UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics & Computer Science

Enhancement in Process Mining

Guideline for Process Owner and Process Analyst

Fitri Almira Yasmin
Master Thesis
February 2019

Supervisors:

dr. Faiza A. Bukhsh
Faculty of Electrical Engineering,
Mathematics and Computer Science

dr. Alphonsus B.J.M. Wijnhoven
Faculty of Behavioural, Management
and Social sciences

University of Twente
Enschede, The Netherlands

Abstract

Author
Fitri Almira Yasmin

Title of the Paper
Enhancement in Process Mining: Guideline for Process Owner and Process Analyst.

Keywords
Data Science, Process Mining, Process Model, Process Discovery, Social Network Analysis, Performance Analysis, Conformance Check.

Description
Business process in an organization consists of numerous activities performed by different actors. A process model is a representation of process executions. In practices, a process model is typically created through meetings and interviews with various stakeholders in the organization. This traditional approach usually takes up to several years to complete. On the other hand, process mining offers an automatic means to develop a process model. The process model discovered by process mining is based on actual process behavior recorded in the event log. However, process mining is a relatively young field, and there is a lack of attention about how to perform a process mining project. In this thesis, we proposed a three-phase process analysis approach using process mining techniques involving process owner and process analyst. The application of the proposed approach is demonstrated using real-life data sets. The approach elaborations and result of the demonstration is combined into a "guideline" document in the form of a white-paper. For evaluation purpose, the guideline is presented to potential process mining users.

Acknowledgment

The submission of this thesis marks the end of my study at the University of Twente. A roller coaster-like journey filled with excitements, doubts, ambitions, thrills, and fears all at once. Two years away from tropical weather, crowded city, and spicy foods have taught me about myself more than I could've imagined.

First, I would like to express my special gratitude to my thesis supervisors who are like Yin and Yang. Faiza with her technical background and optimistic dispositions. Fons with his business expertise and critical questions.

My master study would have not been easy without my family. Thank you, Mom and Dad, for teaching me about tenacity, faith, and bravery. I'm deeply grateful for having Fahmi as my best critique, lifetime supporter, arch-enemy and brother. Thank you for carrying my load of responsibility back at home and taking care of our younger siblings.

I would like to show my gratitude to all my friends who provide me with the support I needed during my master study. Thanks to Riris, Feby, Ika, Fania, Eva, Dzul, Marilena, Agathi, Fajri, Raisa, Yaumi, and Margaretha, my life in Enschede are full of stories. I also would like to thank Khalida, Rizka, Tia, Parput, Johan, Maya, Yemi, Aisy and Nurul for answering my hours-long phone call and replying to my trivial text messages despite the time difference between the countries we live in.

Lastly, I would like to thank questionnaire respondents, process mining community members, experiment participants and all of the contributors that have made the completion of this master thesis possible.

Enschede, February 2019

Fitri Almira Yasmin

Document Change Control

Table 0. Document Versions Description

Revision Index	Date of Issue	Description of Change
Version 1.0	01-11-2018	The initial version of the document containing the first three chapter. Chapter 1 introduces the process mining field. Chapter 2 sourced from a research topic document titled "Process Enhancement in Process Mining Field: A Literature Review" by the same author of this research. The remaining chapters are yet to be written. Chapter 3 describes the research methodology of this study.
Version 2.0	10-01-2019	The addition of cover and table of content. The merging of chapter 1 and chapter 3 from the previous version. Thus, Chapter 1 in this version includes both introduction and research methodology. Chapter 2 is the same as the previous version but minor change and addition. Chapter 3 contains the objective of the artifact is written, as well as the draft version of the artifact's design in Chapter 4. Chapter 5, 6 and 7 is yet to be written.
Version 3.0	04-02-2019	The finalization of Chapter 4. The addition of Chapter 5. Chapter 6 and 7 is yet to be written.
Version 4.0	21-02-2019	The addition of intro and summary in each chapter. Chapter 2 and Chapter 3 in the previous version is merged. Therefore, in this version the document structured as follows: Chapter 1 introduces some basic concepts of process mining and research plan. Chapter 2 presents the literature review. Chapter 3 contains the design and development of the artifact. Chapter 4 demonstrate the implementation of the artifact into a real-life data sets. Chapter 5 evaluates the artifact using experimental evaluation. Based on the experiments then the ProM tools user acceptance is evaluated. Lastly, Chapter 7 concludes the findings, limitation, and recommendation for future works.
Version 5.0	26-02-2019	The final version of the document. Grammar and plagiarism checked. The content is the same as the previous version.

Table of Contents

List of Figures

List of Tables

List of Definitions

1. Introduction

This chapter starts with the introduction of data science and the positioning of process mining within the field. It then continues with the introduction of process mining. After that, it elaborates the motivation of the study, lists the research questions, and the research plan and approaches. The end of the chapter summarizes the research design and the outlines of the thesis document.

1.1. Data Science

In the late 1980s, the presumably most talked story of data analytics came to light. The story goes like this: somewhere in the Midwest, a retail chain analyzed their checkout-counter data and found a correlation between beers and diapers (Rao, 1998). The store then puts the two product's aisles closer to one another, and it leads to a significant increase in sales.

The discovery of the association between different items like the "Beer-diaper Syndrome" exhibits one of many data analytics capabilities. Especially with the advancement of new technologies and techniques for inexpensively working with a very large data sets, the interest in data analytics has grown significantly (Harris et al., 2013). In recent years, the interest incites the use of popular buzzwords like "data science," "big data," "data analytics," etc.

The term "data analytics" and "data science" are, by some people, used interchangeably. The opinions split into two, people who argue that there is no difference between data analytics and data science, and people who believe the opposite. Despite the contrasting views, the word "interdisciplinary" is widely agreed upon, and it represents the main characteristics of the topic. Aalst (2016) describes **data science** as an interdisciplinary field aiming to turn data into real value.

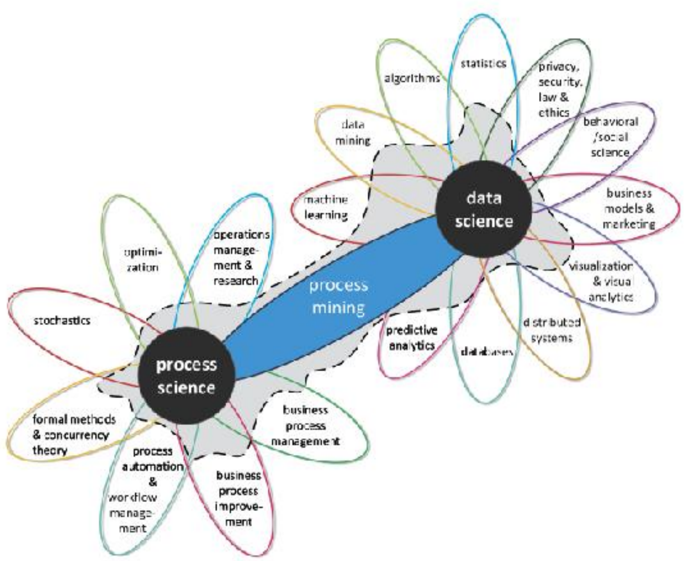


Figure 1.1 The Position of Process Mining
(Source: Aalst, 2016)

The inclusion of the disciplines that are considered as the part of data science may vary from one literature to another. For example, Harris et al. (2013) took a survey and use clustering to define the subfields of data science. They include business, machine learning, big data, math, operation research, programming, and statistics. On the other hand, Aalst (2016) lists statistics, algorithm, data mining, machine learning, predictive analysis, databases, distributed systems, visualization and visual analytics, business model and marketing, behavioral and social science, process mining, and privacy, security, law and ethics as the overlapping subdisciplines of data science.

Our study is about process mining, one of the subfields of data science. Besides, process mining is also a part of process science. Figure 1.1 illustrates the positioning of process mining between the data science and process science fields. **Process science** is an umbrella term for a broad discipline that combines knowledge from information technology and knowledge from management sciences to improve and run operational processes (Aalst, 2016). Operation research, workflow management, business process management, and improvement are the examples of the process science's subfields.

1.2. Process Mining

As the information technology proliferates, more activities are executed digitally. At the same time, more data are constantly collected by information systems. From the moment people wake up until they are going back to sleep, a lot of behavior is recorded. For instance, a mobile health tracker application can automatically detect the sleep duration of the phone users based on the last time they checked their phone at night, until the first moment they check their phone on the next day. It can also automatically record the phone user's physical activity such as walking and running.

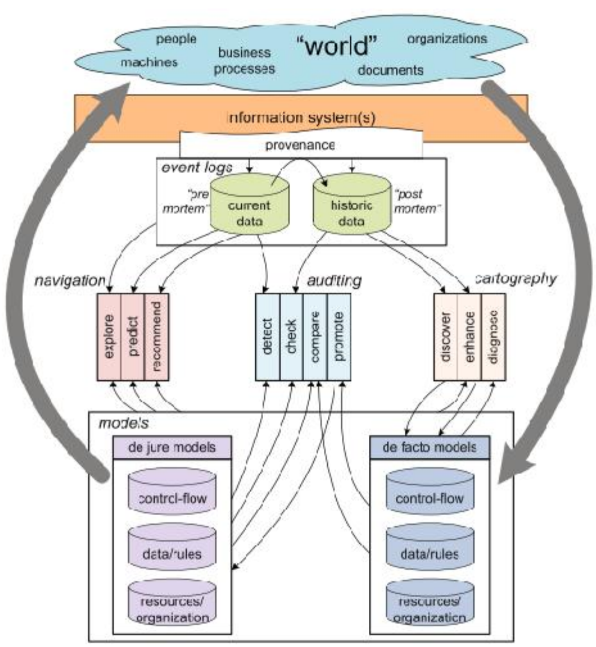


Figure 1.2 Process Mining Spectrum
(Source: Aalst, 2016)

In a business context, the activities performed by the employees and customers can be logged into an information system. For example, in an e-commerce website, the information about the customer's journey since the time they register, to their purchases, until they decided to de-register from the website. Not only that, some activities that executed automatically by a system, e.g., the cancellation of an order if the payment is not received within a certain time limit, are also captured in the event log.

Aalst introduced process mining in 2004 (Aalst & Weijters, 2004). The idea of process mining is to see the behavior recorded in the event log, then creates a process model that represents the process. In other words, **process mining** is a data analytics technique aim to extract process-related information (Aalst, 2016). Process mining is a field with a very broad spectrum. Figure 1.2 depicts the scope of process ming topics.

A business process consists of a set of activities that are performed in coordination in an organizational and technical environment that jointly realize a business goal (Weske, 2012). A business process entails different organizational aspects such as functions, business artifacts, humans, and software systems (Dumas et al., 2013). These different aspects of the process can also be called process "perspectives." Below are the process perspectives that can be identified through process mining:

- The **control-flow perspective** is about the ordering of activities. The goal of mining this perspective is to find a good characterization of all possible paths, e.g., expressed in terms of a Petri Net or some other notation (Aalst, 2016).
- The **organizational perspective** indicates who or what performs which activity. This perspective can be also called as the "resource" perspective. Resource is a generic term to refer to anyone or anything involved in the performance of a process activity, such as process participant, i.e., an individual person like the employee John Smith; a software system, for example a server or a software application; or an equipment, such as a printer or a manufacturing plant (Dumas, et al., 2013).
- The **time perspective** is focused on the timing and frequency of events. The presence of timestamps enables the discovery of bottlenecks, the analysis of service levels, the monitoring time of resource utilization, and the prediction of remaining processing times of running cases (Aalst, 2016).
- The **case perspective** focuses on the properties of cases beyond the path it takes (control-flow) or their originators (resource). Specifically, it focuses on the behavior, properties and data elements associated with individual process instances (Ferreira & Alves, 2012).

Event log is the input for process mining. Provenance is a term that describes a systematic, reliable, and trustworthy recording of event log (Aalst, 2016). Event data that are stored in the provenance can be divided into "post-mortem" and "pre-mortem" data. Post-mortem data is a historical data, i.e., it records the process instances that have completed. Meanwhile, if the process instances are still running, then the event data is a pre-mortem data.

Using the recorded event data, several process mining activities can be performed. There are three categorizations of process mining activity. First, the cartography activities focus on describing the operational process through the constructions of a process model. Process discovery, conformance check, and enhancement are the three main activity types of process mining. They all belong to the category of cartography. Therefore, they only analyze post-mortem data.

Second, in the auditing category, the activities are used to check whether business processes are executed within certain boundaries set by managers, governments, and other stakeholders (Hee et al., 2010). Lastly, the remaining process mining activities aim at business process navigation, e.g., to make predictions about the future of a particular case and guide the user in selecting suitable actions (Aalst, 2016). The last two categories require pre-mortem event data. The activities that belong to the auditing and navigation category will not be discussed in this study.

