as a research method as it allows for fixed questions about the SLR outcome and more open questions about their own experience with process mining. This approach is chosen as it allows for adaptation to each domain expert, as well as follow-up questions that can help to clarify why certain challenges and opportunities are specific to that domain [36]. The data gathered from the interviews will be used for qualitative analysis.

3.2.1 Expert profile

The participants of the interviews are selected based on their experience with process mining applications. At least one expert will be found for each domain. They must have been involved with at least two process mining projects. Experts can have experience in multiple domains. This is beneficial as they can compare opportunities and challenges between the domains. Table 3.4 shows the experts, the domains that have experience in, and if they applied it as a consultant or for research.

Participant	Domains	PM Experience in
Expert 1	Finance, Logistics, Manufacturing	Consultancy
Expert 2	Finance, ICT, Manufacturing	Consultancy
Expert 3	Healthcare, Finance, Logistics	Research and Consultancy
Expert 4	Manufacturing, Finance	Research and Consultancy
Expert 5	Finance, Logistics	Consultancy
Expert 6	Manufacturing, Finance, Logistics	Consultancy
Expert 7	Finance, Healthcare	Consultancy
Expert 8	Education, Healthcare	Research
Expert 9	Finance, Healthcare, Manufacturing, Logistics	Research and Consultancy

Table 3.4: Domain Experts

3.2.2 Interview setup

The interview consists of three main parts, the experience of the experts, a discussion of the SLR outcome, and a comparison of multiple domains. The protocol used for the interview can be found in appendix A. Prior to the first part, a short introduction to the research is given as the experts can give more useful answers knowing what the research is about. The introduction should not guide the expert to certain answers. Therefore examples of challenges and opportunities can not be given. The experts are asked to answer the questions from the perspective of their domain. Meaning that their own experience, but also the experiences of colleagues from the same domain can be used. If an expert has also worked in other domains, they should clearly state what domain they are talking about when answering the questions.

In the first part of the interview, the experts are asked about their experience with process mining within their domain(s). They are asked what opportunities process mining brought them, and why they applied it. They are also asked if they ever ran into challenges whilst applying process mining, and how they dealt with them. If they have experience in multiple domains, it is asked how these opportunities and challenges compare between domains. These questions are asked before sharing the SLR results to get an unbiased insight into the experiences of the experts, possibly resulting in new opportunities and challenges.

The second part of the interview is about the outcome of the systematic literature review. This outcome could contain opportunities and challenges that have not been mentioned in the first part. The experts are asked if they have experience with these opportunities and challenges and how they view them. If the SLR has an outcome that contradicts the experience of the expert, this is discussed to find out where the difference comes from.

The third part of the interview compared the domains. If experts have experience with multiple domains, questions are asked to compare the opportunities and challenges of those domains. After this part, some finishing questions are asked to conclude the interview.

3.3 Combining results

The interviews are recorded and transcribed. These transcriptions are used to create a spreadsheet with all opportunities and challenges, whether they have been confirmed or denied, and by which expert. It also includes a list of interesting comments by experts. Newly found opportunities or challenges are added as well. This list is used to validate the SLR outcome. The results of both the SLR and site interviews are compared and discussed in the Chapter 4.

3.4 Recommendation Infographics

To answer RQ5, the outcomes of the SLR and the expert interviews are combined to create six infographics, one for each domain. These infographics will show inexperienced process miners what opportunities they can expect, and what challenges can occur when applying process mining in their domain. It will recommend techniques and methods that are suited to overcome specific challenges. These techniques and methods will be found by searching existing literature and the papers selected from in SLR. They can also come from the experience of the experts in the interview, as they are asked how they dealt with the challenges they faced.

Chapter 4

Results

This section presents the results of the systematic literature review and the expert interview. It aims to answer RQ1 and RQ2 by analyzing the papers that resulted from the first selection round. Their titles and abstracts have been analyzed to look into the types and perspectives of process mining that have been used. They have also been filtered for the second selection round, resulting in 117 papers. To answer RQ3 and RQ4, these papers have been analyzed on the opportunities and/or challenges that they mentioned. The number of papers per round, per domain has been given in table 3.3 of the previous chapter.

4.1 SLR Results - Types and Perspectives

The three types of process mining are discovery, conformance, and enhancement. Each type has a different goal. Discovery aims to construct a process model, conformance compares existing models of expected behavior with models of observed behavior, and enhancement aims to extend or improve a process based on undesired behavior. Whilst these are three different types, they overlap. If you want to check conformance, you also need to discover the as-is process, and if you want to enhance a process, you could need to check the conformance to find out what needs to be improved. Because of this overlap, the analysis of types of process mining used looks at the main application of that paper. A paper on the improvement of a financial audit is classified as enhancement, even though the process also had to be discovered before it could be enhanced. However, some papers are focused on multiple types. For example, a paper that analyzed a course for conformance in one chapter, and looked for possible enhancements in another. Because of this, some papers have been classified as multiple types. The same goes for the perspectives, being control-flow, organizational, case, and time. Some articles looked for process improvements from multiple perspectives, like time and organizational. They have been classified with multiple perspectives. There were also a few papers that could not be classified. For example, one paper looked into repairing missing event logs without applying process mining [100]. When it is not applied, the type and perspective cannot be determined. The classification of the types and perspectives is based on the titles and abstracts of the papers. When this did not provide enough information, the paper was scanned until the type and perspective were confirmed.

For the healthcare domain, an existing mapping study already looked into the types and perspectives of process mining within that domain [21]. These results were reused with some slight alterations. In their results, they also included an 'other' section. This was removed as papers without types or perspectives were not taken into account when analyzing other domains in this research. The mapping study also made more detailed distinctions between perspectives like *Predictive analysis for case, disease, and care management* and *Deviation detection - outliers.* Both belong to the case perspective and have been combined for this research. A division of the types per domain can be seen in table 4.1. The table shows the absolute number of articles that contained the process mining type. Figure 4.1 shows the same results in percentages. The same can be seen for the perspectives in table 4.2 and figure 4.2.

The data in table 4.1 and figure 4.1 can be used to answer RQ1: *What types of process mining are being applied in the different domains?* On average, the most used type of process mining is discovery, being used in 52% of papers. Enhancement comes in second with 30%, followed by conformance with 18%. The results show little variation between the domains. When looking at types, the use of discovery stands out for the educational domain. Many papers looked into the behavior of students to understand their behavior. Some papers did look to improve the learning process, but most only looked into the behavior of students. This could be why education uses more discovery and less enhancement compared to the average. What also stands out is that logistics had a higher percentage of conformance compared to the average. This could be explained by the complex nature of logistical processes. A deviation is more likely to occur in a large, complex process. Checking the conformance to find these deviations could therefore be more logical in complex logistical processes, compared to other domains.

	Discovery	Conformance	Enhancement
Healthcare	97	47	67
ICT	13	5	11
Manufacturing	60	19	35
Education	59	12	19
Finance	11	3	6
Logistics	18	9	10
Total	258	95	148

Table 4.1: Number of papers containing process mining types per domain

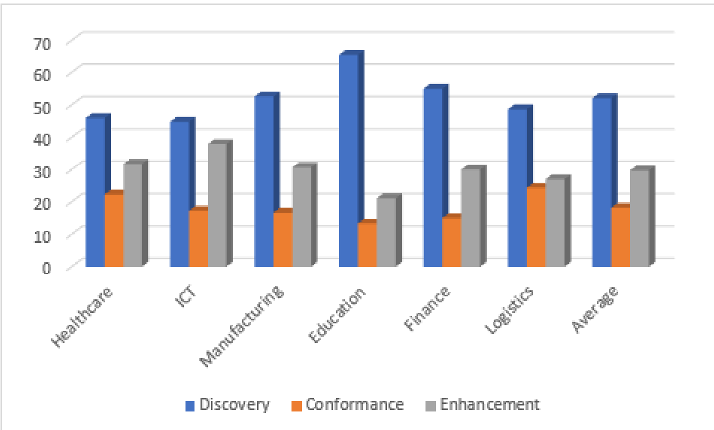


Figure 4.1: Process mining types per domain in percentages

The data in table 4.2 and figure 4.2 can be used to answer RQ2: *What process mining perspectives are used in the different domains?* The most used perspective is control-flow, covering 45% of cases. 28% uses process mining to look from an organizational perspective, followed by time with 18%, and case with 10%. Comparing the perspectives of individual domains, it stands out that manufacturing, finance, and logistics are relatively focused on the time perspective. This could be explained by the large financial benefits that result from lowering process times in those domains. Education is more case-focused as a number of papers researched the specific characteristics of students that resulted in a dropout or failing a course. They also have a lower focus in the time perspective. Possibly because education is more focused on the quality of learning, not on the speed of learning.

	Control-Flow	Organizational	Case	Time
Healthcare	36	26	6	14
ICT	17	9	3	4
Manufacturing	57	30	15	33
Education	47	33	19	8
Finance	13	8	3	6
Logistics	19	13	2	11
Total	189	119	48	76

Table 4.2: Number of papers containing process mining perspectives per domain

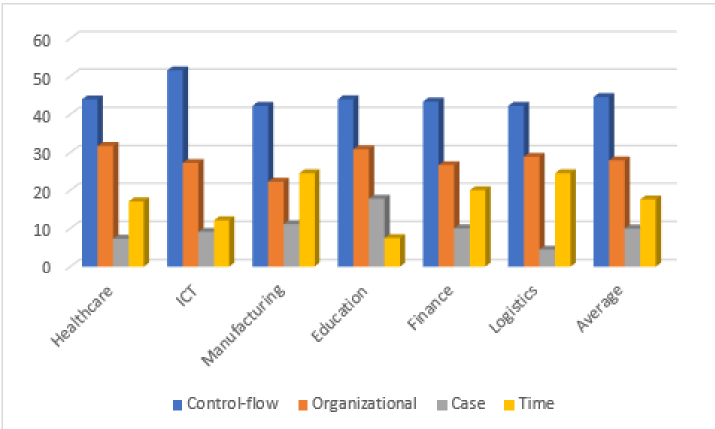


Figure 4.2: Process mining perspectives per domain in percentages

4.2 SLR Results - Opportunities and Challenges

This section aims to answer RQ3: *What are the major opportunities when applying process mining in each domain?* and RQ4: *What are the major challenges when applying process mining in each domain?* by showing the results of this analysis per domain. It does this by analyzing the 117 papers that were selected for the second round of the SLR. The opportunities and challenges mentioned in these papers are analyzed as described in section 3.1.3. Specific opportunities and challenges have been generalized to make it possible to compare them between domains. If two opportunities are in the same category this does not mean they are the same, there are still domain-specific nuances. For example, tracking the performance of a process could look at a machine in manufacturing or a student in education. All nuances are explained per domain in the sections below. An overview of all opportunities and challenges that resulted from the SLR can be seen in table 4.3 and table 4.4, with references to the papers that mentioned them.

Opportunities	Healthcare	ICT	Manufacturing	Education	Finance	Logistics
Optimize the process (Cost and time)	[21]	[6][136][34] [147][83] [43][57]	[72][2][19] [35][12][75] [116][118][110] [123][92] [70][86]	[88][51][98][10] [3][76][41][25] [89][24][32][42] [115][125][17]	[5][73][145] [14][117][143] [140][50][94] [141][37]	[132][9][96] [144][58][106] [59][56][53] [134][135][67]
Optimize the process (Quality of result)	[21]	[6][66] [7][39]	[87] [29] [69][68]	[82][31][16] [32][91]	[138][137][140]	
Check conformance with regulations	[21]		[29][68][69]	[98][120][111]		[62][126][56]
Analyze communication		[128]				
Track performance			[61][113][107] [60][87] [118] [70][55] [11][109]			[95][49][126]
Identify and predict fraud cases					[14][74][141]	

Table 4.3: Opportunities per domain

4.2.1 Healthcare

The systematic literature review for the healthcare domain was simplified due to an already existing mapping study [21]. This mapping study analyzed papers on process mining in healthcare and summarized the opportunities and challenges that depict the domain. The inclusion criteria used are comparable to the criteria for the SLR in this thesis. All papers conducted from the SLR that have been published after the mapping study have been analyzed on the inclusion criteria for the second round. Seven of those papers ended up containing opportunities and/or challenges that depict the healthcare domain. However, all have been mentioned in the mapping study, meaning no new opportunities or challenges have been found in later literature.