Table 1.1 Event Log Example

Case id	Event id	Timestamp	Activity	Customer id	Quantity
1	35654423	30-12-2010:11.02	Registration	102910	Null
	35654424	30-12-2010:12.30	Product Order	102910	10
	35654425	30-12-2010:12.33	Payment	102910	Null
	35654426	30-12-2010:12:45	Purchase Confirmed	102910	Null
	35654427	31-12-2010:10:11	Product Received	102910	Null
2	35654433	31-12-2010:08.33	Registration	102293	Null
	35654434	31-12-2010:08.37	Product Order	102293	100
	35654435	31-12-2010:12.33	Purchase Rejected	102293	Null

Table 1.1 displays an example of a post-mortem event log. Simple event log is the minimum of process mining, it contains multi-set of “traces” over some set of “events” (Aalst, 2016). Events are the smallest data unit in process mining and occur when an activity in a process is executed (Rafiei et al., 2018). A trace is a sequence of events and represents for one instance how a process is executed (Rafiei et al., 2018). In other words, simple event log contains at least “cased id” and event classifier column such as “event id” or “activity” in the event log. In process mining, simple event log can be used for process discovery. In **process discovery**, a process model that captures the behavior of an event log in a representative way is constructed (Dumas et al., 2013).

The constructed model in process discovery must represent the actual control-flow perspective, i.e., the ordering of activity executions of the process. However, since the construction of the model in process mining relies on the discovery algorithm, there is a likelihood that the constructed model does not conform to reality. Therefore, after the process discovery, the process model needs to be compared to the event log of the same process so that the discrepancies between them can be identified. This activity is called **conformance check**.

Most process mining approaches focus only on the control-flow perspective, e.g., process discovery and conformance check (De Leoni & Aalst, 2013). However, oftentimes, the event log contains more than the information about activity order. For example, the attribute timestamps in the event log can be used to know about the time perspective. If the event log contains information about resources, resource-related models can be constructed, e.g., a social network showing how people work together in an organization (Aalst, 2016). Sometimes, the event log also has more attributes like the size of an order, the name of the supplier, etc. These attributes might determine the path taken by each case. For example, the online purchase in Table 1.1, if the order quantity exceeds 10, then the order will be rejected. This is an example that describes the case perspectives.

The analysis of the time, organizational and case perspectives can provide more insight into the process. In process mining, the incorporation of these additional perspectives is called the **enhancement** activity.

1.3. Motivation

The traditional way of process modeling through meetings and discussions usually takes from a half to two years, depending on the size of the organization and the time the managers participating in the team can allocate to do the work (Harmon, 2014). In addition, most process models that are made by hand are not based on a rigorous analysis of existing process data (Aalst, 2013). Few organizations have the resources required to model all their processes in detail, to rigorously analyze and redesign each of them, to deploy automation technology in order to support each of these processes, and finally to continuously monitor the performance of all processes in detail (Dumas et al., 2013).

On the other hand, discovering a process model using process mining provides a process model based on the actual observed behavior, without the need to perform interviews (Adriansyah & Buijs, 2013). However, process mining is a field with a broad spectrum. Numerous process mining techniques, tools, and plug-ins are available. Just ProM, the leading open-source process mining framework, already provides more than 1500 plug-ins (Aalst, 2016).

The end-users of process mining software can be categorized based on their knowledge about process mining. There may be experts that need to be able to answer one-of-a-kind questions requiring ad-hoc data extractions, complex data transformations, and sophisticated analysis techniques (Aalst, 2016). In contrast, there can be also some end-users with limited knowledge in process mining, and they want to look at the overall process overviews of their business through a quick process diagnostics. The vast options of process mining practices might be overwhelming for novice process mining users to get started in using process mining tools. They might be confronted with the concern about which process mining approach to choose and how to execute them.

The inability of novice users to getting started in performing process analysis using process mining tools can result in a low "perceived ease-of-use." Perceived ease of use is the degree to which a person believes that using a particular system would be free of effort (Davis, 1989). In contrast, users who experienced successful interaction with a system have a positive "self-efficacy" towards it. Self-efficacy relates to one's belief in one's capability to perform a task (Bandura, 1977). Both self-efficacy and perceived ease of use can influence how a person believes that using a particular system would enhance his or her job performance, i.e., perceived usefulness of the system (Davis, 1989). If the users feel that the process mining tools are not useful and hard to operate, it might result in the refusal to use the tools.

1.4. Research Goal

This research serves dual purposes. First, like scientific research in general, we want to answer our research questions and propose a solution to the research problem we identified. In addition to that, this study also has an academical purpose as the final project of the main author to obtain a master's degree in business information technology at the University of Twente.

The main goal of this study is to produce an artifact that can guides enhancement activity in process mining, through the exploration of the current state of the art of enhancement practices in the process mining field. Whereas, the objective of the artifact is to demonstrate a successful interaction with the process mining tools so that it can increase the perceived ease of use and self-efficacy of novice users towards process mining tools. The following research questions constructed with the intention of achieving our research goal:

> **RQ1: What is the current state-of-the-art of enhancement practices in process mining?**
> First and foremost, we explore the current state of the art of enhancement in the process mining field. In chapter 2, a systematic literature review that shows the trends of process mining practices within the scope of enhancement is presented. Section 2.5 summarizes the findings of the literature review.
>
> **RQ2: How can we assist novice process mining users in conducting enhancement in process mining?**
> Chapter 3 elaborates the design and the development of the artifact based on the current state-of-the-art of enhancement in process mining identified in chapter 2. The development phase produces the initial design of the artifact. Next, the demonstration of the artifact application explained in chapter 4. Because design science methodology is iterative, the information gained in the demonstration phase is an input for the second version of the artifact design. Lastly, findings from the validation are used to create the final version of the artifact design.
>
> **RQ3: Does the proposed artifact applicable in the real-life case?**
> Lastly, we apply the artifact in a real-life case to validate it. First, the author himself/herself apply it to real-life data sets. After that, the artifact is evaluated by two novice process mining users. The result of their evaluation is then compared to the results of two more experienced process mining users. In addition, the artifact is also evaluated by the User Acceptance Test. The evaluation process of our research can be found in chapter 5.

1.5. Research Plan

Design science is the design and investigation of artifacts in context (Wieringa, 2014). The idea of design science is to interact the artifacts to a problem context so that it can improve something in that context. In this study, we use the design science research methodology proposed by Puffers et al. (2007). Figure 1.3 shows the activities involved in the design science research for information system field.

At the start of the project, we did not restrict our scope to any specific part of process mining. Instead, we sent research collaboration invitations to approximately 15 different companies. The example of the invitation letter can be found in Appendix A. Unfortunately, more than half of the invitee did not respond to our request. The companies that responded were either concerned about data confidentiality or had a complicated research collaboration procedure. In consideration of the project's time limitation, we decided to do a problem exploration in the form of a systematic literature review. To ensure the systematicness of the problem exploration, we use a framework to conduct a scoping study introduced by Arksey and O'Malley (2005). The literature review provides an answer for RQ1. After problem exploration, the objective of the solution is identified. Then, the design and demonstration phase provide the answer to the RQ2. Afterward, the evaluation phase answers the RQ3.

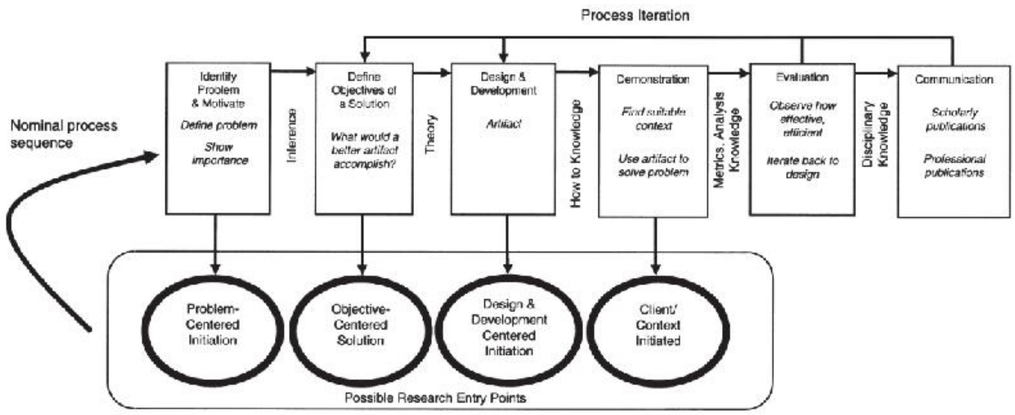


Figure 1.3 Design Science Research Methodology
(Source: Puffer, 2007)

1.6. Research Approach

Various research approaches are used in this study. Below is the description of the approaches used in this study:

- Scoping Study
  Scoping study is a form of a systematic literature review that aims to extract as much relevant data from each piece of literature as possible, including methodology, finding, variables, etc. (Xiao & Watson, 2017). The aim of the scoping study is to provide a snapshot of enhancement in process mining and a complete overview of what has been done. Scoping study is used for problem exploration phase.
- Elimination by Aspects
  EBA is a heuristic decision making where decision makers gradually reduce the number of alternatives in a choice set, starting with the most important one. One cue is evaluated at a time until fewer and fewer alternatives remain in the set of available options (Tversky, 1972). This approach is used for the selection of process mining techniques and plugins in the design and development phase.
- Narrative Review
  A narrative review is a study that focused on gathering relevant information that provides both context and substance to the author's overall argument (Xiao & Watson, 2017). This approach is used to support the selection of process mining techniques and plugins in the design and development phase.
- Case Mechanism Experiment
  A case mechanism experiment in research is a test in which the researcher applies stimuli to a model and explains the response in terms of mechanisms internal to the model (Wieringa, 2014). In this study, after the creation of the artifact, we demonstrate the artifact in real-life data set. The data set is an open data provided by the Business Process Intelligence (BPI) community from the University of Eindhoven.

- Technical Action Research
  In technical action research, the researcher uses an artifact to help a client and explains the outcome architecturally in terms of the case (Wieringa, 2014). In this study, we present our artifact to two students at the University of Twente who are novice process mining users. The result of their process analysis is compared to another two students who is a more experienced user in process mining.
- User Acceptance Test
  The first group of two students, who are a novice process mining users, also evaluate the perceived usefulness of our artifact and the perceived usefulness of the ProM tools by filling out a questionnaire.

1.7. Chapter Summary

This study uses the design science research methodology to conduct the research. First, a scoping study is conducted in the problem exploration phase. It results in the identification of the problem and the objective of the solution. Then, EBA and narrative review are used to develop the artifact design. Through case mechanism experiment, the application of the artifact using real-life data sets is demonstrated. Lastly, the artifact is evaluated through technical action research and user acceptance test.

This thesis document is structured as follows: Chapter 2 is about the state-of-the-art enhancement activity in the process mining field. Then, we present the design and development of the artifact in chapter 3. After that, we demonstrate the use of our artifact in chapter 4. In chapter 5, we evaluate the usability of our artifact. Lastly, we discuss our conclusion, limitation, and recommendation for future work in chapter 6.

2. Background and Related Works

This chapter is about the current state-of-the-art of enhancement in process mining. First, the literature review method is presented. Then, it discusses the most cited definition of enhancement in process mining. It continues with information about process enhancement application in various fields and industries. Lastly, the variety of process enhancement practices in terms of enhancement activity, event log mechanism, process mining tools, and process modeling language is discussed.

2.1. Literature Review Method

In the previous chapter, the three main activities in process mining are indicated. The last activity, extension, described as "the extension or improvement of an existing process model using information about the actual process recorded in some event log" (Aalst, 2016). Meaning that the enhancement activity follows the process discovery and conformance check, i.e., it can be conducted only if the process discovery and conformance check are performed beforehand.

Enhancement activity aims to produce an analysis of a process from various perspectives such as the control-flow discovery, time, organizational and case perspective. However, there are numerous techniques and practices that can be used to produce the analysis. Therefore, this chapter explores the variety of process enhancement practices through problem exploration in the form of a systematic literature review. The literature study elaborated in this chapter has been presented at the 8th International Symposium on Data-Driven Process Discovery and Analysis[1].