Opportunities In healthcare, a process can be optimized with two goals in mind. Firstly to reduce the cost and time of the process. This can be done by finding bottlenecks, such as long

Challenges	Healthcare	ICT	Manufacturing	Education	Finance	Logistics
Handling data from heterogeneous sources	[21]	[136]	[113][12][101] [112][27]	[124][93]	[18][50]	[95][49][59]
Low data quality	[21]	[136][147][54] [83][133][133]	[35][12][60]	[51][99][90] [111][24]	[74][18][140] [50][121]	[77][96][58] [106][59][135]
Handling variable data	[21]			[124][24]		
Handling private data	[21]	[15]	[84]			
Lacking visualizations		[7]		[51][76]	[138][140] [141][37]	[77]
Large computational costs		[46]				
Reluctant employees			[119]			
Complex processes				[124][3][24]		[49][97][59] [8]
Low granularity of timestamps					[94]	[22][58][59]

Table 4.4: Challenges per domain

waiting times for a simple approval, or a waste of resources. Whilst reducing cost and time also benefits the patient, process mining can also help to improve the quality of the medical care itself. Recognizing patterns in medical conditions can help doctors by suggesting diseases that were found in patients with similar conditions. Whilst these results are by no means medical advice, it could help to find a solution quicker. The quality of a medical process could also be improved by lowering waiting times for appointments.

Challenges The challenges in the healthcare domain are all data-related. The sector is characterized by the use of many different systems, each with its own way of logging data. This results in varying data formats that are difficult to combine. However, combining this data is necessary to get a complete picture of the patient flow. Applying process mining in healthcare therefore involves a lot of pre-processing. This pre-processing is also needed for another challenge, which is the low data quality. Using low-quality data for process mining can lead to unreliable results that could be dangerous when applied in healthcare. The privacy of the patients is also considered a challenge. Medical data is private and sensitive, thus handling this data must be done with the patient's privacy in mind. The last challenge is the variability of patient data. All patients have different bodies that behave and react in different ways. Two patients with the same condition might get different treatments based on their circumstances. This results in many different patient flows, creating spaghetti models that make it difficult to recognize patterns.

4.2.2 ICT

For the ICT domain, the SLR resulted in 17 papers that contain opportunities and/or challenges. Most are shared with other domains, but there are some that characterize the ICT domain.

Opportunities The opportunities in the ICT domain mostly speak for themselves. Process mining applications are used to analyze how a process can be improved regarding cost and time [6][136][34][147][83][43][57], or quality [6][66][7][39]. The quality aspect regards the users that use the ICT systems. Data is mined on how users interact with a system, which can be used to introduce quality changes. An opportunity that is specific to the ICT domain is to analyze the communication between software components to check conformity with the desired behavior [128].

Challenges The ICT domain has some data-related challenges. Low-quality data, specifically noisy data and missing timestamps, make it difficult to get accurate results from a process mining project [136][147][54][83][133][133]. The data can not only be low quality but also be different in formats. Just like in other domains, this can occur when data is gathered from heterogeneous sources [136]. One paper focused on the privacy of data, and how it hinders process mining applications [15]. As they are mostly performed by external parties, the companies have to share their data with them. Trusting other parties with your confidential or private data, along with the legal complications of sharing it can cause some companies to step away from process mining analyses. The scalability of ICT projects also poses a challenge, as large data sets can take many resources to analyze. Both in terms of time and computational costs [46]. It was also mentioned that the visualizations to present process mining results are lacking, making it difficult to share the outcome with users without a process mining background [43].

4.2.3 Manufacturing

In the manufacturing domain, 30 papers have been found that contain opportunities and/or challenges.

Opportunities The largest opportunity in manufacturing is the optimization of a process. An optimization can be in terms of production costs [35][123], manufacturing speed [2][110][92][86], and reducing waste materials [19]. But also a combination of these [72][12][19][75][116][118][70]. It can also be optimized to produce a higher quality product [29][87][68][69]. Another opportunity that eventually leads to an optimized process is the prediction of machine failure to prevent failure [55][11][109]. By tracking the performance of machines, it can be predicted when they need repairs. Parameters can be tracked that might indicate wear, but the total running time of a machine can also be an indicator. Planning the maintenance is also part of this opportunity. Downtime is never optimal, but certain timings have less effect on the production. Maintenance of multiple machines can also be simultaneous, reducing the total downtime of a factory. Tracking KPI's for the performance and conformance of a manufacturing process can show if it still produces the desired results [61][113][107][116][118][70], and if it complies with regulations [29][68][69].

Challenges Just like all other domains, data coming from heterogeneous sources is a challenge in manufacturing, as multiple machines with different logs are involved in a process [113][12][101][112][27]. The same goes for low data quality [35][12][60]. Data is not always suited for process mining and its availability is limited. A different challenge that is not seen in other domains is the reluctance of employees toward process mining initiatives [119]. When employees are ill-informed, they could fear that the analysis results in a change to the workflow they are accustomed to. In worse cases, they fear that their work will be automated. In lesser cases, they feel watched and hindered in their privacy. Hence the challenge of handling private data in this domain [84]. Luckily a case study showed that employees are no longer reluctant to process mining applications once it is explained to them [119]. Being ensured that their work will not be replaced and that workflow changes will only make their work easier is an effective way to counteract this challenge. Proving that the data is anonymized and cannot lead back to individual employees helps with their privacy concerns.

4.2.4 Education

For the education domain, a total of 27 papers have been found that contained opportunities and/or challenges.

Opportunities When process mining is applied to analyze the learning behavior of students, it can be on a small or large scale. The large scale analyzes the student's academic career. Here, process mining can help to track the progress of students and recognize patterns that can lead to dropouts or course failure [3][111][24]. The small scale looks at student behavior within one course or assignment. It can analyze how students make assignments or tests to see what questions are difficult, but also to check when students guess the answer [91]. The career of a student can be optimized in terms of time and costs [88][51][98][10][3][76][41][25][89][24][32][42][115][125][17] by identifying bottleneck courses that have high failure rates and can lead to graduation delays[16]. The quality of that course can then be improved to prevent the bottleneck [82][31][16][32][91]. It can also be used to check course compatibility with regulations and guidelines [98][120][111].

Challenges Similar to healthcare, educational process mining is largely focused on analyzing human behavior. Because each human is unique, the learning processes are highly variable [124][24]. Even when two students follow the same set of courses, they can be taken in a different order, further adding to the variability. Most students take multiple courses at once, resulting in concurrent processes that are complex to analyze [124][3][24]. The domain also struggles with data coming from multiple sources [124][93]. A standard for collecting data is clearly missing, resulting in different data formats that take time to combine. As in the other domains, the data needs to be of high quality in order to get reliable results. Low-quality data therefore poses a challenge [51][99][90][111][24]. Other domains also talk about the lack of good visualizations to present the models. With the high variability of educational data and the spaghetti models it produces, a good visualization is difficult to achieve [51][76].

4.2.5 Finance

Regarding the financial domain, 16 papers have been found that contain opportunities and/or challenges.

Opportunities The application of process mining on financial processes presents three opportunities. Firstly, it is used to identify fraud cases [14][74][141]. By analyzing existing fraudulent and regular cases, the likelihood of a new case being fraudulent can be predicted. Besides the prediction of fraud cases, financial processes such as loan applications are also analyzed. This can detect bottlenecks or inefficiencies that can be optimized to save costs and time [5][73][145][14][117][143][140][50][94][141][37]. Saving time also improves customer satisfaction as they receive their results faster [138][137][140].

Challenges Loan applications are processes that involve a bank and a client. These two parties both perform actions, but only the system of the bank produces logs accessible to the bank. When the granularity of timestamps is too low, for example when only the end time of an activity is logged, bottlenecks on the bank's side can be mistaken for late replies from clients [94]. Whilst low granularity might also occur in other domains, the two-way communication between bank and client turns it into a challenge. As in some other domains, financial institutions also use multiple systems that produce logs in different formats [18][50]. Lacking visuals to present the results of the process mining project has also been mentioned as a challenge [138][140][141][37], along with low data quality [74][18][140][50][121].

4.2.6 Logistics

When it comes to logistics, 20 papers have been found that contain opportunities and/or challenges.

Opportunities The Logistic domain shares similarities with manufacturing. Some papers even discussed both domains as they have comparable characteristics [49][8]. All logistical opportunities are also present in manufacturing. Process mining is used to track how a process performs, searching for changes that have a negative impact on its performance [95][49][126]. The complex logistical models can be analyzed to find bottlenecks that can help whole supply chains when removed, resulting in time and cost savings [132][9][96][144][58][106][59][56][53][134][135][67]. Conformance with regulations can also be checked by logging logistical data [62][126][56].

Challenges The challenges in logistics are a bit more unique due to the complex and fluctuating nature of its processes [49][97][59][8]. They can involve multiple systems and can change over time. This can result in complex models that are difficult to understand. Combining the data from these different sources also poses a challenge [95][49][59]. The low granularity of timestamps is seen as a challenge in the logistical domain as some supply chains are scheduled by the minute. Not knowing when exactly a transport is finished hinders this tight scheduling [22][58][59]. Lacking visualizations representations of results [77] and lower data quality have also been mentioned [77][96][58][106][59][135].

4.3 Interview Results

To validate the outcome of the SLR, interviews were conducted with process mining experts. This section discusses the results of the interviews. The setup of the interview was discussed in section 3.2. This section also contains table 3.4, showing the experts, along with the domains they have experience with.

The experts have been asked about their experience with the opportunities and challenges that resulted from the SLR. They agreed with each other in most cases, confirming or denying the opportunity or challenge in their domain. There are however some interesting cases where their options contradict, and some experts came up with new challenges that have not been found in the SLR. As some experts had experience in multiple domains, it was possible to compare challenges between domains. The sections below give further explanations of the interview results.

4.3.1 Healthcare

With regards to the challenges in the healthcare domain, there is one confirmed challenge that is also seen in most other domains, being that data often comes from heterogeneous sources. A comparison of this challenge between the domains is discussed in section 4.3.7. Another data-related challenge that was confirmed is the variety of the data. As patients can have many factors that influence their treatment, it becomes difficult to generalize their treatment processes. The challenge where patient data is difficult to acquire due to privacy regulations was also confirmed. Expert 1 (with experience in finance, logistics, and manufacturing), noted that this is most difficult when projects are still a proof of concept. When they become full projects supported by management, it becomes easier to access data. This could also be true in healthcare. As for opportunities, it was confirmed that process mining is used in healthcare to optimize cost, time, and patient care and also to check compliance with regulations.

4.3.2 ICT

Only one expert was found with experience in the ICT domain, though its main expertise is in finance. This expert did not encounter most challenges but was not experienced enough to deny they existed. The process mining projects of the expert were not big enough to experience large computational costs when using large datasets. The other challenges were acknowledged, but not solely in ICT. These challenges could therefore not be confirmed as ICT-specific. The expert never used process mining to analyze communication between software components. This does not rule out its use, but this opportunity was neither confirmed nor denied. Optimizing the process in terms of cost, time, and quality was acknowledged, but this was also seen in other domains and not confirmed as ICT-specific.

4.3.3 Manufacturing

The confirmed challenges in the manufacturing domain are on handling data from heterogeneous sources, and the low quality of that data. Whilst literature hinted that some employees might be reluctant to cooperate with a process mining project due to a fear of being replaced, experts 1 and 3 did not agree with this standpoint. In their experience, process mining is not being applied to track the performance of individual employees, and process changes more often lead to different or optimized workflows rather than layoffs. Because process mining is not applied to track individuals, there is no need to store private data about employees without anonymizing it. In addition, most data in this domain is produced by machines, so working with private data in manufacturing is not seen as a challenge by the experts. The opportunities in manufacturing have been confirmed, as process mining is being used to track the performance of machines and predict failure. Processes are also being analyzed to optimize them and to ensure conformance with regulations.

4.3.4 Education

The typical challenges of process mining in education that have been found in literature and confirmed by experts are mostly data-related. Data is spread out between varying systems and collecting and combining this data poses a challenge. Whilst the data from heterogeneous sources is a challenge, expert 8 says a standard for data gathering is missing. Combining data from multiple sources would not be as big of a challenge if all systems followed the same standard. This standard is not just missing between, but also within educational institutes. The complex nature of the data itself is also a confirmed challenge. When it comes to analyzing data on courses and study paths, the processes can have concurrent activities, for example when a student follows multiple courses in the same period. Sometimes two students can take the same courses, but in a different order, or have mostly the same program with a difference of one or two courses. This leads to high variability in the learning processes, as many unique combinations can be made. A new challenge that was not found in literature is oversimplification, brought up by expert 8. This challenge became clear from the interviews. Oversimplification in this domain occurs because conclusions are drawn based on human actions. For example, when a student fails a course, it is easy to assume that the student struggles with the materials of that course. However, there can also be other unpredictable reasons that can lead to failing a course such as traumatic experiences or illness. Low-quality data and visualizations are present in education but are not seen as a domain-specific challenge according to expert 8. All opportunities that were found in the SLR have been confirmed by experts. Process mining is used in education to track the performance of students. The behavior of students during exams and exercises is analyzed to improve the questions. Their behavior is also analyzed through-out their academic career to recommend courses and prevent dropout. It is also used to check if a course still complies with the guidelines and regulations to ensure the quality of the course.

4.3.5 Finance

Experts 2, 4, and 5 confirmed that data coming from heterogeneous sources is also a challenge in the financial domain. A comparison of the domains regarding this challenge is given in the section shared challenges. The challenge where a lack of timestamps makes it impossible to determine if an activity is delayed or waits on a response from the customer is acknowledged by the experts but was not seen as a big issue. A new challenge in the financial domain that became clear from the interview with expert 4 is that legacy systems are still in use. Retrieving data from these older systems can be difficult as some systems were developed in the 80s, before process mining existed. The experts do not agree with each other when it comes to data privacy. Experts 6 and 8 argued that privacy regulations made it more difficult to acquire data. Experts 4 and 7 on the other hand never encountered this problem, as personal data is often not needed for analysis. There is no need to use personal data, as this does not add value to the analysis of a process. Cases of the same person can be linked with a case ID. As for opportunities, it was confirmed that process mining is used in the financial domain to identify and predict fraudulent cases. Analyzing loan applications to improve their cost and quality is also confirmed.

4.3.6 Logistics

In the logistics domain, the challenge of data coming from heterogeneous sources was recognized by all domain experts. Another challenge that was confirmed is the low granularity of timestamps. For example, only after an activity has been completed was a timestamp logged. This makes it difficult to calculate the activity time. An easy fix would be logging a timestamp when an activity starts, and one when it is completed. Expert 6 suggests this is not being logged in systems as the logs are not created with process mining in mind. Changing the way processes are logged can be a big ask for the person who executes a process mining project. As they are often a third party that is asked to analyze a process, the company that hires them is not always willing to change their systems just for a process analysis. It was also confirmed that logistical processes can be quite complex. For example, transporting a container can take months and could take over 300 activities to complete. Analyzing processes of this size can be challenging. The challenges of low data quality and lacking visualizations were not seen as domain-specific. When it comes to opportunities in the logistical domain, process mining is being used to reduce export time and determine the performance of processes. It is also confirmed that process mining is used to track conformance with regulations in logistical processes.

4.3.7 Shared challenges

Heterogeneous data
Data that comes from heterogeneous sources, also known as data triangulation, is a challenge that occurs in all domains. This makes it a challenge that is more general for process mining, but there are still domain-specific nuances. Some experts with experience in multiple domains compared this challenge between them. They agreed that logistics suffers the most from this challenge. Expert 6 explained that a logistical process, like shipping a container, is a large process that can consist of up to 300 steps. These steps are performed by various systems, as a container is handled by multiple companies. Compared to a financial process, which often has smaller high-frequency processes,

it is likely that logistics suffers more from data triangulation. Expert 8 compared this challenge between the educational and healthcare domains, where healthcare seems to struggle more with it. The reasoning was that the healthcare sector has a larger focus on saving costs, resulting in more competition and therefore, more different systems. Healthcare systems are also updated more often, resulting in more changes in the log output.

Handling private data

Handling private data poses a challenge in multiple domains. Not only because the data should be treated with care, but also because it can be difficult to access. Privacy in healthcare, ICT, and manufacturing have already been discussed. Comparing these domains, it is clear that healthcare deals with this challenge more than the others. Patient data is private and sensitive. Getting approval to use the data for process mining is therefore a difficult and time-consuming task. In manufacturing some thought needs to be put into minimization, but most data is produced by machines and does not contain private information. In ICT the challenge was not seen as domain specific. A few experts also started to talk about privacy in other domains. Whilst not being a challenge according to the SLR results, experts in finance had some interesting thoughts on it. Experts 6 and 8 argued that the financial data of customers can be difficult to access, as this is private data that can have big consequences when leaked. Experts 4 and 7 did not disagree with the importance of the data, but in their experience, financial data was anonymized to the extent that it did not pose issues. Data containing private information is not needed to get the desired process mining result. Combining these arguments, handling private data is a challenge in the financial domain. It can however be overcome by anonymizing the data, as is already being done in some organizations.

Low data quality and Lacking visualizations

When it comes to low data quality, no clear distinction between domains can be made. Expert 4 suggests that older systems produce lower-quality data, but this is not always the case. This challenge seems to be a more general challenge for process mining that applies to all domains. The same goes for the visualizations of results. Sharing these results with management is a challenge that experts have experienced, as it is difficult to present the results so that they can be understood by people without process mining experience. They did not see this as a challenge that differs per domain.

Chapter 5

Infographics

The objective of this research is to make new process mining users aware of the opportunities and challenges that can arise when applying process mining by listing the relevant opportunities and challenges that are specific to the six main application domains. The systematic literature review and the expert interviews resulted in a detailed answer to this objective. This chapter aims to answer RQ5: *Considering the major opportunities and challenges in each domain, what should be taken into account when applying process mining?* by visualizing those results. It does this by creating an infographic per domain that summarizes the opportunities and challenges. Its goal is to inform inexperienced process miners about the major opportunities that process mining can bring to their domain, as well as the challenges that they can encounter. Recommendations are made to deal with those challenges. The infographics can be seen on page 27. The sections below will provide reasoning behind certain recommendations that have been given in the infographics.

Recommendation for private data

The experts had mixed opinions on the topic of data privacy. There is an agreement on its importance, but it does not always lead to a challenging situation. When it does, accessing the data is difficult because of regulations on sharing private data, but also because the data must be handled with care to protect people's privacy. Both challenges can be mitigated by anonymizing the data. When the dataset cannot be used to get information of individuals, their personal information is protected. Experts also said that data is more accessible when you only request the data you need. Often, non-anonymized personal data is not needed for a process mining analysis, as this mostly looks at the whole process. This is a general recommendation, as all domains could encounter personal data.

Recommendation for complex data

The logistics domain needs to deal with large, complex, dynamic, and noisy processes. The challenge of dealing with this data can be mitigated by applying certain methods. Using a clustering method can help with the complex and dynamic nature of the data [49]. Clustering puts traces that behave in a similar way or show similar patterns in one group. By doing this, multiple slightly different flows can be seen as one. This lowers the number of different flows, making the model less complex. It can also help with the identification of concept drift [97]. To deal with the noise that is often in logistical data, a miner heuristic, genetic, or fuzzy miner can be used as these can handle noisy data well. For logistics, a fuzzy miner is recommended as can handle complex and dynamic processes better than a heuristic or genetic miner [106] [64].

Dealing with complex data is also a challenge in healthcare and education, especially the variety of the data. A clustering method can also be applied in these domains to get more structured results [4]. A fuzzy miner is also effective in these domains when dealing with complex data [91].

Recommendation for low quality data

Improving the quality of data can be done by pre-processing it. This takes time but is needed to get accurate results [74][51]. The *garbage in = garbage out* principle applies here. When you use low-quality data for process mining, the result will be of low quality [99]. The data quality can also be improved by changing the way in which systems log data. This can be difficult to change and only impacts newly generated data, but it could be worthwhile for continuous process mining applications.

The Opportunities and Challenges of Process Mining in Healthcare

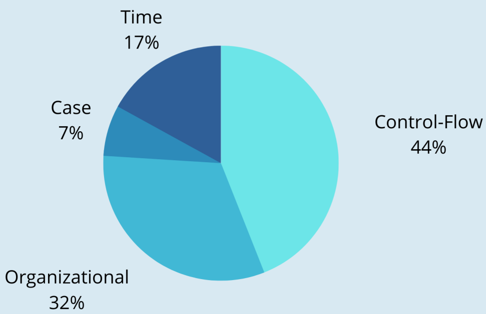
The healthcare domain covers clinical paths, patient treatment, and the primary processes of a hospital. Process models in this domain are characterized by high variability, complexity, security, and privacy.

Opportunities

Just as in other domains, the application of process mining in healthcare can help to optimize a process in terms of cost and time by analyzing the process and determining bottlenecks. More specifically for the domain, it can also help with identifying flows of patients with certain diseases, estimating chances of treatment, finding correlations among treatments, and improving the quality of treatments. Process mining can also be used to check if medical processes and procedures are following the safety regulations correctly.

Perspectives

A process mining perspective is the aspect of that process that is being analyzed. The control-flow perspective looks at the order of the events (or steps) of a process. The organizational perspective is about the actors involved. The case perspective looks at the properties of a case. And the time perspective looks at the frequency and timing of events.



The same can be said about the perspectives used in healthcare. The division of perspectives is comparable to the average of process mining in general. The case perspective is used quite little as it often needs private information. With permission from the patient, it can lead to valuable insights though. To use the time perspective, you must compare timestamps. Remember the challenge of data from different systems, they can produce timestamps in different formats.

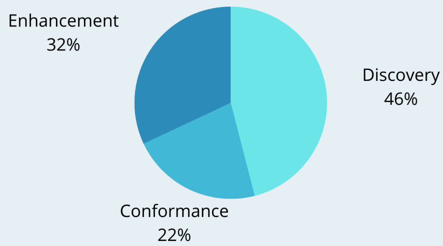
Challenges

Hospitals and other medical institutions need multiple systems to operate. These systems all produce their own logs, often in different formats. When multiple systems are involved in one process, it can be challenging to combine the datasets. Make sure to properly combine the datasets before analyzing them, or if possible, try to change the systems to produce the same logs.

Not only the systems are different. The patients are also unique, resulting in variable data. Every patient has different factors that influence their health. Making generalizations can be difficult, as it is impossible to log all factors. Keep in mind that each patient is different and could be influenced by unknown factors. Be cautious and take time when drawing conclusions. Always do so in association with a medical professional. To deal with the variety of patient data, you can use a clustering method and a fuzzy mining algorithm to help with generalizations. Also, be aware that accessing data can take longer as medical data is private and sensitive. Personal data is not always needed for analysis. Asking for anonymized data or data without personal information can help getting access.

Types

There are three types of process mining that can be applied. Discovery aims to construct a process model based on an event log. Conformance makes a comparison between the intended behavior of a model and the actual behavior. Enhancement aims to improve a process based on information from the event logs.



Whilst discovery is the most used type in healthcare, it is used a bit less compared to other domains. The healthcare domain uses conformance and enhancement relatively more often on the other hand. All types are applicable in healthcare. To choose one, think about what you want to achieve with your process mining application.