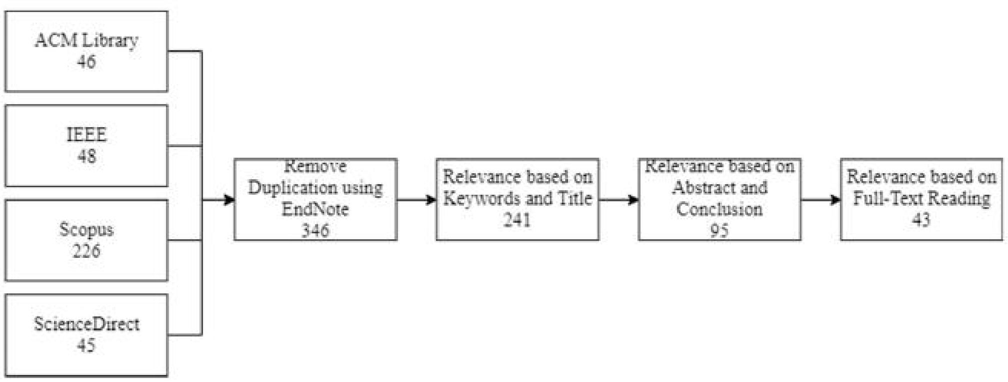


Figure 2.1 Literature Search Statistics

Figure 2.1 depicts the literature search statistics. The literature search started by querying relevant phrases through several online databases. We acquired 46 articles from ACM Library by searching articles containing at least of these phrases "Process Enhancement," "Enhancement," "Perspective" and "Repair"; and limiting the results with the keyword "Process Mining." We found another 48 Articles from IEEE Xplore and 226 from Scopus by applying the same query. Lastly, we searched for articles with the phrase "Process Enhancement" and the keywords "Process Mining" in ScienceDirect, returning 45 articles. The returned search lists then imported to the reference management software named EndNote to further underwent the selection process. After

[1] the published short-paper version of the study is available in http://ceur-ws.org/Vol-2270/short4.pdf

utilizing the automatic duplication identifier in EndNote, a total of 346 articles identified. These articles then selected with the criteria of:

Include only the literature that is written in English.

Include only the latest version of the article.

Include only literature that discusses process enhancement that falls within the process mining field.

Include only the articles that implicitly indicate the utilization of process mining techniques, conception or tools in the scope of process enhancement.

Exclude duplicating article. For instance, in some cases, there are identical articles published by two different publishers. In that case, we eliminated the older version.

Exclude the articles that the full-text version is not available via the University of Twente's access portal.

The initial relevance determined by the title and the keywords of the literature. This phase aims to eliminate the articles that fall outside the scope of process mining and to eliminate duplications. After initial screening, 241 articles are identified. The next step is to narrow the scope down to the topic of enhancement within the process mining field by reading the article's abstract and conclusion; this reduced the number of articles to 95. After the full-text reading, the final total of 43 articles included in the study. We want to underline that the 43 articles are referred as the numbered citation in **square brackets**, even though the referencing format used in this document is APA. The differentiation is intended to ensure conciseness and readability. The table that includes the indexing and the complete summarization of the article lists can be found in Appendix B.

2.2. Enhancement Definition

The word "enhances", according to Merriam-Webster dictionary, defined as the increase or improve in value, quality, desirability, or attractiveness (Merriam-Webster, n.d.). Out of 43 articles included in our literature study, 10 explicitly mention the definition or concept of enhancement in process mining. Six among them [1] [7] [9] [11] [16] [36] adopted Aalst definition of enhancement, and the other three [20] [21] [25] described the conception of extension task instead of the process enhancement itself. Process extension is "a form of process enhancement where apriori model is extended with a new aspect or perspective" also defined by Aalst (Aalst, 2016). Another definition of enhancement is "a mining process that aims to improve the mining performance, as well as the readability of discovered models" as expressed by Schönig, S., et al. [27]. Their definition, however, is the enhancement definition with the use of tools called DeclareMiner.

2.3. Process Enhancement Application

In this section, we discuss the process enhancement application in various domains. Table 2.1 indicates the distribution of the process enhancement within various industrial domains, where the medical field is the most studied one. From an industry context, there are two categories of approaches. First, the proposed framework or approach is domain-specific. Second, the approach only applied in a specific domain(s).

The most studied field within process enhancement is Medical. Different outcomes are expected out of different approaches proposed within this field. For instance, both [20] and [21] focuses on delivering insights into the Careflow by looking at the control-flow, organizational and performance perspective. They both use a general approach that is easy to implement, therefore, good as a starting point for medical institutions for getting to know how process mining works. However, the results produce understandable spaghetti-like models showing all details, instead of high-level information.

Table 2.1 Process Enhancement Application per Industry

Industry	Articles
Medical	Dutch Academic Hospital [3] [4] [21] [29], Private Dental Practice [20], Hospital in Sao Sebastiao [13], Chinese Hospital [37], Other [36] [30]
Governmental	Dutch Municipality [12] [26] [28] [29], Dutch UWV [10], Dutch Public Works Department [32]
Financial	Dutch Financial Institute [1] [3] [23] [4] [39], Bank in Bosnia and Herzegovina [11]
Educational	University Business Trip Management [27] [41] University in Thailand [2]
Law Enforcement	Road Fines Management in Italy [5] [18] [19] [28] [29]
Other	Service Delivery [6], Social Media [17], Guangzhou Port [33], Telecommunication Company [35], Volvo IT Belgium [4] [16] [23], Bottle Manufacturer [40]

Another domain-specific approach is the discovery of the main flow in the invoice handling process [32] and the method of applying process mining to the distribution of non-alcoholic beverage manufacturer [40]. The former [32] is an approach that analyzed the processing of invoices sent by the various subcontractors and suppliers from three different perspectives: the process perspective, the organizational perspective, and the case perspective. The latter [40] guides the application of much of the knowledge generated in process mining for a specific industry, in their case non-alcoholic beverages manufacturer. They're aiming to increase the quality of service delivered to customers by making the process more transparent [40]. Besides the approaches above, the rest of the articles fall into the second category where the approaches are potentially applicable to the other field but in the study, uses a specific event log from a specific industry. In this case, the distinctive variable of those approaches lies in their objectives and constraints.

Table 2.2 Real-life Datasets Sources

Datasets	Articles
Dutch Municipality - BPI 2012	[12] [26] [29]
Dutch Academic Hospital - BPI 2011	[3] [4] [21] [29]
Dutch Financial Institute - BPI 2012	[1] [3] [23] [4] [39]
Dutch Governmental Corp	[10] [8] [32]
Volvo IT Belgium - BPI 2013	[4] [16] [23]
Road Fines Management Italy – Fines 2015	[5] [18] [19] [28] [29]

Another interesting finding that we acquired while examining the process enhancement application throughout different industries is the contribution of BPI Community, in this case, to the availability of event logs. Majority of articles examined in this study propose an approach of process enhancement; therefore, in order to validate their approaches, the researcher need to use real-life event log. Table 2.2 shows that a lot of these studies, indeed, validate their proposal by using the event log made available through BPI community over the years.

2.4. Enhancement Variety

The relevant literature demonstrates the wide spectrum of process mining topics. Even with the same goal to extract process-related information, the implementation of process mining techniques is different in each article. They are different in terms of the enhancement activity they perform, and the mechanism of event data attainment, the tools they use, and the process modeling language of their output.

2.4.1. Enhancement Task

There are two task types of enhancement: repair and extension (Aalst, 2016). After the process model discovery, the conformance of the model to the log can be checked. This conformance checking process describes whether the process model reflects the actual process in the recorded event log. If the process model does not reflect reality, then the model needs to be repaired so they can replay most of the event log is the next step. The model can also be extended by additional perspectives. We argue that the mining of those additional perspectives is important to process enhancement, regardless of the incorporation of the perspective to the a priori model.

> **Definition 1:** Enhancement in Process Mining
> The effort of improving the business process of an entity through the discovery of control-flow perspective, and the incorporation of additional insight of other perspectives, namely organizational, time and case perspective.

Repair consists of modifying the model to better reflect reality (Aalst, 2016). A total of nine articles discusses the process model repair approach with different objectives and constraint. See Table 2.1. for details about the distribution of the articles in process model repair.

Vahland and Aalst [12] conceptualized the term Model Repair in their article as repairing a process model with regard to a log such that the resulting model can replay the log (i.e., conforms to it) and is as similar as possible to the original model. Their article is a good starting point for practitioners who wish to get started in understanding the concept of model repair. In addition to that, Cervantes et al. [5] advocates for an interactive and incremental approach to process model repair, where differences between the model and the log are visually displayed to the user, and the user repairs each difference manually based on the provided visual guidance. This approach is an advisable next step for understanding the practices of model repair; the manual and step-by-step procedure makes it easy to understand. Another manual process model repair is proposed by Adriansyah and Buijs [1], they iteratively repair process model through the alignment between execution traces in the event log and the process models to identify non-deviating activity executions as much as possible and use them to measure the performance of process executions. Their approach is suitable for practitioners whose goal is to have reliable performance information.

Table 2.3 Article Distribution on Process Enhancement Types and Perspectives

Type	Perspective	Articles
Repair	Control-flow	[1] [4] [5] [7] [12] [22] [23] [24] [36]
Extension	Organization	[2] [3] [8] [9] [10] [11] [13] [14] [15] [16] [17] [18] [19] [20] [21] [25] [26] [27] [28] [29] [30] [31] [32] [33] [35] [37] [38] [40] [41] [42] [43]
	Time	[6] [10] [15] [16] [18] [19] [26] [28] [29] [32]
	Case	[10] [15] [18] [19] [26] [28] [32] [34] [39]

On the other hand, when practitioners discovered a spaghetti-like process model, they can consider adopting Appice and Malerba [4] algorithm by handling variability in the recorded behavior of existing logs and facilitating process model discovery. The non-conformance that is resulted by variability in recorded behavior can also be solved by [22], a repairing model that decomposes the process model and event log into model fragments and sub-logs, then selectively repair the fragments. This allows keeping the initial model structure as much as possible. However, their approach has its own limitations. The approach is yet to be tested in real life and it also significantly reduces precisions. Besides that, the non-conformance might also be a result of the incompleteness in the event log. In a case where the recorded event logs some events are missing time information, Solti et al. [24] propose a method to make them available by combining stochastic Petri Nets, alignments, and Bayesian networks. Contrasting with sub-logs and sub-process approaches like described by [4] and [22], Zhang, et al. propose a repair approach using information of causal relation and a concurrent relation of the process models. The uses of clustering approach in process model with causal and concurrent relations might result in the repaired model with a lower precision value, and a lot of useless traces generated by the repaired model cannot be observed in the event log. Lastly, if out of all approaches above, practitioners still not sure which model repair they need, we suggest adopting the approach [23] where they check all possible combinations of repair actions, i.e., inserting and skipping activities, to find an optimal repaired model within the range set by the available repair resources.

Extension is another type of process enhancement consists of adding a new perspective to the process model by cross-correlating it with the log 34 article that discusses process model extension. From literature analysis, we have found that there are three perspectives that can be added to the process model: time, case and organizational perspective. For details about the distribution of the articles in process extension, see Table 2.1.

- Time perspective. Articles like [15] [16] [29] does not exclusively discuss the time perspective, but they incorporate it to their approach. This incorporation can be further utilized as performance analysis. The article that only specifically uses time perspective is Ballambettu et al. [6] in which they include time perspectives such as flow time and activity execution times as determinant for process variances.
- The case perspective. Werner [34] introduce a specification of Colored Petri Nets that enables the modeling of the data perspective for a specific application domain and Van Eck, et al. [39] present an approach to provide insights into processes that can be considered from multiple state-based perspectives.

- The organizational. As shown in the table, the mostly studied perspective is the organizational perspective, or sometimes referred to as resource perspective. The total of 30 out of 43 articles considered the organizational perspective in their approach. Articles like [2] [8] [14] [16] [17] [20] [21] [25] [30] [32] [33] [35] [38] [40] incorporate organizational perspectives in the form of role analysis. When the event log contains the information about who executed the event, the role analysis then can take place. Even though [2] and [8] the main objectives are to provide a quick process overview, the incorporation of organizational perspectives provides the opportunity to discover information like resource groups or hierarchy. This can be further utilized to find factors like the division of work within departments of the organization. Zhao et al. [37] capture various types of patterns by mining resource characteristics and task preference from past process executions so that the right resources could be recommended to improve resource utilization. Another approach [13] propose the discovery of groups, but they use working-together metrics instead of task similarity metrics. In other words, they discover user groups based on actual user interactions. Appice et al. [3] propose a plug-in called TOSTracker that not only can discover communities, but they also consider the dynamic event log. The plug-ins enable the discovering and tracking changes in resource communities (organization structures) of a business process over time.

2.4.1. Event Log Mechanism

Event log is the requirement for process mining. However, the raw data is usually hidden in all kind of data sources. A data source may be a simple flat file, an Excel spreadsheet, a transaction log, or a database table (Aalst, 2016). An information system that produces event log is called Process-Aware Information Systems. Examples of PAISs are workflow management system, ERP system, case handling system, CRM system, etc. (Song & Aalst, 2008).