The Opportunities and Challenges of Process Mining in ICT

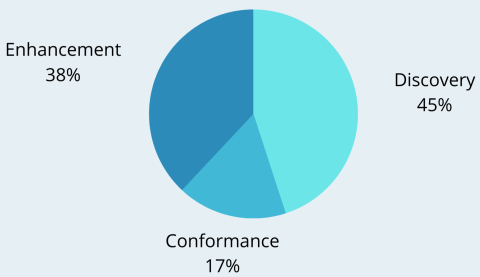
The Information and Communication Technology (ICT) domain covers process improvement, evaluation, and agile task management, on IT systems.

Opportunities

In this domain, process mining can be used to improve processes in terms of cost, time, and quality. It can also evaluate and analyze processes between systems that communicate directly with each other. As no human is involved with this interaction, process mining can be used to check if this communication is being performed as planned. It has also been shown capable of improving the maturity levels of software processes, and it is used to analyze and improve maintenance processes using bug report logs.

Types

There are three types of process mining that can be applied. Discovery aims to construct a process model based on an event log. Conformance makes a comparison between the intended behavior of a model and the actual behavior. Enhancement aims to improve a process based on information from the event logs.



Whilst conformance is the least used type in ICT, its percentage is similar to the average of all process mining domains. Discovery is used the least compared to other domains, and enhancement is the most used type when compared. The process behind an ICT system is often thought out before the system is built. As process is already known, making enhancements to the process more attractive. If you want to know how a system is intended to operate, you can often ask the designers.

Challenges

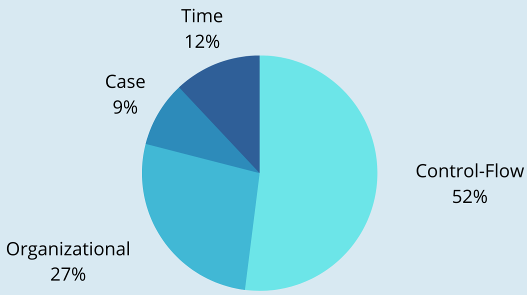
The ICT domain has some data-related challenges. Low-quality data, specifically noisy data and missing timestamps, make it difficult to get accurate results from a process mining project. Cleaning and preparing the data before analyzing it will provide better and more accurate results.

The data can not only be low quality but also be different in formats. Just like in other domains, this can occur when data is gathered from multiple sources. Combining data with different formats can be challenging, but taking the time to do so will help you later on in the analysis. If possible, try to change the systems to produce the same logs.

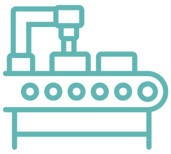
ICT systems can produce large datasets that take up much processing time to analyze. You can save time by using a subset of the data for some quick tests. Once you have dialed in your settings, you can analyze the full dataset.

Perspectives

A process mining perspective is the aspect of that process that is being analyzed. The control-flow perspective looks at the order of the events (or steps) of a process. The organizational perspective is about the actors involved. The case perspective looks at the properties of a case. And the time perspective looks at the frequency and timing of events.



Of all process mining domains, ICT looks the most at the control-flow. Finding out how software processes run can help to understand them better. The organizational perspective looks at the actors involved. In ICT, these actors can also be systems.



The Opportunities and Challenges of Process Mining in Manufacturing

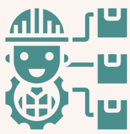
The manufacturing domain covers processes used in the creation of products. They often take place in factories and can combine machinery, as well as factory employees.

Opportunities

The main opportunity in manufacturing is the optimization of a process. An optimization can be in terms of production costs, manufacturing speed, and reducing waste materials. Another opportunity is the prediction of machine failure to prevent failure. Parameters can be tracked that might indicate wear, such as the total running time of a machine. This makes it possible to plan maintenance at the best time. Downtime is never optimal, but certain timings have less effect on the production. Maintenance of multiple machines can also be simultaneous, reducing the total downtime of a factory.
KPI's can be tracked to determine if the performance of a manufacturing process is still on par. Other parameters can be checked to ensure that the process adheres to regulations.

Challenges

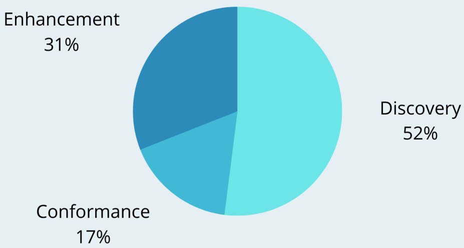
Manufacturing processes involve multiple machines that can produce different logs. Combining data with different formats can be challenging, but taking the time to do so will help you later on in the analysis. If possible, try to change the systems to produce the same logs.



When analyzing a process that involves employees, make sure to inform them of the goals of the analysis and ensure the anonymity of their data. If uninformed, they might fear losing their job due to automation. They can also feel watched and hindered in their privacy. Be honest about the process mining implementation and show how it can benefit them.

Types

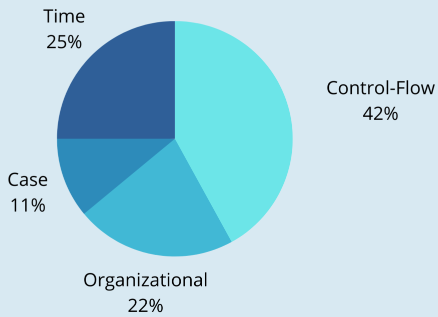
There are three types of process mining that can be applied. Discovery aims to construct a process model based on an event log. Conformance makes a comparison between the intended behavior of a model and the actual behavior. Enhancement aims to improve a process based on information from the event logs.



The manufacturing domain almost perfectly represents the average division of types of all process mining domains. Discovery is used most as it is the simplest to implement and can be used when there is no existing model of the process. Enhancement can provide financial benefits for a manufacturing company. A process can be improved by removing bottlenecks, thus increasing its efficiency. When a bottleneck is removed, it can benefit the entire process, not just that one step.
Conformance can be used to check if a process adheres to the safety regulations. Not just to ensure the safety of workers, but also to prevent fines from regulating organizations.

Perspectives

A process mining perspective is the aspect of that process that is being analyzed. The control-flow perspective looks at the order of the events (or steps) of a process. The organizational perspective is about the actors involved. The case perspective looks at the properties of a case. And the time perspective looks at the frequency and timing of events.



The perspectives show a lot less similarities to the other domains. The organizational and case perspectives are used a bit less, but this is made up by a large focus on time. Manufacturing processes are optimized to produce as many products as possible. This leaves little to no time for delays and downtime. Especially as a small delay early on in a process can hinder lead to large time losses at the end of the process.

The Opportunities and Challenges of Process Mining in Education

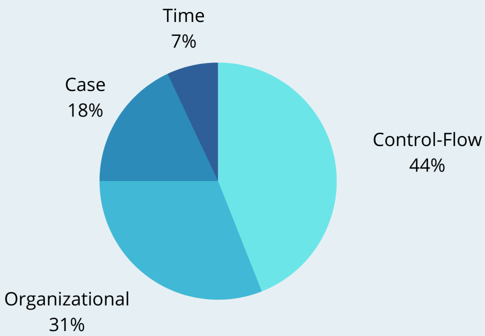
The educational domain covers processes that occur at a school or in a learning program. The focus can be on the students, but also on the learning material.

Opportunities

In education, process mining can analyze the learning behavior of students on a small or large scale. The large scale analyzes the student's academic career. Here, process mining can help to track the progress of students and recognize patterns that can lead to dropouts or course failure. The small scale looks at student behavior within one course or assignment. It can analyze how students make assignments or tests to see what questions are difficult, but also to check when students guess the answer. The performance of courses can be tracked to identify bottleneck courses that have high failure rates and can lead to graduation delays. The quality of that course can then be improved to prevent the bottleneck. It can also be used to check course compatibility with regulations and guidelines

Perspectives

A process mining perspective is the aspect of that process that is being analyzed. The control-flow perspective looks at the order of the events (or steps) of a process. The organizational perspective is about the actors involved. The case perspective looks at the properties of a case. And the time perspective looks at the frequency and timing of events.



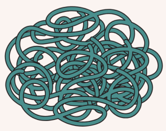
The use of the control-flow and organizational perspectives are comparable to other domains. What stands out is that education has a relatively high use of the case perspective, even the most of all domains. Looking at individual characteristics of students to explain their behavior can lead to valuable insights, but ensure the privacy of the students when you do so. Finding time optimizations is often done to save money. In educations, a time reduction is less about the money, and more about a better experience for the students.

Challenges

Educational process mining is largely focused on analyzing human behavior. Because each human is unique, the learning processes are highly variable. Even when two students follow the same set of courses, they can be taken in a different order, further adding to the variability.

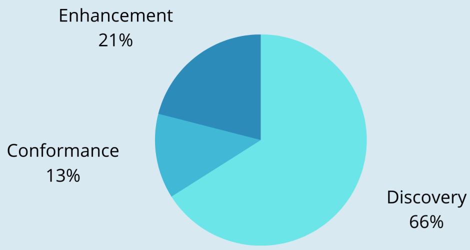
Most students take multiple courses at once, resulting in concurrent processes that are complex to analyze. The data can also come from different sources, in different formats, as a standard for collecting data is missing. Combining data with different formats can be challenging, but taking the time to do so will help you later on in the analysis. If possible, try to change the systems to produce the same logs.

The complex nature of educational data can result in messy 'spaghetti' models. These models are difficult to read as there is too much going on. It can help to generalize by clustering the data before analyzing.



Types

There are three types of process mining that can be applied. Discovery aims to construct a process model based on an event log. Conformance makes a comparison between the intended behavior of a model and the actual behavior. Enhancement aims to improve a process based on information from the event logs.



The division of types used in education deviates the most from the average compared to other domains. It has a large focus on discovery, as process mining is often used to discover how students behave. Whist this is used in two-thirds of process mining applications, this does not mean other types are less valuable. Enhancement is used to improve learning methods and conformance is applied to ensure adherence to regulations.

The Opportunities and Challenges of Process Mining in Finance

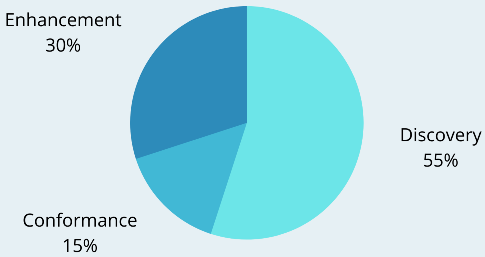
The financial domain covers processes from financial institutions such as banks and insurance companies.

Opportunities

The application of process mining on financial processes can be used to identify fraud cases. By analyzing existing fraudulent and regular cases, the likelihood of a new case being fraudulent can be predicted. It can also look at financial processes such as loan applications. Analyzing these processes can detect bottlenecks or inefficiencies that can be optimized to save costs and time. The time saving factor also improves the customer's satisfaction as they are helped quicker.

Types

There are three types of process mining that can be applied. Discovery aims to construct a process model based on an event log. Conformance makes a comparison between the intended behavior of a model and the actual behavior. Enhancement aims to improve a process based on information from the event logs.



The financial domain is quite similar to the average of all domains when it comes to the distribution of types used. Discovery is used most often to find out how financial processes actually take place. Enhancement is being used to improve those processes in terms of cost and time, but also to improve the customer's satisfaction. If the processes are already known, conformance can be used to check if they are still being performed as planned.

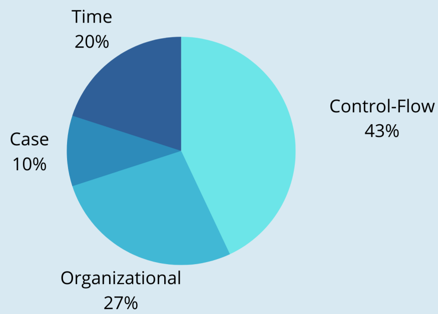
Challenges

Loan applications are processes that involve a bank and a client. These two parties both perform actions, but only the system of the bank produces logs accessible to the bank. When the granularity of timestamps is too low, for example when only the end time of an activity is logged, bottlenecks on the bank's side can be mistaken for late replies from clients. If possible, add timestamps to the start and end of an activity.

Financial institutions also use multiple systems that produce logs in different formats. Combining data with different formats can be challenging, but taking the time to do so will help you later on in the analysis. If possible, try to change the systems to produce the same logs.

Perspectives

A process mining perspective is the aspect of that process that is being analyzed. The control-flow perspective looks at the order of the events (or steps) of a process. The organizational perspective is about the actors involved. The case perspective looks at the properties of a case. And the time perspective looks at the frequency and timing of events.



As in all other domains, the control-flow and organizational perspective take up most of the financial cases. The case perspective is not used that often, despite its potential. In finance, it can be used to compare fraudulent cases with new cases to determine their likelihood of also being fraudulent.

The Opportunities and Challenges of Process Mining in Logistics

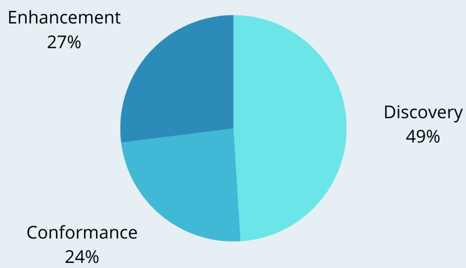
The logistical domain covers logistical processes about industrial shifts, warehouse layouts, traffic control, and resource allocation.

Opportunities

The main opportunity in logistics is the optimization of a process. An optimization can be in terms of costs and time, like shortening a delivery process. Other parameters can be checked to ensure that the process adheres to regulations, such as the driving hours of truck drivers. Some processes have bottlenecks. One activity might cause a delay that slows down the entire process. Finding and resolving these bottlenecks can result in large time and money savings. The processes can also be checked to ensure compliance with regulations. For example, the driving hours of a truck driver can be checked to ensure he takes enough breaks.

Types

There are three types of process mining that can be applied. Discovery aims to construct a process model based on an event log. Conformance makes a comparison between the intended behavior of a model and the actual behavior. Enhancement aims to improve a process based on information from the event logs.



Almost half of the process mining applications in logistics use the discovery type to find out how processes run. The other half is split up. Enhancement is used to improve the logistical processes in terms of time and costs. The conformance type is used the most in logistics compared to all other domains. In logistics, it can help to find out if processes actually run how they are supposed to. Given the large and complex nature of logistical processes, this becomes much easier with process mining.

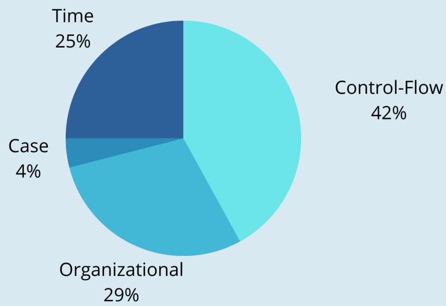
Challenges

Logistical processes are quite complex in nature. They can involve multiple systems that result in hundreds of steps. Over time, these steps or even the systems might be changed out for new ones.
Working with this complex data can be challenging. It can help to cluster the data in pre-processing and apply a fuzzy mining algorithm to deal with this. Clustering will group comparable traces, thus lowering the number of traces and making it more overseeable.

As some systems only log when an activity is completed, it can be hard to determine if an activity is waiting, or in progress. If possible, it helps to change the logging to include a start and end time. This is especially important in logistics, as tight planning and strict schedules are required for a optimal process.

Perspectives

A process mining perspective is the aspect of that process that is being analyzed. The control-flow perspective looks at the order of the events (or steps) of a process. The organizational perspective is about the actors involved. The case perspective looks at the properties of a case. And the time perspective looks at the frequency and timing of events.



Just as in all other process mining domains, control-flow and organizational perspectives are used the most in logistics. The time perspective however is relatively the largest in logistics. Time savings are almost directly linked to cost savings, making them interesting to logistical companies.

Chapter 6

Discussion

This chapter discusses the outcome of this research by comparing the results to existing literature. Threats that can influence the validity of this research are discussed, and suggestions for future research are given.

6.1 Results compared

6.1.1 Opportunities and Challenges

The systematic literature review from this thesis resulted in a set of six opportunities and nine challenges. Comparing this set to the 30 opportunities and 32 challenges by Martin et al. [81] shows some interesting differences. Martin et al. found a lot more practical opportunities and challenges. For example, the challenge *Lack of management support* about the struggle to convince management of the benefits of process mining. Their approach, a Delphi study with experts, might explain this. Martin et al. constructed their list of opportunities and challenges based on brainstorming sessions with experts. These experts were both practitioners and researchers. In this thesis, the list of opportunities and challenges was constructed by findings from papers and journals. This gives the set by Martin et al. a practical and research perspective, and the set in this thesis only a research perspective. The expert interviews from this thesis did bring some practical insights, like the use of legacy systems in finance. However, these insights came after the construction of the opportunities and challenges.

Comparing the sets is possible, though difficult due to the difference in size. The biggest opportunities from Martin et al. are quite similar to those in this thesis. Their largest opportunities are *Evaluating business process performance, Enhancing business process improvement and redesign, and understanding business process compliance.* these are comparable to *Track performance, Optimize the process (cost and time), and Check conformance with regulations* respectively. However, it should be made clear that the opportunities in this thesis are a lot broader.

Comparing the challenges sees more differences. The largest challenges from Martin et al. are *Poor data quality, Unavailability of data, Insufficient technical skills, Insufficient analytical skills, Insufficient process orientation.* The first two can be linked to *Low data quality* from this thesis. The other three look at more practical challenges. It seems that the literature used for the SLR assumed that process mining is being applied by experts, as it did not mention a lack of skills as a challenge. This could also be due to the difference in perspectives, as Martin et al. also included a more practical perspective in their set.

Another contradiction is that the SLR from this thesis concluded that multiple domains struggle to visualize the data in a way that is understandable for people without process mining experience. Martin et al. on the other hand found that over two-thirds of experts see *Generating intuitive visualizations for business users* as an opportunity of process mining. It could be that visualizations of a process are not always good enough, but process mining does enable these visualizations. Having a way to present a process is better than no way at all.

The challenges from this thesis can also be compared to a set of challenges that are listed in the process mining manifesto [129]. Two of these challenges are drivers behind this thesis, being improving the usability and understandability of process mining for people without process mining experience. These are challenges for process mining as a concept. This thesis looked challenges of the implementation of process mining, making some challenges incomparable. What can be compared is the challenge on *Finding, merging, and cleaning data.* This challenge from the manifesto covers both *Handling data from heterogeneous sources, Low data quality, and low granularity of timestamps.* The manifest also has a challenge on *Dealing with complex event logs having diverse characteristics*, which is similar to *Complex processes.* Their other challenge on *Cross-organizational mining* also overlaps with *Handling data from heterogeneous sources.*

6.1.2 Types

A paper by Stefanovic et al. [118] also researched the types of process mining used, specifically within the manufacturing domain. They found that all cases used discovery, 57% of cases used enhancement, and 28% used conformance. This differs from the percentages in this thesis, as they looked at all types used. In this thesis, the focus was on the main type used. When you apply enhancement, you need to discover the model first. Comparing the discovery type therefore becomes difficult, but the ratio between enhancement and conformance can be compared. With 28% and 57%, Stefanovic et al. found that enhancement is used about twice as much as conformance in manufacturing. The results of the SLR are quite similar, with 17% conformance and 31% enhancement, which is also about double. Note that these are percentages, meaning they are influenced by the use of the discovery type. Because of this, the ratio between conformance and enhancement is used.

6.2 Threats to validity

Domain experts

Finding enough experts that are willing to participate in the interview was not a challenge. However, finding experts in education was difficult. Only one expert was found, giving only one person's opinion on the results for the education domain. Whilst more experts would have been desirable, the expert was very experienced within the domain, having spent over 10 years applying educational process mining. The same applies to the ICT expert.

Paper databases

Only one database, Scopus, is used to search for papers in the systematic literature review. This database is chosen as it provides relevant results when searching for process mining papers. A study [40] showed that Scupus covers 73% of relevant papers on process mining. Adding a second database such as Google Scholar would bring this up to 96%. A test was run using Google Scholar and Scopus, but this resulted in too many papers that needed to be analyzed, making it not feasible to use for this project. Using Scopus resulted in 874 papers, and adding Google Scholar resulted in roughly 8000. Given that most relevant papers would still be found with Scopus, Google Scholar was not used for the systematic literature review.

Dependency

This research is heavily dependent on a paper by Garcia et al [26]. It is used for the selection of the process mining application domains. The top six domains that resulted from this research form the basis of this thesis. Given that the author is well respected within the process mining research community, and the paper is cited quite often considering its publication date, it is trusted that this dependency does not hinder the quality of this research.

Research vs Practice

Whilst the top six process mining domains in research come from a reliable source, it does not mean that it is comparable to the top six application domains in practice. An expert had his doubts about the top six during the interview. This was mostly due to the order, which does not hinder the results of this thesis. It was suggested that in practice, governance is a domain that could take the spot of education in the top six. This is the number 7 in the ranking by [26]. Whilst there would have been some value to adding governance, a cutoff point had to be made somewhere, and including all domains would exceed the scope of this thesis.

ICT Domain

The ICT domain came with some difficulties during this research. As all process mining projects need digital systems, it was difficult to distinguish between ICT-specific and general opportunities and challenges. This also explains the lack of domain-specific results for ICT.

6.3 Future work

Domains or process types

Future research could look into the difference in process mining in another way. This thesis made decisions based on the domain in which process mining is applied. The type of process can also be used to look at differences. For example, a financial process can be applied in multiple domains. If it is applied in healthcare, that process likely has more financial characteristics than healthcare characteristics. A financial process within healthcare could have similar opportunities and challenges to the results of the financial domain in this thesis. However, it could be interesting to see if there are differences.

Growing process mining

Another topic that future research can look into is the more practical side of process mining implementation. Process mining projects often start out as a proof of concept that needs to prove its capabilities. Persuading management to use process mining should not be as difficult as it currently is, given the opportunities it can bring. This could be because process mining is relatively young and unknown. Future research could look for ways to make process mining more popular, and make it grow as a method. It can be compared to other methods to see why they are chosen. Finding out how process mining can be improved to be a preferred method would be interesting.

Easier implementations

This thesis aimed to make process mining more understandable for people without process mining experience. Actually implementing process mining is another step. There are already easy-to-use tools that can help you to analyze a data set. But how does a new user get that dataset? Researching easy methods to extract data from systems and preparing that data could be interesting. Coming up with a method that makes this simple for new users can make process mining more accessible to a wide audience.

Chapter 7

Conclusion

This thesis set out to make inexperienced process miners aware of the opportunities and challenges that can arise when applying process mining by listing the relevant opportunities and challenges that are specific to their application domain. It did so by performing a systematic literature review that included papers from the six main process mining domains: Healthcare, ICT, Manufacturing, Education, Finance, and Logistics. The papers were analyzed to find out what type of process mining is being used the most. Across all domains, discovery is used for 52% of cases, conformance 18%, and enhancement 30%. There were some minor differences in these numbers for the individual domains. The same accounts for the process mining perspectives, where in all domains control-flow was used 41%, organizational 25%, case 10%, and time 15%. The systematic literature review further analyzed papers that mentioned opportunities and challenges that occurred during a process mining application. General opportunities have been found that apply to all domains, such as optimizing the process to save cost and time and checking a process to ensure conformance with regulations. General challenges that have been found are on handling data that originated from heterogeneous sources, and low-quality data. Besides these general findings, each domain had specific opportunities and challenges that originate from their domain characteristics.

Bibliography

[1] Karim Abouelmehdi, Abderrahim Beni-Hessane, and Hayat Khaloufi. Big healthcare data: preserving security and privacy. *Journal of big data*, 5(1):1–18, 2018.

[2] Santiago Aguirre, Lina Zuñiga, and Michael Arias. Predictive method proposal for a manufacturing system with industry 4.0 technologies. *Communications in Computer and Information Science*, 1685 CCIS:109 – 121, 2022. Cited by: 0.