Definition 2: Event Log
Event log contains data related to a single process with the assumption of (Aalst, 2016):

- A process consists of cases.
- A case consists of events such that each event relates to precisely one case.
- Events within a case are ordered.
- Events can have attributes. Examples of typical attribute names are activity, time, costs, and resource.

We found two types of mechanism that are used in enhancement activity in process mining through our literature review which is from database and file. First, event log can be stored in a **file** like CSV, Excel file, or XES. We found that the most used file format to store event data is XES, the successor of MXML (Aalst, 2016). The log can contain an arbitrary number of trace objects, and each of them describes the execution of one specific instance, or case, of the logged process (Fluxicon Process Laboratories, 2009).

Second, event log can be also loaded from a **database** via tools that support periodical event loading. It can be loaded through an adapter of particular application like SAP and SalesForce, or loaded from a stream of events emitted through an event bus or web services (Aalst, 2016). We found a sparse article that discusses process mining with the event log stored in SQL databases. Stroinski et al. [38] algorithm enable the discovering process models from event log of Communicating Resource System, (a set of independent and hierarchically composed resources communicate to realize a business process like RESTful Web services and cloud computing platforms) including the feature of resource perspective. In addition, Schönig et al. [43] introduce a mining approach that incorporates the organizational perspective that directly works on relational event data by querying the log with conventional SQL.

2.4.2. Tools

Another attribute to be highlighted is the tools used in practices. Process mining software tools are capable of creating a process model from event logs, existing process templates, or user applications found in enterprise information (Devi, Kumudavalli & Sudhamani, 2017). As shown in Table 2.2, most articles use ProM as their tools, followed by Disco. The rest of the studies use various tools such as SQL, Weka, DpiL Miner, etc. The rest six articles are do not uses any tools in their approach, or they do not explicitly mention what tools they used in their study.

Table 2.4 Tools in Process Enhancement

Tools	Articles
ProM	[1] [2] [4] [6] [8] [9] [10] [11] [12] [13] [15] [16] [17] [18] [19] [20] [21] [22] [23] [26] [27] [34] [35] [36] [38] [40] [41]
Disco Fluxicon	[14] [15] [35] [40]
Other	SQL [15] [28] [43], Weka [10], MP Declare [29], R [30], Proview [25], BPMN Oracle [11], Apromore [5], DpiL Miner [27], SPADA [31]
Unavailable	[3] [7] [24] [37] [38]

ProM is a project by the Process Mining Group, Eindhoven Technical University that results in an extensible framework that supports a wide variety of process mining techniques in the form of plug-ins. It is platform independent as it is implemented in Java and can be downloaded free of charge (Process Mining Group, 2014). **Disco** by Fluxicon is another commercial software in Process Mining, known as an enterprise version of the ProM.

Table 2.5 shows the difference of feature between ProM and Disco. The support of a tool for a given feature is labeled by the mark ✓ and features that are not supported are labeled by the mark ✗. Section 2.4.1 explains that event log can be loaded from a database or it can be stored in a file. Both Disco and ProM use event log that is stored in a file. ProM supports the MXML and its successor the XES format (Process Mining Group, 2014). ProM has no limit for the log size. Meanwhile, Disco supports a wider range of event log import formats including CSV, MS Excel, MXML, XES, FXL (Fluxicon, n.d.). Although, Disco limit the log size up to five million events.

Both tools support data filtering features. Data filtering can be used for limiting the scope of the log to be analyzed. For example, users can choose the process instances that executed in a certain period of time.

Table 2.5 Comparison between Disco and ProM

Features	ProM	Disco
Import Type Support	MXML, XES	CSV, XLS, XES, MXML and FXL
Import Log Size	Unlimited	Up to 5 million events
License	Open Source	Commercial
Output Model Notation	BPMN, WF, Petri Nets, EPCs, Transition System, and Heuristic Nets	Fuzzy model
Filtering Data	✓	✓
Process Discovery	✓	✓
Process Visualization	✓	✓
Conformance Checking	✓	✗
Performance Reporting	✓	✓
Social Network Mining	✓	✗
Decision Rule Mining	✓	✗

Note. Data for import Disco import and license from Fluxicon (n.d.), for ProM import and license from Process Mining Group (2014), and for features from Devi et al. (2017)

ProM and Disco can discover control-flow perspective of the process since they both have features that support process discovery and process visualization. The process discovery produces a static process model. Whereas in process visualization, process executions are animated (Devi et al., 2017).

Performance check is supported in both tools. For performance analysis, the fitness of the process model with respect to the event log needs to be high (Adriansyah & Buijs, 2013). However, the conformance check is not supported in Disco. Therefore, the information about the fitness is unknown.

ProM also has features to discover organizational and case perspective. The organizational perspective can be identified through social network analysis. The case perspective can be identified through decision rule mining.

2.4.3. Process Modelling Language

A process model is the output of the process discovery activity. We identified two types of process modeling language namely declarative and procedural. Procedural languages only allow for activities that are explicitly triggered through control-flow, while in a declarative language "everything is possible unless explicitly forbidden" (Aalst, 2016). A process modeling language is **declarative** when it explicitly takes into account the business concerns that govern business processes (Goedertier & Vanthienen, 2007).

Definition 3: Process Model
A process model is a type description that captures the properties common to a class of processes (Munch et al., 2012). It represents a blueprint for a set of process instances with a similar structure.

Aalst (2016) briefly discussed the declarative process model in the form of language called Declare, a language that uses a graphical notation and semantics based on Linear Temporal Logic. LTL is an example of a temporal logic that, in addition to classical logical operators, uses temporal operators such as: always (□), eventually (◊), until (␣), weak until (*W*), and next time (○) (Clarke, Grumberg, & Peled, 1999). LTL can express business rules such as "a specific activity A should be executed eventually if activity B is executed," "after three months of activity A, activity B should be executed," etc.

Table 2.6 Differences between procedural and declarative modeling language

Aspect	***Procedural modeling***	***Declarative modeling***
Business concerns	Implicit	Explicit
Execution scenario	Explicit	Implicit
Execution mechanism	State-driven	Goal-driven
Modality	What *must*	What *must, ought, can*
Rule enforcement	Procedural (what, when, how)	Declarative (what)
Communication	Explicit (how)	Implicit (what)

Table 3.1 summarizes the difference between procedural and declarative process modeling language. The distinguishing difference between the two are (Meersman et al., 2007):

- The information about how the process model should be executed is defined explicitly in the procedural business process model, but the reasoning behind the design choice is only implicitly tracked. In declarative process modeling, the business process is governed explicitly through business rules and vocabulary.
- The enforcement of business rules through control-flow-based modeling construct is possible in procedural modeling; while declarative process modeling does not make use of control-flow to indicate when and how business rules are to be enforced.
- Since procedural process models describe the ordering of events, it has a necessity modality (what must) attached, whereas declarative process languages allow for other modalities like intention (what ought), possibility (what can) and certainty (what is).
- The communication purpose of procedural process models is to inform external stakeholder about the execution of business events, while declarative process model intended to describe business events and business concepts.

Only a few articles focus on enhancement using declarative language, among them is [27] which propose a process mining approach to discover resource-aware declarative process models. The framework by [41] supports the discovery of patterns related to resource assignment. The possible reason to explain the lack of discussion is that to discover a declarative process model, it may need more attributes than procedural modeling language. In section 1.2, we explain that "simple event log," an event log that contains "case id" and event classifier, can be used to know about activity ordering in the process.

Hence, a simple event log can be used to discover a procedural model. However, for the reinforcement of business rules in the declarative model, such as "activity B executed when the supply is below 100 units" or "'activity C will be canceled if the payment is not received within 24 hours", attributes such as timestamp and order quantity is needed. In other words, simple event log is not enough to discover such model.

Procedural modeling language aims to describe end-to-end processes and allow for activities that are explicitly triggered through control-flow (Aalst, 2016). Most of the articles in the literature with a procedural modeling language use Petri Net and the Business Process Modeling Notation (BPMN) as their modeling notation.

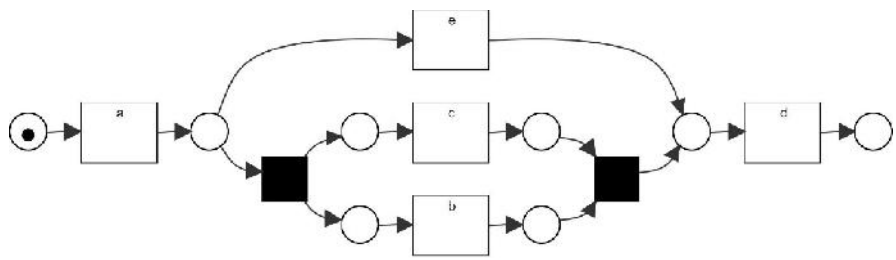


Figure 2.2 Example of Petri Net Model

A **Petri Net** is a bipartite graph with two types of nodes: transitions and places. The transitions represent activities that can be executed, and places represent states (intermediate or final) that the process can reach (Burattin, 2015). Figure 2.2 shows the example of a Petri Net model. When modeling business processes in Petri Nets, a subclass of Petri Net known as workflow net is used. A workflow net is a Petri Net with a dedicated source place where the process starts and a dedicated sink place where the process ends (Aalst, 2016). Meaning that it should clearly indicate the initiation and the termination of the process.

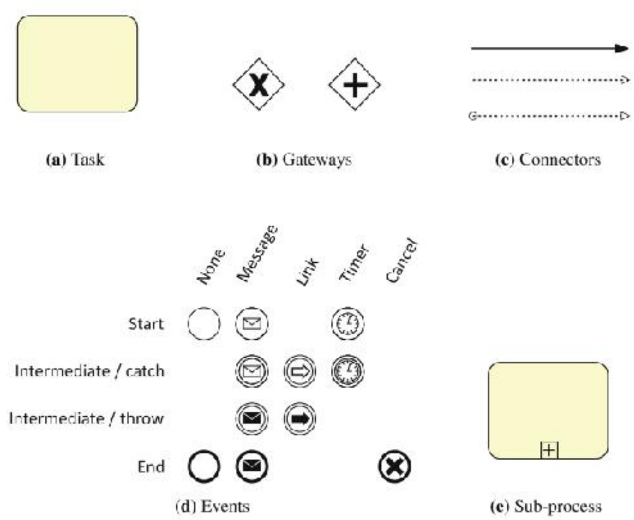


Figure 2.3 Basic Component of BPMN Notation
(Source: Burattin, 2015)

Business Process Modeling and Notation (BPMN) is another modeling notation that are mostly discussed in the literature after Petri Net. BPMN is the result of an agreement among multiple tool vendors, that agreed on the standardization of a single notation. Therefore, it is used in many real cases, and many tools adopt it daily (Burattin, 2015). Burattin et al. [9] and Djedovic et al. [11] discusses the enhancement through organizational perspective in the process represented by the BPMN model. The identification of roles based on the detection of the handover of roles in the BPMN model is discussed in [9]. In addition, [11] integrate the statistical analysis into organization perspective so they can find the distribution of users work and distribution of instance generation in the process.

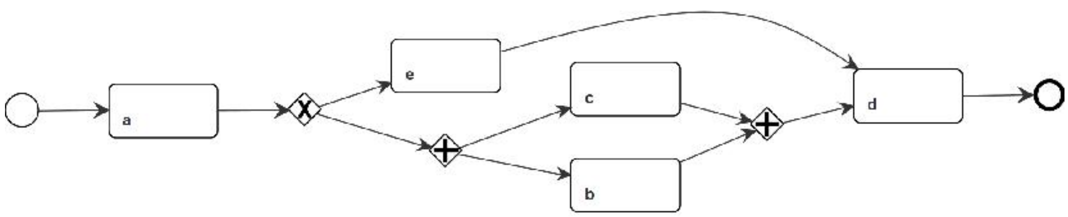


Figure 2.4 Example of BPMN Model

A standard BPMN provides businesses with the capability of understanding their internal business procedures in a graphical notation and will give organizations the ability to communicate these procedures in a standard manner (Object Management Group, n.d.). Figure 2.3 shows the basic components that are used in the BPMN notation. The main components of a BPMN diagram are:

- Task is an atomic activity that is included within a process (Object Management Group, 2006). In other words, task is a single unit of work, and if the task requires many steps, it will be called as an activity (Dumas, et al., 2013). Subprocess can be used to hide different level of abstraction of work (Burattin, 2015).
- Gateway is modeling elements that are used to control how Sequence Flows interact as they converge and diverge within a process (Object Management Group, 2006). Unlike Petri Net, one cannot have events with multiple incoming or outgoing arcs using BPMN; splitting and joining needs to be done using gateways (Aalst, 2016).
- Connectors represent the flow of between component of the graph (Burattin, 2015). There are three types of connectors: a Sequence Flow is used to show the order that activities, a Message Flow is used to show the flow of messages between two entities, and an Association is used to associate data, information, and artifacts with flow objects (Object Management Group, 2006).
- Events correspond to things that happen atomically, meaning that they have no duration, e.g., the arrival of a shipment in the customer's house (Dumas, et al., 2013). Event is something that "happens" during the course of a business process that can affect the flow of the process, and usually have a trigger or a result (Object Management Group, 2006). Event can start, interrupt, or end the flow of the process. An event is comparable to a place in a Petri Net (Aalst, 2016).