[3] Hameed Alqaheri and Mrutyunjaya Panda. An education process mining framework: Unveiling meaningful information for understanding students' learning behavior and improving teaching quality. *Information (Switzerland)*, 13(1), 2022. Cited by: 10; All Open Access, Gold Open Access.

[4] Hanane Ariouat, Awatef Hicheur Cairns, Kamel Barkaoui, Jacky Akoka, and Nasser Khelifa. A two-step clustering approach for improving educational process model discovery. page 38 – 43, 2016. Cited by: 23; All Open Access, Green Open Access.

[5] Poohridate Arpasat. Data-driven analysis of loan approval service of a bank using process mining. volume 2022-November, 2022. Cited by: 0.

[6] Peyman Badakhshan, Bastian Wurm, Thomas Grisold, Jerome Geyer-Klingeberg, Jan Mendling, and Jan vom Brocke. Creating business value with process mining. *Journal of Strategic Information Systems*, 31(4), 2022. Cited by: 2.

[7] Arjel D. Bautista, Syed M. Kumail Akbar, Anthony Alvarez, Tom Metzger, and Marshall Louis Reaves. Process mining in information technology incident management: A case study at volvo belgium. volume 1052, 2013. Cited by: 1.

[8] Till Becker and Wacharawan Intoyoad. Context aware process mining in logistics. volume 63, page 557 – 562, 2017. Cited by: 38; All Open Access, Gold Open Access.

[9] Rob Bemthuis, Niels van Slooten, Jeewanie Jayasinghe Arachchige, Jean Paul Sebastian Piest, and Faiza Allah Bukhsh. A classification of process mining bottleneck analysis techniques for operational support. page 127 – 135, 2021. Cited by: 4; All Open Access, Green Open Access, Hybrid Gold Open Access.

[10] Anis Bey and Ronan Champagnat. Analyzing student programming paths using clustering and process mining. volume 2, page 76 – 84, 2022. Cited by: 1; All Open Access, Hybrid Gold Open Access.

[11] Rehet Bhogal and Anchal Garg. Anomaly detection and fault prediction of breakdown to repair process using mining techniques. page 240 – 245, 2020. Cited by: 2.

[12] Alexander Birk, Yannick Wilhelm, Simon Dreher, Christian Flack, Peter Reimann, and Christoph Gröger. A real-world application of process mining for data-driven analysis of multi-level interlinked manufacturing processes. volume 104, page 417 – 422, 2021. Cited by: 4; All Open Access, Gold Open Access.

[13] Grady Booch, Ivar Jacobson, James Rumbaugh, et al. The unified modeling language. *Unix Review*, 14(13):5, 1996.

[14] Rizal Broer Bahaweres, Jainaba Trawally, Irman Hermadi, and Arif Imam Suroso. Forensic audit using process mining to detect fraud. volume 1779, 2021. Cited by: 4; All Open Access, Bronze Open Access.

[15] Andrea Burattin, Mauro Conti, and Daniele Turato. Toward an anonymous process mining. page 58 – 63, 2015. Cited by: 19.

[16] Juan Antonio Caballero-Hernandez, Juan Manuel Dodero, Ivan Ruiz-Rube, Manuel Palomo-Duarte, Jose Fidel Argudo, and Juan Jose Dominguez-Jimenez. Discovering bottlenecks in a computer science degree through process mining techniques. 2018. Cited by: 4.

[17] Awatef Hicheur Cairns, Joseph Assu Ondo, Billel Gueni, Mehdi Fhima, Marcel Schwarcfeld, Christian Joubert, and Nasser Khelifa. Using semantic lifting for improving educational process models discovery and analysis. volume 1293, page 150 – 161, 2014. Cited by: 16.

[18] Najwa Chaydy and Abdellah Madani. An overview of process mining and its applicability to complex, real-life scenarios. 2019. Cited by: 2.

[19] Minsu Cho, Gyunam Park, Minseok Song, Jinyoun Lee, Byeongeon Lee, and Euiseok Kum. Discovery of resource-oriented transition systems for yield enhancement in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 34(1):17 – 24, 2021. Cited by: 5.

[20] R. Conforti, M. Leoni, de, M. La Rosa, W.M.P. Aalst, van der, and A.H.M. Hofstede, ter. A recommendation system for predicting risks across multiple business process instances. *Decision Support Systems*, 69:1–19, 2015.

[21] Marcelo Dallagassa, Cleiton Garcia, Edson Scalabrin, Sergio Ioshii, and Deborah Carvalho. Opportunities and challenges for applying process mining in healthcare: a systematic mapping study. *Journal of Ambient Intelligence and Humanized Computing*, 13, 02 2021.

[22] Ajitesh Das, Kasandra Dominguez, Aly Elbanna, Quintin Williams, Jeremiah Abiade, and Jida Huang. Utilizing business process re-engineering for optimization of a third-party logistics company. page 217 – 222, 2021. Cited by: 0.

[23] Tonia de Bruin and Michael Rosemann. Using the delphi technique to identify bpm capability areas. page 47, 2007.

[24] Galina Deeva, Johannes De Smedt, Pieter De Koninck, and Jochen De Weerdt. Dropout prediction in moocs: A comparison between process and sequence mining. *Lecture Notes in Business Information Processing*, 308:243 – 255, 2018. Cited by: 17.

[25] R. Divya Sri and Malini M. Patil. Study of learners behaviour in virtual learning environment using process mining. 2021. Cited by: 0.

[26] Cleiton dos Santos Garcia, Alex Meincheim, Elio Ribeiro Faria Junior, Marcelo Rosano Dallagassa, Denise Maria Vecino Sato, Deborah Ribeiro Carvalho, Eduardo Alves Portela Santos, and Edson Emilio Scalabrin. Process mining techniques and applications – a systematic mapping study. *Expert Systems with Applications*, 133:260–295, 2019.

[27] Simon Dreher, Peter Reimann, and Christoph Gröger. Application fields and research gaps of process mining in manufacturing companies. volume P-307, page 621 – 634, 2020. Cited by: 4.

[28] Marlon Dumas, Marcello La Rosa, Jan Mendling, Hajo A Reijers, et al. *Fundamentals of business process management*, volume 1. Springer, 2013.

[29] Le Toan Duong, Louise Travé-Massuyès, Audine Subias, and Nathalie Barbosa Roa. Assessing product quality from the production process logs. *International Journal of Advanced Manufacturing Technology*, 117(5-6):1615 – 1631, 2021. Cited by: 3; All Open Access, Green Open Access.

[30] Maikel Eck, Natalia Sidorova, and Wil Aalst. Guided interaction exploration in artifact-centric process models. pages 109–118, 07 2017.

[31] Darko Etinger. Discovering and mapping lms course usage patterns to learning outcomes. *Advances in Intelligent Systems and Computing*, 1131 AISC:486 – 491, 2020. Cited by: 4.

[32] Darko Etinger, Tihomir Orehovački, and Snježana Babić. Applying process mining techniques to learning management systems for educational process model discovery and analysis. *Advances in Intelligent Systems and Computing*, 722:420 – 425, 2018. Cited by: 4.

[33] Daniela Fishbein, Siddhartha Nambiar, Kendall McKenzie, Maria Mayorga, Kristen Miller, Kevin Tran, Laura Schubel, Joseph Agor, Tracy Kim, and Muge Capan. Objective measures of workload in healthcare: a narrative review. *International Journal of Health Care Quality Assurance*, 33(1):1–17, 2020.

[34] Ilham Akbar Fitriansah, Rachmadita Andreswari, and Muhammad Azani Hasibuan. Business process analysis of academic information system application using process mining (case study: Final project module). page 189 – 194, 2019. Cited by: 3.

[35] Jonas Friederich and Sanja Lazarova-Molnar. Process mining for reliability modeling of manufacturing systems with limited data availability. 2021. Cited by: 2.

[36] Fiona Fylan. Semi-structured interviewing. *A handbook of research methods for clinical and health psychology*, 5(2):65–78, 2005.

[37] Nick Gehrke and Niels Mueller-Wickop. Basic principles of financial process mining: A journey through financial data in accounting information systems. volume 3, page 1590 – 1600, 2010. Cited by: 17.

[38] Kerstin Gerke, Alexander Claus, and Jan Mendling. Process mining of rfid-based supply chains. pages 285–292, 07 2009.

[39] Kerstin Gerke, Konstantin Petruch, and Gerrit Tamm. Optimization of service delivery through continual process improvement: A case study. page 94 – 107, 2010. Cited by: 2.

[40] Mahdi Ghasemi and Daniel Amyot. Process mining in healthcare: A systematised literature review. *International Journal of Electronic Healthcare*, 9:60, 01 2016.

[41] Kanika Goel, Sander J.J. Leemans, Moe T. Wynn, Arthur H.M. ter Hofstede, and Janne Barnes. Improving phd student journeys with process mining: Insights from a higher education institution. volume 3112, page 39 – 49, 2021. Cited by: 0.

[42] K. Grigorova, E. Malysheva, and S. Bobrovskiy. Application of data mining and process mining approaches for improving e-learning processes. volume 1903, page 115 – 121, 2017. Cited by: 6; All Open Access, Bronze Open Access.

[43] Teresa Guarda, Manuel Filipe Santos, Maria Fernanda Augusto, Carlos Silva, and Filipe Pinto. Process mining: A framework proposal for pervasive business intelligence. 2013. Cited by: 3.

[44] Monika Gupta, Alexander Serebrenik, and Pankaj Jalote. Improving software maintenance using process mining and predictive analytics. In *2017 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 681–686, 2017.

[45] Monika Gupta, Ashish Sureka, and Srinivas Padmanabhuni. Process mining multiple repositories for software defect resolution from control and organizational perspective. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, MSR 2014, page 122–131, New York, NY, USA, 2014. Association for Computing Machinery.

[46] Sergio Hernández, Joaquín Ezpeleta, S.J. Van Zelst, and Wil M. P. Van Der Aalst. Assessing process discovery scalability in data intensive environments. page 99 – 104, 2016. Cited by: 10.

[47] G.T.s Ho and H. Lau. Development of an olap-fuzzy based process mining system for quality improvement. volume 228, pages 243–258, 11 2007.

[48] Wacharawan Intayoad and Till Becker. Applying process mining in manufacturing and logistic for large transaction data. In *Dynamics in Logistics: Proceedings of the 6th International Conference LDIC 2018, Bremen, Germany*, pages 378–388. Springer, 2018.

[49] Wacharawan Intayoad, Till Becker, and Otthein Herzog. Process discovery method in dynamic manufacturing and logistics environments. *Communications in Computer and Information Science*, 1278:143 – 163, 2020. Cited by: 1.

[50] Amin Jalali. Aspect mining in business process management. *Lecture Notes in Business Information Processing*, 194:246 – 260, 2014. Cited by: 5.

[51] Renée S. Jansen, Anouschka van Leeuwen, Jeroen Janssen, and Liesbeth Kester. Exploring the link between self-regulated learning and learner behaviour in a massive open online course. *Journal of Computer Assisted Learning*, 38(4):993 – 1004, 2022. Cited by: 9; All Open Access, Green Open Access.

[52] Silvia Jaqueline Urrea-Contreras, Brenda L. Flores-Rios, María Angélica Astorga-Vargas, and Jorge E. Ibarra-Esquer. Process mining perspectives in software engineering: A systematic literature review. In *2021 Mexican International Conference on Computer Science (ENC)*, pages 1–8, 2021.

[53] Lihong Jiang, Jianyi Wang, Nazaraf Shah, Hongming Cai, Chengxi Huang, and Ray Farmer. A process-mining-based scenarios generation method for soa application development. *Service Oriented Computing and Applications*, 10(3):303 – 315, 2016. Cited by: 3; All Open Access, Green Open Access.

[54] M.V. Kamal and D. Vasumathi. Abpmdf: Towards a framework for automated model discovery from process event logs for business intelligence. *International Journal of Advanced Trends in Computer Science and Engineering*, 8(3):913 – 921, 2019. Cited by: 0; All Open Access, Bronze Open Access.