Section 2.4.2 describes that the ProM can produce both Petri Net and BPMN model. However, even though the complete specification of BPMN 1.0 defines 38 constructs and attributes, ProM only uses the constructs that related to control-flow perspectives. It only produces a flat model represented by start and end events, tasks, gateways and sequence flows (Kalenkova et al., 2017). Figure 2.4 depicts an example of a BPMN model that is produced by process discovery using ProM. Moreover, the performance and conformance analysis for BPMN is not supported in ProM. To analyze a process represented as a BPMN model using ProM, the model must be converted to a corresponding Petri Net (Kalenkova et al., 2017). Table 3.2 lists the difference between Petri Net and BPMN notation.

Table 2.7 Comparison between Petri Net and BPMN

Aspects	Petri Net	BPMN
Modeling constructs	Places and Transitions	Task, Event, Pool, Lane, Gateway, Normal Flow, Message Flow, etc.
Represents exclusive decision	Yes	Yes
Represents parallel execution	Yes	Yes
Represents Inclusive decision	Yes	Yes
Represents rework and repetition	Yes	Yes
Performance Analysis in ProM	Available	Not Available

Note. Data for modeling constructs from Aalst (2016), and for activity executions from Brabander & Davis (2007) and Aalst (2016).

Both notations are capable of capturing various activity execution that represents non-sequential ordering. An exclusive decision is when the execution of one activity means that the exclusion of the other, for example, when an order is accepted then the activity "reject order" will not be performed. In Figure 2.2 and Figure 2.4, the exclusive decision is after the execution of activity "A." On the other hand, parallel execution is when multiple activities need to be performed in parallel, i.e., at the same time. In Figure 2.2 and Figure 2.4, the activity "C" and "B" executed in parallel. There are also cases where the decision is not mutually exclusive, but it represents two or more activity performed after a decision activity; this execution is an inclusive decision. Lastly, in term of activity execution, there is a scenario when activity needs to be repeated, due to human or system error. Both Petri Net and BPMN are capable of capturing these behaviors using a permutation of XOR-join, AND-join, XOR-split, and AND-split.

2.5. Chapter Summary

In summary, enhancement in process mining is the effort of improving the business process of an entity through the discovery of control-flow perspective, and the incorporation of additional insight of other perspectives, namely organizational, time and case perspective. XES format is de facto exchange format for process mining and has been adopted by IEEE task force on process mining. ProM and Disco are used the most in process mining practices and both support the analysis of XES event log file. However, Disco only focuses on the control-flow perspective. At the other hand, ProM also supports the analysis for Organizational, Time and Case perspectives. ProM can produce a process model in different notations, but the most used in practices are Petri Net and BPMN. However, conformance and performance analysis for BPMN is not supported in ProM.

3. Design and Development

This chapter is about the development of the proposed artifact. First, it describes the targeted users of the artifact. Then it proceeds to describes the proposed process mining approach that consists of three phases including pre-analysis, analysis, and post-analysis. Lastly, the summary of the chapter provides the visualization of the proposed artifact.

3.1. Artifact's Users

Section 1.4 states that the main goal of this study is to produce an artifact that can guides enhancement activity in process mining, through the exploration of the current state of the art of enhancement practices in the process mining field. But first, we need to know whom our artifact is intended for.

In an organization, process analysis using process mining tools is the task of process analyst. However, the process analyst typically has a limited understanding of the actual process execution. In contrast, process owner typically has limited knowledge in process modeling techniques, and they rely on process analyst to formalize their business process in the form of a process model. Process analyst and process owner have complementary roles in the act of process analysis. Table 3.1 highlight the difference between the process owner and process analyst.

Table 3.1 Typical Profile of Process Analyst and Process Owner

Aspect	Process Analyst	Process Owner
Modeling Skill	Strong	Limited
Process Knowledge	Limited	Strong

Note. Data for the differences of process analyst and process owner from Dumas, Rosa, et al. (2013).

Process owner is responsible on the one hand for planning and organizing, and on the other hand for monitoring and controlling the process (Dumas et al., 2013). Process owner encounters an **organizational** challenge to integrate and embed the technical steps of process mining project into the organization mechanism (Burattin, 2015). If we map the process owner task into process mining project, we can see that the process owner is a crucial stakeholder in the pre-analysis and post-analysis phase. In pre-analysis, process owner determines which process needs to be analyzed. In post-analysis, process owner determines what improvement effort will be taken based on the analysis.

Process analyst is the one who has a profound knowledge of business process modeling techniques; and is familiar with languages like BPMN and skilled in organizing information in terms of a process diagram (Dumas et al., 2013). Process analyst encounters **technical** challenges in the process mining project. First, the challenge is to exploit as much information from the recorded log. As explained in section 1.2, an event log may contain more than just information about the activity executions. Process analyst needs to consider other perspectives of the process. Second, the challenge is the configuration of using process mining tools. In section 1.2, we also discuss that spectrum of process mining is very broad. Therefore, a plethora of approach options to conduct process analysis exists. To tackle this challenge, a framework for conducting process analysis using process mining based on the current state-of-the-art of process mining practices is proposed.

In chapter 2, we found that the XES format is de facto exchange format for process mining and has been adopted by IEEE task force on process mining. ProM and Disco are used the most in process mining practices, and both support the analysis of XES event log file. However, Disco only focuses on the control-flow perspective. At the other hand, ProM also supports the analysis for Organizational, Time and Case perspectives. ProM can produce a process model in different notations, but the most used in practices are Petri Net and BPMN. However, conformance and performance analysis for BPMN is not supported in ProM. Therefore, based on the findings in our literature review, we define these criteria for our guideline: using (1) ProM tools by extracting event log in (2) XES format to produce (3) procedural model of the process, more specifically in the form of (4) Petri Net.

To further select the process mining techniques in detail, e.g., the plugins to be used, we use Elimination by Aspect approach. EBA is a heuristic decision making where decision makers gradually reduce the number of alternatives in a choice set, starting with the most important one. One cue is evaluated at a time until fewer and fewer alternatives remain in the set of available options (Tversky, 1972). First, we identify available options that are mostly used in practices, then reflect each option to the goal of this research.

Table 3.2 Process Mining Challenges and Proposed Solution

Role	Focus	Phase	Challenge	Solution
Process Owner	Organizational	Pre-Analysis	Integrate technical steps into organization mechanism	Process identification and prioritization
		Post-Analysis		Improvement recommendation
Process Analyst	Technical	Analysis	Exploit as much information from the recorded log files	Mining additional perspectives
			The configuration of using process mining tools	Specific guidelines with the help of process enhancement literature graph

Table 3.2 summarizes the challenges, the responsible role, and the proposed solution. These aspects, with the addition of process enhancement trends we identified in Chapter 2 and the criteria of the artifact in Chapter 3, are the ingredients of our guideline. Our primary artifact is a guideline for process owner and process analyst to conduct enhancement activity in process mining that consists of three phases. In the first phase, **pre-analysis**, the process owner first identifies the current key business process by creating the process landscape. After that, a process that chosen to be analyzed is chosen by considering the importance, dysfunction and feasibility criteria of the process. The second phase, **analysis**, is conducted by process analyst. It started with the installation of ProM tools, followed by the event log import. Based on the behavior recorded in the event log, the control-flow perspective of the process is discovered using inductive miner plug-in. After that, the additional perspectives of the process are identified[2]. The social network analysis for organizational perspective, the performance analysis for time perspective, and decision mining for case perspective. last phase, **post-analysis**, process owner takes process improvement decision(s) based on the result in the analysis phase.

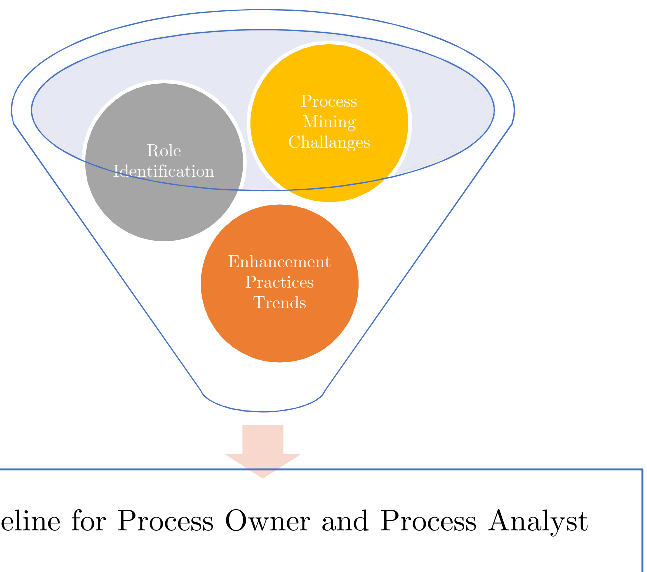


Figure 3.1 Artifact's Ingredient

We want to emphasize that our guideline does not intend to restrict the practice of enhancement in process mining. Therefore, as a secondary artifact, we created a graphical chart that visualizes the available process enhancement literature. This artifact can be used by process analysts who wishes to use practices beyond the one prescribed in the guideline. The graph can be found in Appendix C.

3.2. Pre-Analysis

The modeling of an enterprise-wide business process would be a very complex project and time-consuming. Therefore, a process mining project focus on one key process within the organization. In this phase, we want to help process owner to identify that one process by answering questions like "what key processes are executed within the

[2] definition of process perspectives is in Background and Related Works

organization?" and "which process should we focus on?". These questions can be answered through process identification and process prioritization consecutively.

Process identification is a set of activities aiming to systematically define the set of business processes of a company and establish clear criteria for prioritizing them (Dumas et al., 2013). The output of process identification is process architecture. A business process architecture is an organized overview of business processes that specify their relations, which can be accompanied by guidelines that determine how these processes must be organized (Dijkman et al., 2016). There are many approaches to create process architecture. Historically, the most popular way to define a company's processes has been to put a group of managers in a room and discuss how things get done (Harmon, 2014). This discussion can be refined and accelerated with the support of a formal approach. In a systematic literature review that explores various framework to build a process architecture, it describes that the function-based and object-based structuring approaches appear to be the most useful. In this study, we use Dijkman's function-based method of process identification in consideration of its simplicity and systematicness.

After the process architecture is created, we gain information about what processes are executed within the organization, but another question remains: "which process should we focus on?". This is when the process prioritization takes place. In this study, we use the prioritization approach described in Dumas's book about business process management.

3.2.1. Process Identification

Process architecture can be classified into three different levels of abstraction. Figure 4.2 portrays the abstraction level of process architecture. A lot of process architecture design approach is available. However, most practitioners seek a lower-level guideline for designing business process architectures, in contrast to all-encompassing approaches (Dijkman et al., 2011). In addition, most of the approaches describe only their informal methodology on the relations and structure of the process model collection but do not specify a concrete technique how to get from process models to the certain process architecture level and vice versa (Sabbagh, 2015).