[55] T.S. Karthik and B. Kamala. Cloud based ai approach for predictive maintenance and failure prevention. volume 2054, 2021. Cited by: 2; All Open Access, Gold Open Access.

[56] Sagit Kedem-Yemini, Naor S. Mamon, and Gal Mashiah. An analysis of cargo release services with process mining: A case study in a logistics company. volume 2018, page 726 – 736, 2018. Cited by: 3.

[57] Mathias Kirchmer, Francisco Gutiérrez, and Sigifredo Laengle. Process mining for organizational agility. *Industrial Management (Norcross, Georgia)*, 52(1):19 – 24, 2010. Cited by: 0.

[58] Dino Knoll, Gunther Reinhart, and Marco Prüglmeier. Enabling value stream mapping for internal logistics using multidimensional process mining. *Expert Systems with Applications*, 124:130 – 142, 2019. Cited by: 64.

[59] Dino Knoll, Julian Waldmann, and Gunther Reinhart. Developing an internal logistics ontology for process mining. volume 79, page 427 – 432, 2019. Cited by: 14; All Open Access, Gold Open Access, Green Open Access.

[60] Wolfgang Koehler and Yanguo Jing. Trace induction for complete manufacturing process model discovery. *International Journal of Advanced Manufacturing Technology*, 110(1-2):29 – 43, 2020. Cited by: 0; All Open Access, Green Open Access.

[61] Mahesh Kumbhar, Amos H.C. Ng, and Sunith Bandaru. Bottleneck detection through data integration, process mining and factory physics-based analytics. volume 21, page 737 – 748, 2022. Cited by: 1; All Open Access, Gold Open Access, Green Open Access.

[62] Valeriy Kurganov, Aleksey Dorofeev, Mikhail Gryaznov, and Mikhail Yakimov. Process mining as a means of improving the reliability of road freight transportations. volume 54, page 300 – 308, 2021. Cited by: 2; All Open Access, Gold Open Access.

[63] Angelina Prima Kurniati, Owen Johnson, David Hogg, and Geoff Hall. Process mining in oncology: A literature review. In *2016 6th international conference on information communication and management (ICICM)*, pages 291–297. IEEE, 2016.

[64] Geetika T. Lakshmanan and Rania Khalaf. Leveraging process-mining techniques. *IT Professional*, 15(5):22 – 30, 2013. Cited by: 9.

[65] Geetika T Lakshmanan, Davood Shamsi, Yurdaer N Doganata, Merve Unuvar, and Rania Khalaf. A markov prediction model for data-driven semi-structured business processes. *Knowledge and Information Systems*, 42:97–126, 2015.

[66] Lijun Lan, Ying Liu, and Wen Feng Lu. Learning from the past: Uncovering design process models using an enriched process mining. *Journal of Mechanical Design*, 140(4), 2018. Cited by: 7; All Open Access, Bronze Open Access, Green Open Access.

[67] H.C.W. Lau, G.T.S. Ho, Y. Zhao, and N.S.H. Chung. Development of a process mining system for supporting knowledge discovery in a supply chain network. *International Journal of Production Economics*, 122(1):176 – 187, 2009. Cited by: 41.

[68] C.K.H. Lee, K.L. Choy, G.T.S. Ho, and C.H.Y. Lam. A slippery genetic algorithm-based process mining system for achieving better quality assurance in the garment industry. *Expert Systems with Applications*, 46:236 – 248, 2016. Cited by: 40.

[69] C.K.H. Lee, G.T.S. Ho, K.L. Choy, and G.K.H. Pang. A rfid-based recursive process mining system for quality assurance in the garment industry. *International Journal of Production Research*, 52(14):4216 – 4238, 2014. Cited by: 42.

[70] Seung-Kyung Lee, Bongseok Kim, Minhoe Huh, Sungzoon Cho, Sungkyu Park, and Daehyung Lee. Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry. *Expert Systems with Applications*, 40(1):83 – 95, 2013. Cited by: 34.

[71] Artini M. Lemos, Caio C. Sabino, Ricardo M. F. Lima, and Cesar A. L. Oliveira. Using process mining in software development process management: A case study. In *2011 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1181–1186, 2011.

[72] Federico A. Lievano-Martínez, Javier D. Fernández-Ledesma, Daniel Burgos, John W. Branch-Bedoya, and Jovani A. Jimenez-Builes. Intelligent process automation: An application in manufacturing industry. *Sustainability (Switzerland)*, 14(14), 2022. Cited by: 2; All Open Access, Gold Open Access, Green Open Access.

[73] Audrius Lopata, Rimantas Butleris, Saulius Gudas, Kristina Rudžionienė, Liutauras Žioba, Ilona Veitaitė, Darius Dilijonas, Evaldas Grišius, and Maarten Zwitserloot. Financial process mining characteristics. *Communications in Computer and Information Science*, 1665 CCIS:209 – 220, 2022. Cited by: 1.

[74] Audrius Lopata, Rimantas Butleris, Saulius Gudas, Vytautas Rudžionis, Kristina Rudžionienė, Liutauras Žioba, Ilona Veitaitė, Darius Dilijonas, Evaldas Grišius, and Maarten Zwitserloot. Financial data preprocessing issues. *Communications in Computer and Information Science*, 1486 CCIS:60 – 71, 2021. Cited by: 2.

[75] Rafael Lorenz, Julian Senoner, Wilfried Sihn, and Torbjørn Netland. Using process mining to improve productivity in make-to-stock manufacturing. *International Journal of Production Research*, 59(16):4869 – 4880, 2021. Cited by: 19; All Open Access, Green Open Access, Hybrid Gold Open Access.

[76] Martin Macak, Daniela Kruzelova, Stanislav Chren, and Barbora Buhnova. Using process mining for git log analysis of projects in a software development course. *Education and Information Technologies*, 26(5):5939 – 5969, 2021. Cited by: 10.

[77] Dennis G.J.C. Maneschijn, Rob H. Bemthuis, Faiza A. Bukhsh, and Maria-Eugenia Iacob. A methodology for aligning process model abstraction levels and stakeholder needs. volume 1, page 137 – 147, 2022. Cited by: 1; All Open Access, Green Open Access, Hybrid Gold Open Access.

[78] R. S. Mans, M. H. Schonenberg, M. Song, W. M. P. van der Aalst, and P. J. M. Bakker. Application of process mining in healthcare – a case study in a dutch hospital. In Ana Fred, Joaquim Filipe, and Hugo Gamboa, editors, *Biomedical Engineering Systems and Technologies*, pages 425–438, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[79] R.S. Mans, W.M.P. Aalst, van der, and R.J.B. Vanwersch. *Process mining in healthcare : opportunities beyond the ordinary.* BPM reports. BPMcenter. org, 2013.

[80] R.S. Mans, W.M.P. Aalst, van der, and R.J.B. Vanwersch. *Process mining in healthcare: evaluating and exploiting operational healthcare processes.* SpringerBriefs in Business Process Management,. Springer, Germany, 2015.

[81] Niels Martin, Dominik A Fischer, Georgi D Kerpedzhiev, Kanika Goel, Sander JJ Leemans, Maximilian Röglinger, Wil MP van der Aalst, Marlon Dumas, Marcello La Rosa, and Moe T Wynn. Opportunities and challenges for process mining in organizations: results of a delphi study. *Business & Information Systems Engineering*, 63:511–527, 2021.

[82] Pablo Martinez, Oscar Montañes, Juan Manuel Serralta, and Libertad Tansini. Modelling computer engineering student trajectories with process mining. volume 3059, page 48 – 57, 2021. Cited by: 0.

[83] Mohammed Mesabbah and Susan McKeever. Presenting a hybrid processing mining framework for automated simulation model generation. volume 2018-December, page 1370 – 1381, 2019. Cited by: 7; All Open Access, Green Open Access.

[84] Judith Michael, Agnes Koschmider, Felix Mannhardt, Nathalie Baracaldo, and Bernhard Rumpe. User-centered and privacy-driven process mining system design for iot. *Lecture Notes in Business Information Processing*, 350:194 – 206, 2019. Cited by: 20.

[85] Jorge Munoz-Gama and Isao Echizen. Insuring sensitive processes through process mining. In *2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing*, pages 447–454, 2012.

[86] Gašper Mušič and Primož Rojec. Process mining of production management data for improvement of production planning and manufacturing execution. page 483 – 488, 2012. Cited by: 2.

[87] Zsuzsanna Nagy, Agnes Werner-Stark, and Tibor Dulai. Using process mining in real-time to reduce the number of faulty products. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11695 LNCS:89 – 104, 2019. Cited by: 3.

[88] Anake Nammakhunt, Parham Porouhan, and Wichian Premchaiswadi. Creating and collecting e-learning event logs to analyze learning behavior of students through process mining. *International Journal of Information and Education Technology*, 13(2):211 – 222, 2023. Cited by: 1; All Open Access, Gold Open Access.

[89] Konstantin Nikitin. Educational game analysis using intention and process mining. volume 2795, page 117 – 125, 2020. Cited by: 0.

[90] Kingsley Okoye and Samira Hosseini. Educational process intelligence: A process mining approach and model analysis. *Advances in Intelligent Systems and Computing*, 1180 AISC:201 – 212, 2021. Cited by: 0.

[91] Zacharoula Papamitsiou and Anastasios A. Economides. Process mining of interactions during computer-based testing for detecting and modelling guessing behavior. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9753:437 – 449, 2016. Cited by: 10.

[92] Minjeong Park, Minseok Song, Tae Hyun Baek, SookYoung Son, Seung Jin Ha, and Sung Woo Cho. Workload and delay analysis in manufacturing process using process mining. *Lecture Notes in Business Information Processing*, 219:138 – 151, 2015. Cited by: 26.

[93] Mykola Pechenizkiy Jr., Nikola Trčka, Ekaterina Vasilyeva, Wil Van Der Aalst, and Paul De Bra. Process mining online assessment data. page 279 – 288, 2009. Cited by: 62.

[94] Edward M.L. Peters, Guido Dedene, and Jonas Poelmans. Understanding service quality and customer churn by process discovery for a multi-national banking contact center. page 228 – 233, 2013. Cited by: 5.

[95] Jean Paul Sebastian Piest, Jennifer Alice Cutinha, Rob Henk Bemthuis, and Faiza Allah Bukhsh. Evaluating the use of the open trip model for process mining: An informal conceptual mapping study in logistics. volume 1, page 290 – 296, 2021. Cited by: 5.

[96] A.A. Popov, S.N. Masaev, D.A. Edimichev, and O.V. Pomolotova. Analytical treatment of transport logistics business processes by the process mining technology. volume 1679, 2020. Cited by: 0; All Open Access, Bronze Open Access.

[97] Frans Prathama, Bernardo Nugroho Yahya, Danny Darmawan Harjono, and E.R. Mahendrawathi. Trace clustering exploration for detecting sudden drift: A case study in logistic process. volume 161, page 1122 – 1130, 2019. Cited by: 8; All Open Access, Gold Open Access.

[98] Andrea Rocco Racca, Emilio Sulis, and Sara Capecchi. Behavioral web tracking in e-learning: An educational process mining application. volume 2022-July, page 269 – 274, 2022. Cited by: 0.

[99] Ria Rahmawati, Rachmadita Andreswari, and Rokhman Fauzi. Analysis and exploratory of lecture preparation process to improve the conformance using process mining. page 461 – 466, 2022. Cited by: 2.

[100] Belén Ramos-Gutiérrez, Ángel Jesús Varela-Vaca, F. Javier Ortega, María Teresa Gómez-López, and Moe Thandar Wynn. A nlp-oriented methodology to enhance event log quality. *Lecture Notes in Business Information Processing*, 421:19 – 35, 2021. Cited by: 3; All Open Access, Green Open Access.

[101] Stefanie Rinderle-Ma and Juergen Mangler. Process automation and process mining in manufacturing. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12875 LNCS:3 – 14, 2021. Cited by: 9; All Open Access, Green Open Access.