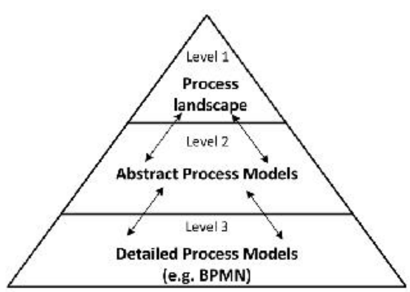


Figure 3.2 Abstraction Level of Process Architecture
(Source: Dumas, et al., 2013)

An approach of process architecture design by Dijkman (n.d.) produces a level-one process architecture called process landscape. His approach is one of the few approaches that describe concrete guidelines to create a process architecture (Sabbagh, 2015). It identifies and hierarchically decomposes business processes based on business functions and case type. Since the creation of process architecture in pre-analysis phase is solely for the purpose of process prioritization, i.e., in choosing what process needs to be analyzed using process mining techniques, the level one abstraction is adequate for pre-analysis. Below are the steps required to produce the process landscape:

i. Identify case types

   In this step, the process owner needs to identify properties to classify various cases within the company. Some properties that are commonly used to classify cases are product type, service type, channel or customer type (Dumas et al., 2013). A company may handle different products and services that result in different organizational behavior. Note that different product and service type does not always lead to different organizational behavior, for example, in a retail shop that sells fresh foods and dairy products may handle the purchase of both product types in an identical manner. In addition, different channel and customer type may also lead to different organizational behavior. Channel describes how a company communicates with and reaches its customer to deliver their services or products (Osterwalder, 2010). Customer type defines the different groups of people or organizations an enterprise aims to reach and serve (Osterwalder, 2010). However, it's important to note that the property is not limited to the four aforementioned properties. For example, a company may have a different process in different office location, e.g., an organization may do things differently in The Netherland than in Indonesia.

ii. Identify functions for case types

   The main organizing concept in the function-based approach is the business function, which is defined as a capability of an organization, such as 'production' or 'procurement (Dijkman et al., 2011). A business function is a collection of business behavior based on a chosen set of criteria, typically required business resources and/or competences, closely aligned to an organization, but not necessarily explicitly governed by the organization (The Open Group, 2016). In process landscape, the two rules of thumb for decomposing business function are: the functional decomposition should at least be performed down to a level at which functions correspond to different organizational units, and decomposition should include different functions for the different roles in each department (Dumas et al., 2013).

iii. Construct one or more case/function matrices.

   After identifying the case and function dimension in the two previous steps, process owner could construct a function/case matrix by putting the functions as the rows and different case type as the column. After that, the process owner needs to mark "X" in the cell where a certain function can be performed in the corresponding case type.

iv. Identify processes

   In this final step of process landscape design, process owner needs to determine which combination of case and function type that form a specific business process. The process architecture on level one put emphasize on readability. Thus it should show not much more than approximately 20 categories of business processes of an organization (Dumas et al., 2013). Figure 4.3 shows an example of a process landscape. The case property used in the example is product type and location. Within these different types of product, a certain business function can be performed. The example shows that the organization has four major processes: product development, mortgage application, mortgage payment, and mortgage collection. Yet, in this case, the product development and mortgage application handled differently in The Netherland than in Belgium. Moreover, the mortgage application activities for product type simplex and composite are different. Thus, split up results in six major processes.

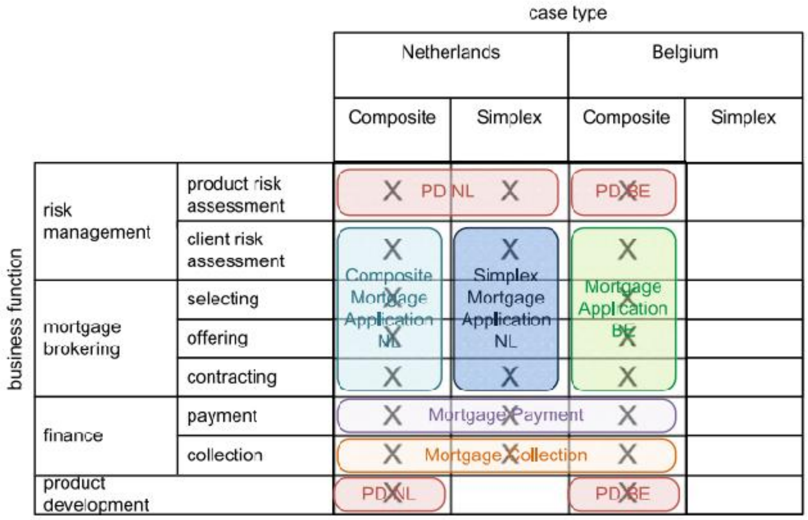


Figure 3.3 Process Landscape Example
(Source: Dijkman, n.d.)

3.2.2. Process Prioritization

Process mining assumes that an event log contains data related to a single process (Aalst, 2016). Hence, after identifying major processes within the organization, the process owner now needs to determine which process they want to analyze before acquiring the event log of the chosen process. Organizations often choose a process in which they expect a positive impact from the acquired insight. A process with a significant improvement potential is also a good starting point. The most commonly used criteria for process prioritization are the following (Dumas et al., 2013):

- Importance. This criterion is concerned about out which processes have the greatest impact on the company's strategic goals, for example considering profitability, continuity, or contribution to a public cause.
- Dysfunction. This criterion is concerned about which processes are in the deepest trouble. This can be identified, for example, by employee or customer complaints, system failures, workarounds occurrences, etc.
- Feasibility. For each process, it should be determined how susceptible they are to be analyzed using process mining. The most important is the process owner should consider the existence of event log as the prerequisite for process mining. There are cases where the event log cannot be obtained because the system used for supporting the process is not process-aware. In other cases, the event log may exist but cannot be acquired for analysis due to privacy concern.

3.3. Analysis

After the process owner decided which process they want to analyze, then the process analyst firstly needs to acquire the event log of the selected process. After that, the process analyst needs to install the tools. ProM is open-sourced and can be downloaded from www.promtools.org. ProM can load XES, MXML, and CSV files. To extract files from other data sources, ProM Import can be used (Aalst, 2016)[3]. Nonetheless, as we defined in chapter 3, the format of the event log we demonstrate in this study is in the form of XES file.

3.3.1. Process Discovery

The discovery task of control-flow perspective in process mining is referred as process discovery. A process discovery algorithm is a function that maps event log onto a process model such that the model is "representative" for the behavior seen in the event log (Aalst, 2016).

There are at least 16 process discovery algorithm plug-ins available in ProM. In our literature review, we found four of the most used plug-ins for process discovery namely alpha miner, inductive miner, fuzzy miner, and heuristic miner. The full literature table that includes the information about the algorithm used in the articles can be found in Appendix B. Fluxicon (2010) recommends fuzzy miner and heuristic miner to be used in practice for users that just getting started in using ProM. However, while both alpha miner and inductive miner can produce Petri Net as their output, fuzzy miner and heuristic miner does not. The alpha algorithm was one of the first process discovery algorithms, but it should not be seen as a very practical mining technique as it has problems with noise, infrequent or incomplete behavior, and complex routing constructs (Aalst, 2016). On the contrary, inductive mining techniques can handle infrequent behavior and deal with huge models and logs while ensuring formal correctness criteria such as the ability to rediscover the original model (Aalst, 2016). Moreover, the inductive miner is one of the few algorithms that guarantees sound process model (Buijs, n.d.). Thus, it will always produce a model that able to replay the whole event log, i.e., fitness is guaranteed (Aalst, 2016). Hence, we use **Mine Petri Net with Inductive Miner** plug-in as our process discovery algorithm in ProM. The algorithm works as follows (Buijs, n.d.):

i. Construct the directly-follows graph based on the event log.
ii. The event log is split from the most prominent operator. The split is based on the exclusive-choice, sequential execution, parallel execution and then redo-loop consecutively[4].
iii. The split is repeated in the sub-logs until a base case, the sub-log with only one activity is reached.

3.3.2. Conformance and Performance Analysis

The discovered model of a process can be compared to the behavior recorded in the event log to see the commonalities and the discrepancies between them. There are two types of a process model in regard to their purpose: normative and descriptive. When a model intended to be descriptive, then the discrepancies indicate that the model needs to be improved to capture reality better (Aalst, 2016). When the discovered model intended to be normative, the discrepancies means that there are deviations in the process execution (Aalst, 2016). For example, when an employee needs to conduct unforeseen activities whose situation is not prescribed in the existing process model; or when an

[3] available at http://www.promtools.org/promimport/

[4] see Section 3.2 about the definition of activity's execution types

employee bypasses a certain procedure and violates regulations. In this case, the process itself, not the model, need to be improved.

It is described that the process model aims to produce a process model that "representative." Representativeness of the process model can be operationalized by requiring that the model is able to replay all behavior in the log (Aalst, 2016). This is the so-called "fitness" requirement. It is the most important quality dimension of a process model (Adriansyah & Buijs, 2013). Process model conformance uses the recorded behavior to verify how well the process model conforms with the observed behavior or vice versa, and it also indicates where the actual execution differs from the process model (Adriansyah & Buijs, 2013).

Performance in a process can be seen from three dimensions: time, cost and quality. Here, we analyze the time dimension of a process. The time perspective is concerned with the timing and frequency of events (Aalst, 2016). By replaying the executed traces on the process model, timing information of the different steps in the process become available (Adriansyah & Buijs, 2013). When events bear timestamps, it is possible to discover bottlenecks, measure service levels, monitor the utilization of resources, and predict the remaining processing time of running cases (Aalst, 2016).

There are at least 10 different plug-ins for conformance or performance analysis in ProM. Two of them, the "Replay Log on Petri Net for Conformance/Performance Analysis" by Adriansyah and "Multi-perspective Process Explorer" by Mannhardt, that can explore both performance and conformance at the same time. Multi-perspective Process Explorer is a plug-in that integrates multi-perspective process mining techniques for discovery and conformance checking (Mannhardt, Leoni & Reijers, 2015). In contrast with the rather simple and static output of Adriansyah's plug-in, the MPE comes with a configuration panel that enables users to choose their measures and mode preference. Hence, we use **Multi-Perspective Process Explorer** for our performance and conformance analysis in ProM.

3.3.3. Social Network Analysis

Organizational mining focuses on the organizational perspective, i.e., the starting point for organizational mining is typically the resource attribute in the event log (Aalst, 2016). Organizational entity in the process mining could be a person, a department, a role or a system that perform the activity. In this study, we use social network analysis to present the entities relationships in the form of a graph. Social network analysis views relationships in terms of network theory consisting of "nodes" that represent individual actors within the networks, and "arcs" that represent the relationships between the actors (Kosorukoff & Passmore 2011). Arcs and nodes may have weights that indicate its importance.

There are several types of social network analysis based on how the construction of the network from the event log, that is:

- Handover of Work: describes work handovers between individuals. The more frequent individual x performed an activity that is causally followed by an activity performed by individual y, the stronger the relation between x and y is (Aalst, Reijers, & Song, 2005).

- Similar Task: detects whether or not resources perform similar tasks. If individuals work together on cases, they will have a stronger relation than individuals rarely working together (Aalst, Reijers, & Song, 2005).
- Subcontracting: detects if one executor subcontracts his work to another (Chamorro & Maturana, 2017). The main idea is to count the number of times an individual x executed an activity in-between two activities executed by individual y (Aalst, Reijers, & Song, 2005).
- Working Together: count how many times two resources have worked on the same case (Aalst, 2016).
- Reassignment: determines activities have been reassigned (Chamorro & Maturana, 2017).

Our scoping study shows that the most used metrics is the handover-of-work (Ferreira & Alves, 2012; Mans et al., 2012; Mans et al., 2008; Aalst, 2008; Wang, 2016; Chamorro & Mataruna, 2018). Hence, we use the **Mine for a Handover-of-Work Social Network** to mine the organizational perspective in ProM.

3.3.4. Decision Mining

Decision mining discovers why particular cases take a particular path (De Leoni & Aalst, 2013). In order to analyze the choices in a business process, we first need to identify those parts of the model where the process is split into alternative branches, also called decision points (Rozinat & Aalst, 2008). In Petri Net, the decision point is located in a place with more than one outgoing arcs. After identifying a decision point, the influence of case data to the decision is evaluated, i.e., whether cases with certain properties typically follow a specific route. The idea is to convert every decision point into a classification problem where the classes are the different decisions that can be made (Rozinat & Aalst, 2008).

A classification technique like decision tree learning can be used to find decision rules (Aalst, 2016). In classification technique, one column corresponds to the target attribute that we try to predict (Rokach & Lior, 2014). The predicted column is called a response variable, and the target attributes are called predictor attributes. There are three main types of predictor attributes (Rokach & Lior, 2014):

- Numeric: describes the quantitative value of the attribute, e.g., the number of order in a particular case.
- Ordinal (or categorical): provides the ordering of the category but without the exact measurement of the distance between them. For example, the three classes of socioeconomic status: poor, middle class, and rich.
- Nominal: the values are merely distinct names or labels with no meaningful order by which one can sort the data. For example, the country origin of customers.

In contrast with control-flow perspective who occupies a lion's share of process mining research focuses, there is very limited support and less attention for the case perspective of the process (De Leoni & Aalst, 2013; Rozinat & Aalst, 2008). The first plug-in in ProM that supports case perspective is Decision Miner implemented in 2006 (Rozinat & Aalst, 2006). However, it cannot deal with event logs with deviating behavior and more complex control-flow constructs (De Leoni & Aalst, 2013). A more recent plug-in named "Discovery of the Process Data-Flow (Decision-Tree Miner)" was developed with the intention to overcome the limitation of Decision Miner plug-in. It can discover accurate data flow even in the presence of event logs with non-conforming traces (De Leoni & Aalst, 2013). Hence, we use **Decision-Tree Miner** for decision mining of the case perspective in ProM.

4.3 Post-Analysis

After the analysis, the process owner summarizes the results, and this gives input for further process improvement (Buijs, n.d.). Process mining can result in one or more of the following improvement actions such as (Aalst, 2016):

- Redesign: insights attained in the analysis phase can prompt changes to the process. For example, if the analysis found that the performance of the business process is poor, then it indicates that the process needs to be redesigned.
- Adjust: process mining can result in (temporary) adjustments. Here, the process is not redesigned, only predefined controls are used to adapt or reconfigure the process. For example, a company might define in their contingency plan that if a certain criterion is met, then they will allocate certain resources to accommodate the cases.
- Intervene: process mining may also reveal problems related to particular cases or resources. For example, process owner can decide to terminate a certain problematic case or take disciplinary action for the employee that violates compliance regulations.
- Support: Based on historical information, a process mining tool can predict the remaining flow time or recommend the action with the lowest expected costs.

3.4. Chapter Summary

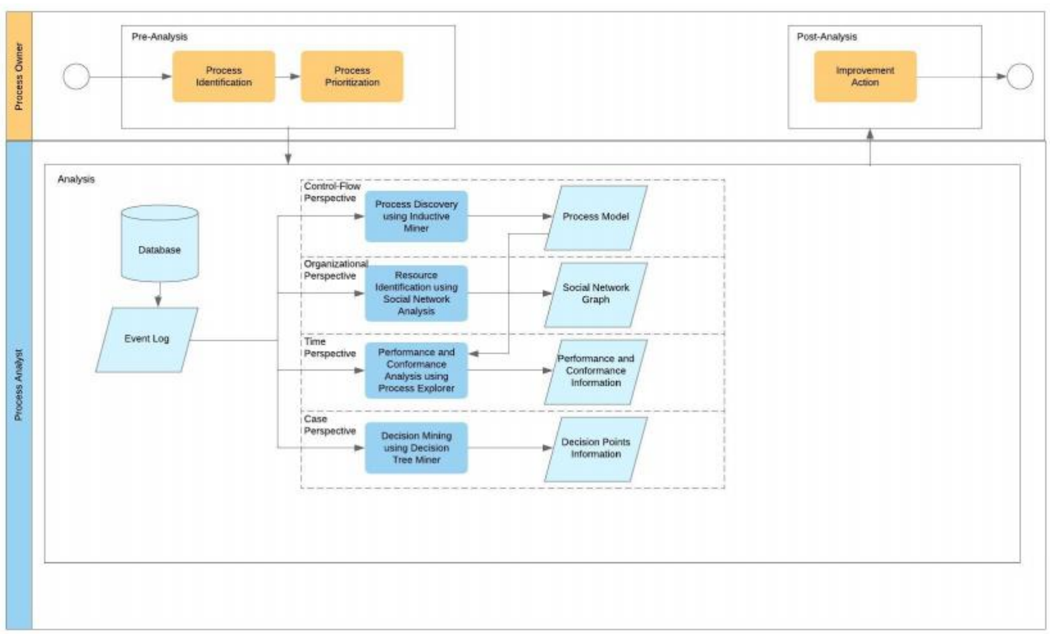


Figure 3.4 Enhancement Approach 1st Version

The analysis phase conducted by process analysts. The control-flow perspective is discovered using inductive miner. The organizational perspective is discovered using social network analysis. The timing and frequency of the event then replayed on top of the model to identify the conformance and performance of the process. The case perspective about the decision points in the process is identified using decision-tree miner. Figure 3.4 illustrates the initial version of the proposed approach. In the next chapter, we will demonstrate the analysis phase implementation in ProM.

4. Demonstration

The chapter starts with an overview of ProM tools. It then demonstrates the application of the proposed approach in the real-life data set. The first case is about insurance claim via telephone. The second case is about building permit application in Dutch's Municipality. The last part of this chapter summarizes the demonstration phase of the research.

4.1. ProM Tools Overview

The requirement of the analysis phase is the installation of the process mining tools. We suggest the installation ProM 6 as the tools. Figure 5.1 shows the page that will appear when users open ProM. The import button is located at the upper right of the page. There are several import options. The naive one uses the most memory, but it's quite fast, lightweight; sequential is also quite fast but has some limitations; and then disk-buffered by MapDB puts as little information in the memory as possible and therefore is the best choice when you have extremely large event logs (Buijs, n.d.). Naïve is the default option for log import.

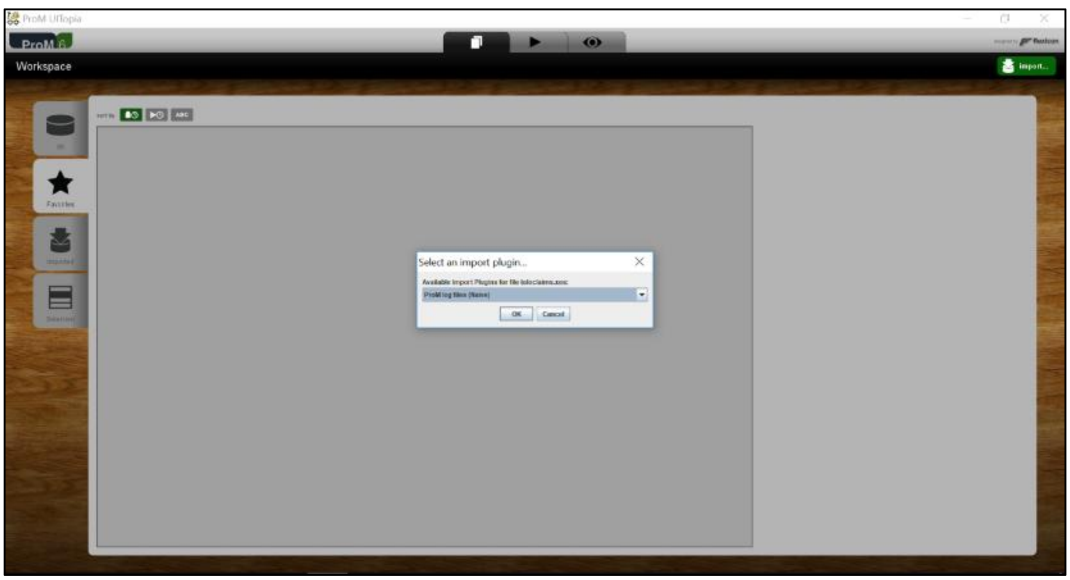


Figure 4.1 Import Event Log in ProM

Figure 5.2 below displays the 'view' tab in ProM. The tab brings the user to view the information about the imported event log. The dashboard displays general information such as the number of cases, events, event classes, and originators. It also displays the minimum, maximum and average number of event per case. In the inspector, the users can view the detailed ordering of events per case. Log summary lists the available event classes, the starting events, the end events, and the resources of the process. Also, in this tab, users can see the dotted chart visualization of the event log. The dotted chart can be seen as an example of a visual analytics technique, where users visually identify patterns and trends in large datasets (Aalst, 2016).

Figure 4.2 Log Info Visualization in ProM

The "action" button is located in the upper middle part of the display with the symbol ►. Figure 5.3 shows the selection of inductive miner plug-in in ProM by typing the plug-ins name on the actions search bar. Event classifier, an inductive miner variant, and a noise threshold need to be configured to use this plug-in. If the noise threshold is set to 0.00 perfect log fitness is guaranteed, meaning all behavior observed in the event log will be present in the process model (Karwehl, 2018). The rule of thumb is to set the noise threshold at 0.20.

Figure 4.3 Inductive Miner in ProM

Multi-perspective Process Explorer is a plug-in that integrates multi-perspective process mining techniques for discovery and conformance checking (Mannhardt, Leoni & Reijers, 2015). The input for the multi-perspective explorer is the imported event log and the discovered Petri Net. Figure 5.4 shows the selection of the plug-in.

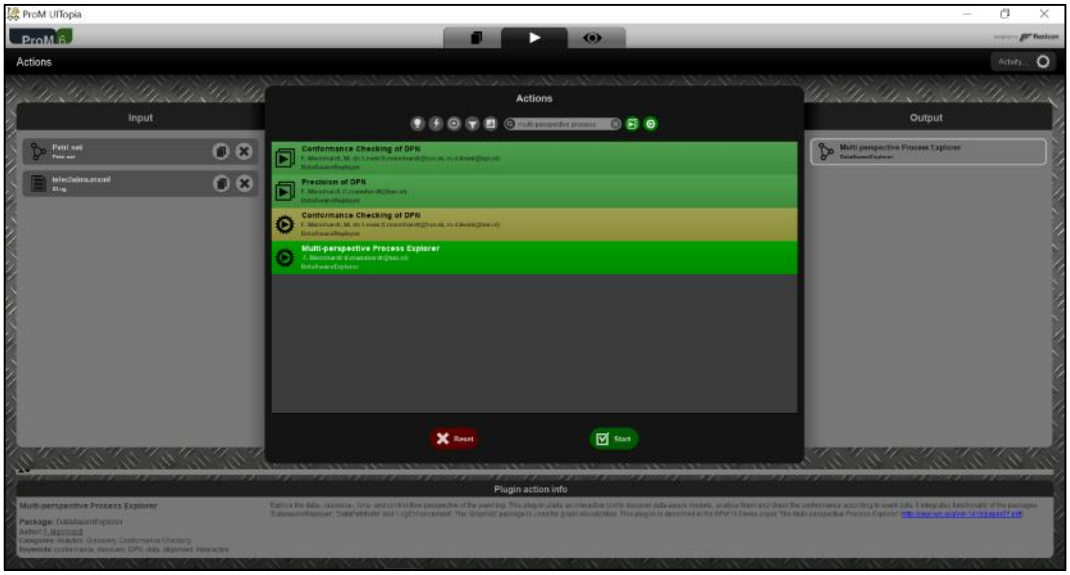


Figure 4.4 Multi-Perspective Process Explorer in ProM

As explained previously, fitness requires the model is able to replay all behavior in the log. After evaluating whether the discovered model is a suitable representation of the process behavior using fitness mode in the MPE plug-in, performance information about the time perspective such as average waiting times can be obtained using the performance mode (Mannhardt, Leoni & Reijers, 2015).

Figure 5.6 shows a series of social network analysis plug-ins. The metric options are a handover-of-work social network, a reassignment social network, similar-task social network, subcontracting social network and working-together social network. Here, we will use the handover-of-work metrics.

Figure 4.5 Social Network Analysis in ProM

Figure 5.7 shows the decision tree miner plug-in in ProM. The decision tree miner plug-in in ProM determines the decision points in a Petri Nets model and deals with hidden and duplicate activities (Kebede, 2015). The input for the decision tree miner is the imported event log and the discovered Petri Net.

Figure 4.6 Decision Tree Miner in ProM

4.2. Case 1: Teleclaim

The event log named teleclaim.xes records the handling of different types of insurance claim over the phone. The call centers are located in Brisbane and Sydney. Table 5.1 below summarizes the properties of the event log. In this section, we do not show the step-by-step implementation in ProM. Instead, it can be found in Appendix D.

Table 4.1 Event Log Properties of Teleclaim

Aspect	***Value***
Cases	3,512
Events	46,138
Event Classes	11
Average Event per Case	13
Maximum Event per Case	17
Minimum Event per Case	5
Originators	4
Start Date	Jan 1, 1970 1:00:00 AM
End Date	Jan 1, 1970 1:09:18 PM

4.2.1. Process Discovery

We discover the process model using the Inductive Miner plug-in in ProM. Figure 5.7 shows the discovered process model of the teleclaim.xes event log. As shown, the discovered model is concise and can be interpreted easily. We can see that the claim process started by the customer call to Brisbane or Sydney call center. Then, each call center checks whether the information is sufficient. If it is, then the claim is registered. After the assessment, the claim will be either paid or reimbursed. However, if the assessment determines that the claim is not eligible for payment or reimbursement then the case will be ended, i.e., the claim is rejected.

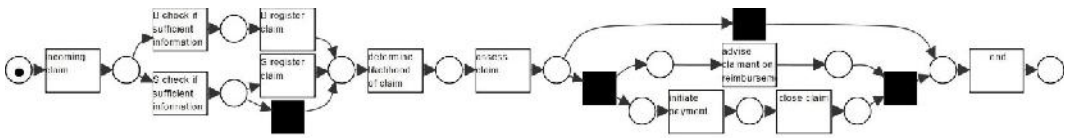


Figure 4.7 Process Model of Teleclaim

4.2.2. Conformance and Performance Analysis

Next, we use the multi-perspective process explorer plug-in to view the fitness information. The average fitness of the discovered process model is 60.9%. Figure 5.8 depicts the fitness view of the discovered Petri Net. From the graph, we can see that more than half of the cases are reimbursed or paid, while the rest rejected.