[102] Eric Rojas, Jorge Munoz-Gama, Marcos Sepúlveda, and Daniel Capurro. Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*, 61:224–236, 2016.

[103] Grzegorz Rozenberg and Joost Engelfriet. *Elementary net systems*, pages 12–121. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[104] Anne Rozinat, I.S.M. Jong, C.W. Gunther, and Wil Aalst. Process mining applied to the test process of wafer scanners in asml. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 39:474 – 479, 08 2009.

[105] Vladimir Rubin, Christian W. Günther, Wil M. P. van der Aalst, Ekkart Kindler, Boudewijn F. van Dongen, and Wilhelm Schäfer. Process mining framework for software processes. In Qing Wang, Dietmar Pfahl, and David M. Raffo, editors, *Software Process Dynamics and Agility*, pages 169–181, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[106] Julia Rudnitckaia, Wacharawan Intayoad, Till Becker, and Tomas Hruska. Applying process mining to the ship handling process at oil terminal. page 552 – 557, 2019. Cited by: 1.

[107] Edson Ruschel, Eduardo de Freitas Rocha Loures, and Eduardo Alves Portela Santos. Performance analysis and time prediction in manufacturing systems. *Computers and Industrial Engineering*, 151, 2021. Cited by: 8.

[108] Edson Ruschel, Eduardo Alves Portela Santos, and Eduardo de Freitas Rocha Loures. Mining shop-floor data for preventive maintenance management: Integrating probabilistic and predictive models. *Procedia Manufacturing*, 11:1127–1134, 2017. 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy.

[109] Edson Ruschel, Eduardo Alves Portela Santos, and Eduardo de Freitas Rocha Loures. Establishment of maintenance inspection intervals: an application of process mining techniques in manufacturing. *Journal of Intelligent Manufacturing*, 31(1):53 – 72, 2020. Cited by: 25.

[110] Hind R'bigui and Chiwoon Cho. Customer oder fulfillment process analysis with process mining: An industrial application in a heavy manufacturing company. page 247 – 252, 2017. Cited by: 12.

[111] Juan Pablo Salazar-Fernandez, Marcos Sepúlveda, and Jorge Munoz-Gama. Describing educational trajectories of engineering students in individual high-failure rate courses that lead to late dropout. volume 2425, page 39 – 48, 2019. Cited by: 0.

[112] G. Schuh, A. Gutzlaff, S. Cremer, S. Schmitz, and A. Ayati. A data model to apply process mining in end-to-end order processing processes of manufacturing companies. volume 2020-December, page 151 – 155, 2020. Cited by: 6.

[113] G. Schuh, A. Gützlaff, S. Schmitz, C. Kuhn, and N. Klapper. A methodology to apply process mining in end-to-end order processing of manufacturing companies. *Lecture Notes in Mechanical Engineering*, page 127 – 137, 2022. Cited by: 0.

[114] Maria Laura Sebu and Horia Ciocarlie. Applied process mining in software development. pages 55–60, 05 2014.

[115] Gayane Sedrakyan, Jochen De Weerdt, and Monique Snoeck. Process-mining enabled feedback: "tell me what i did wrong" vs. "tell me how to do it right". *Computers in Human Behavior*, 57:352 – 376, 2016. Cited by: 48; All Open Access, Green Open Access.

[116] Michael Siek and Ryan Malik Gunadharma Mukti. Process mining with applications to automotive industry. volume 924, 2020. Cited by: 4; All Open Access, Bronze Open Access.

[117] Kamonmas Sirisong, Prajin Palangsantikul, Poohridate Arpasat, Sarayut Intarasema, and Sompong Tumswadi. Analysis of a bank's lending approval system using process mining. 2021. Cited by: 2.

[118] Darko Stefanovic, Dusanka Dakic, Branislav Stevanov, and Teodora Lolic. Process mining in manufacturing: Goals, techniques and applications. *IFIP Advances in Information and Communication Technology*, 591 IFIP:54 – 62, 2020. Cited by: 3; All Open Access, Green Open Access.

[119] Florian Stertz, Juergen Mangler, Beate Scheibel, and Stefanie Rinderle-Ma. Expectations vs. experiences – process mining in small and medium sized manufacturing companies. *Lecture Notes in Business Information Processing*, 427 LNBIP:195 – 211, 2021. Cited by: 7; All Open Access, Green Open Access.

[120] Pisit Sukanjanachot, Wipavan Narksarp, Norranut Saguansakdiyotin, and Wichian Premchaiswadi. Procedure analysis of courses offered by universities using process mining. volume 2022-November, 2022. Cited by: 0.

[121] Suriadi Suriadi, Moe T. Wynn, Chun Ouyang, Arthur H. M. Ter Hofstede, and Nienke J. Van Dijk. Understanding process behaviours in a large insurance company in australia: A case study. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7908 LNCS:449 – 464, 2013. Cited by: 56; All Open Access, Bronze Open Access, Green Open Access.

[122] N. Trcka, M. Pechenizkiy, and W.M.P. Aalst, van der. *Process mining from educational data*, pages 123–142. Chapman amp; Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, 2011.

[123] Thi Bich Hong Tu and Minseok Song. Analysis and prediction cost of manufacturing process based on process mining. 2016. Cited by: 26.

[124] Rahila Umer, Teo Susnjak, Anuradha Mathrani, and Suriadi Suriadi. Data quality challenges in educational process mining: building process-oriented event logs from process-unaware online learning systems. *International Journal of Business Information Systems*, 39(4):569 – 592, 2022. Cited by: 1; All Open Access, Green Open Access.

[125] Mehrnoosh Vahdat, Luca Oneto, Davide Anguita, Mathias Funk, and Matthias Rauterberg. A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9307:352 – 366, 2015. Cited by: 45; All Open Access, Green Open Access.

[126] R. M. E. Ruud Van Cruchten and H. Hans Weigand. Process mining in logistics: The need for rule-based data abstraction. volume 2018-May, page 1 – 9, 2018. Cited by: 10.

[127] Wil van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, Berlin, 2011.

[128] Wil Van Der Aalst. Big software on the run: In vivo software analytics based on process mining (keynote). volume 24-26-August-2015, page 1 – 5, 2015. Cited by: 22.

[129] Wil van der Aalst et al. Process mining manifesto. In *Business Process Management Workshops*, page 169–194. Springer, 2011.

[130] W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, and H.M.W. Verbeek. Business process mining: An industrial application. *Information Systems*, 32(5):713–732, 2007.

[131] B. F. van Dongen, A. K. Alves de Medeiros, and L. Wen. *Process Mining: Overview and Outlook of Petri Net Discovery Algorithms*, pages 225–242. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[132] Yuri Van Midden, Jeewanie Jayasinghe Arachchige, and Faiza Allah Bukhsh. Optimal scenario mining for business strategy decision-making through process mining. page 311 – 316, 2021. Cited by: 1.

[133] Jianyi Wang, Lihong Jiang, and Hongming Cai. Scenario-based method for business process analysis and improvement in soa. page 19 – 25, 2014. Cited by: 5.

[134] Jie Wang, Bing Zhu, Ying Wang, and Lei Huang. Mining organizational behaviors in collaborative logistics chain: An empirical study in a port. 2016. Cited by: 3.

[135] Ying Wang, Filip Caron, Jan Vanthienen, Lei Huang, and Yi Guo. Acquiring logistics process intelligence: Methodology and an application for a chinese bulk port. *Expert Systems with Applications*, 41(1):195 – 209, 2014. Cited by: 50.

[136] Sven Weinzierl, Verena Wolf, Tobias Pauli, Daniel Beverungen, and Martin Matzner. Detecting temporal workarounds in business processes–a deep-learning-based method for analysing event log data. *Journal of Business Analytics*, 5(1):76 – 100, 2022. Cited by: 7; All Open Access, Hybrid Gold Open Access.

[137] Michael Werner. Process model representation layers for financial audits. volume 2016-March, page 5338 – 5347, 2016. Cited by: 9.

[138] Michael Werner. Materiality maps - process mining data visualization for financial audits. volume 2019-January, page 1045 – 1054, 2019. Cited by: 3.

[139] Michael Werner and Nick Gehrke. Multilevel process mining for financial audits. *IEEE Transactions on Services Computing*, 8(6):820–832, 2015.

[140] Michael Werner and Nick Gehrke. Multilevel process mining for financial audits. *IEEE Transactions on Services Computing*, 8(6):820 – 832, 2015. Cited by: 27.

[141] Michael Werner, Nick Gehrke, and Markus Nüttgens. Business process mining and reconstruction for financial audits. page 5350 – 5359, 2012. Cited by: 20.

[142] Fitri Almira Yasmin, Faiza Allah Bukhsh, and Patricio De Alencar Silva. Process enhancement in process mining: A literature review. *CEUR workshop proceedings*, 2270:65–72, December 2018. 8th International Symposium on Data-driven Process Discovery and Analysis 2018, SIMPDA 2018 ; Conference date: 13-12-2018 Through 14-12-2018.

[143] Ibrahim E. Yazici and Orhan Engin. Use of process mining in bank real estate transactions and visualization with fuzzy models. *Advances in Intelligent Systems and Computing*, 1029:265 – 272, 2020. Cited by: 2.

[144] Pierluigi Zerbino, Davide Aloini, Riccardo Dulmin, and Valeria Mininno. Towards analytics-enabled efficiency improvements in maritime transportation: A case study in a mediterranean port. *Sustainability (Switzerland)*, 11(16), 2019. Cited by: 15; All Open Access, Gold Open Access, Green Open Access.

[145] Aimilia Zisimou, Ioanna Kalaitzoglou, Georgia Theodoropoulou, Alexandros Bousdekis, and Georgios Miaoulis. Evaluation of public funding processes by mining event logs. 2021. Cited by: 0.

[146] Álvaro Rebuge and Diogo R. Ferreira. Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, 37(2):99–116, 2012. Management and Engineering of Process-Aware Information Systems.

[147] Güzin Özdağoğlu and Ece Kavuncubaşı. Monitoring the software bug-fixing process through the process mining approach. *Journal of Software: Evolution and Process*, 31(7), 2019. Cited by: 6.

Appendices

Appendix A

Interview Protocol

The interviews with the process mining experts followed the following protocol.

A.1 Opening

- Small welcome word
- Ask the interviewee to fill in the consent form and ask permission to record
- Start recording
- Explain the setup of the interview (It starts with general questions, and then the SLR findings are presented and discussed)
- Explain the goal of the interview and the research and explain what opportunities and challenges are

A.2 General questions

- A Few basic questions on their process mining history:
  - When did you start using process mining?
  - How often do you use it?
  - What was your role in the application?
  - In what domains did you apply it?
- If they have experience in multiple domains, ask what was different between the domains (Ask about opportunities and challenges if they do not bring it up themselves)
- For each domain they have experience in, ask for what goal they apply process mining (opportunities), and what hindered them during the application (challenges). Ask if they think these opportunities and challenges are specific to their domain and why.

A.3 SLR findings

For each domain that the expert has experience in, ask the following questions. If an expert already talked about an opportunity or challenge in the previous section, it can be skipped. Except if the SLR came to another conclusion, then discuss this with the expert.

- Discuss all opportunities of a domain and ask if the expert agrees or disagrees and why
- Discuss all challenges of a domain and ask if the expert agrees or disagrees and why
- Ask if any opportunities or challenges came to mind that have not been mentioned in the previous section, nor in the SLR findings

A.4 Comparing domains

For the experts who have experience with multiple domains, ask the following questions. This is similar to the *General questions* section, but they might remember new insights after the SLR results.

- Are there differences between the opportunities in domains X and Y (and Z...)?
- Are there differences between the challenges in domains X and Y (and Z...)?
- Are there differences in the nature of the data/models between the domains?
- Are there other differences between the domains that could be interesting?

A.5 Final questions

- If time allows it, ask what advice they would give to an inexperienced process miner who just started applying process mining.
- Ask if they would like to add or say anything else, some concluding thoughts
- Thank them for their time and collaboration
- End recording