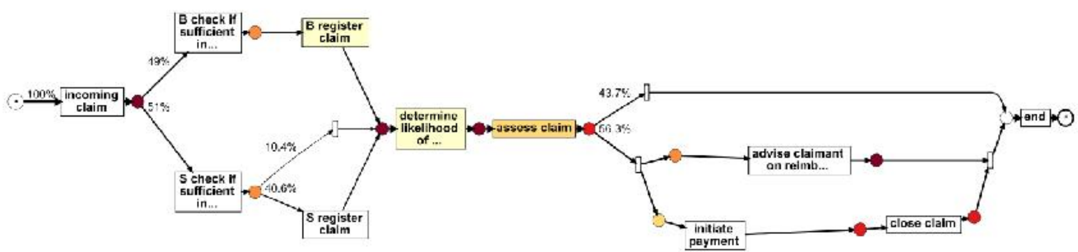


Figure 4.8 Fitness of Teleclaim Process Model

For performance analysis, since only timing information of traces that can actually be replayed on the process model can be taken into account, the fitness of the process model with respect to the event log needs to be high (Adriansyah & Buijs, 2013). Adriansyah & Buijs (2013) suggest that the fitness of the model to the log shoul be at least 99% to have a reliable performance result.

In the previous step, we find that the average fitness of the discovered model is not high. Therefore, we utilize our secondary artifact, the literature graph to solve this problem. We follow the suitable branch of the graph in Appendix B, starting from following the branch XES format file of the event log, then the procedural modeling language, and after that the Petri Net notation. We explore the existing approaches with the objective of improving the discovered process model. For the purpose of performance analysis, we found that the article numbered [1] is the suitable approach. They suggest that before calculating the performance, **if the fitness is below 99%, then the event log needs to be aligned with the model**. This conditional requirement will be incorporated into our artifact design.

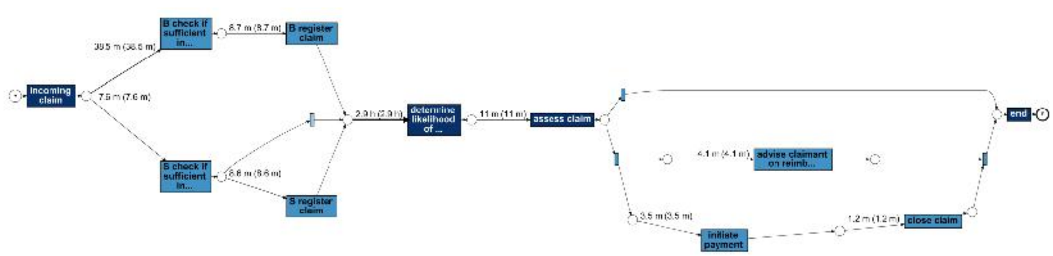


Figure 4.9 Performance of Teleclaim Process

Users can use the "Align Log to The Model" plug-in to increase the fitness of the model to the log. After the alignment, the conformance between the log and the model is 100%. Figure 5.9 shows the performance analysis based on the aligned event log. The darker blue color indicates the longer average time of the activity. As shown, the activity "determine the likelihood of claim" take the longest average time with 2.9 hours per case.

4.2.3. Social Network Analysis

Figure 5.10 shows the social network graph of teleclaim.xes based on the handover-of-work metrics. The graph provides visual imagery behind the basic concept of relations between the actors (Scott, 2012). The size of the nodes indicates the weight of the nodes. The weight of the nodes is written in numbers. It can be seen that the customer can contact the call center in Sydney or Brisbane. After that, both branches direct the claim to the claim handler.

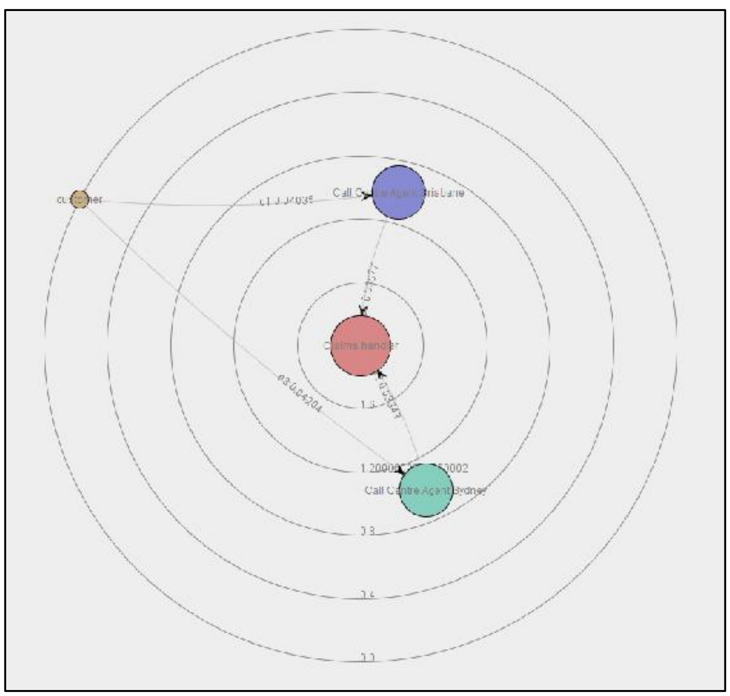


Figure 4.10 Social Network Analysis of Teleclaim Process

Through process discovery, the activities and their execution ordering are identified. Social Network Analysis provides the visualization of relationships between resources. However, these two steps do not relate the resource attribute to the event they perform.

Therefore, as an additional step **to provides more insight into the organizational perspectives, the dotted chart of events to the resources is created**. This additional step will be incorporated into our artifact design. Below is the dotted chart for Teleclaim process.

Figure 4.11 Resource and Event Dotted Chart of Teleclaim

4.2.4. Decision Mining

Figure 5.13 shows the Data Petri Net discovered by using Decision Tree Miner plug-in in ProM. In decision mining, since the data properties of the case is very important, it is prudent to check the data types of the attributes in the log. Therefore, **before using the decision-tree miner, the event log needs to be repaired.** This additional requirement will be incorporated into our artifact design.

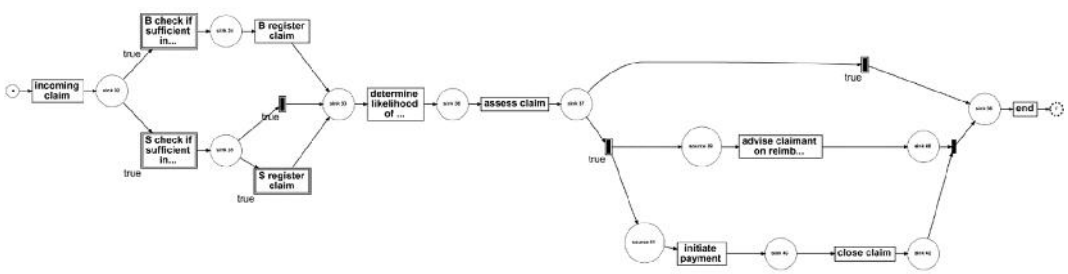


Figure 4.12 Decision Tree of Teleclaim

In ProM, users can use "Repair Type of Event Attribute in the Log" plug-in for the purpose. The table below shows the output of the decision-tree miner. The plug-in discovers guards for transitions at decision, instead of decision logic for an entire decision point (Massimiliano, 2015; Peeters, 2016). However, the decision logic that is extracted by the decision-tree miner returned in very unclear formats thus significantly reducing the utility of these plug-ins for daily practice (Peeters, 2016). In fact, many of the following algorithms either directly use or are based upon open source algorithms like those found in the Weka library.

4.3. Case 2: Dutch Municipality

Here, we evaluate our guideline by demonstrating its implementation using an event log made available by 3TU data repository[5]. In this section, we do not show the step-by-step implementation in ProM. Instead, it can be found in Appendix E. Table 4.3 shows the information about the event log named Municipality_1.xes. The event log we use is from Dutch municipalities building permit application.

In order to build, rebuild or renovate in The Netherland, in most cases, one will need a building permit. There are two ways to apply for the permit, online application or by sending a printed form via mail. Most municipalities require planned construction to be checked for aesthetic value by a commission which controls compliance with regulations regarding the external appearance of a building (VROM, 2018). These regulations are laid down on the local level and differ from municipality to municipality.

4.3.1. Process Discovery

Just like the previous case, we discover the process model of the event log using the Inductive Miner in ProM. Figure 5.13 depicts the Dutch Municipality process model. As shown, the process model is very complex and hard to understand at a glance.

Table 4.2 Example Event Log Properties

Aspect	Value
Cases	832
Events	44.354
Event Classes	410
Average Event per Case	53
Maximum Event per Case	132
Minimum Event per Case	1
Originators	11
Start Date	Tue, June 29th, 2010
End Date	Wed, March 4th, 2015

A process model can be a “lasagna process” if with limited efforts it is possible to create an agreed-upon process model that has a clear structure and most cases are handled in a prearranged manner (Aalst, 2016). The process model discovered in Case 1 is a lasagna process. However, in practice, a process model can be spaghetti-like, unstructured and hard to understand. Service-related processes are typically spaghetti-like (Aalst, 2016). The discovered process model of the Dutch Municipality business process is an example of the spaghetti process model.

[5] available at http://data.3tu.nl/repository/collection:event_logs/

In a process that is complex by design and nature, there will be a dilemma of either losing detail and not see what is really going on or getting a very complex process model (Ham, 2015). In the spaghetti process model, even though it is hard to identify the exact ordering of the process at a glance, we still can see the pattern of the process. In the case of Dutch Municipality, we can see that there are several initial activities took places before the variation of sequences represented by the big XOR and XOR Loop (See the red circle in the Figure 5.12). After further examine the procedure of building permit application, this variation seems natural since there is a lot of different permits that one can apply for. Even some of the initiation activities of the process relatively the same, but further in the process, different permit requires different procedures and activities.

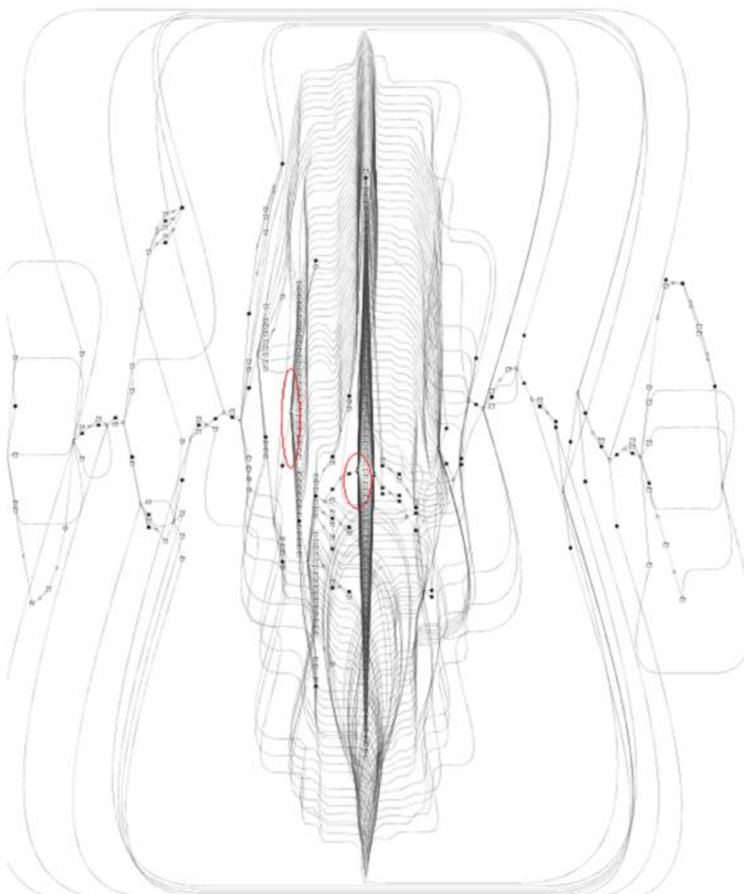


Figure 4.13 Dutch Municipality Process Model

4.3.2. Conformance and Performance Analysis

The fitness mode of the Multi-Perspective Process explorer shows that the average fitness of the discovered model to the log is 94.6%. Here, we suggest for process analyst to look into the problematic activity, which is marked by darker yellow to red color, instead of the overall Petri Net. Figure 5.14 shows some activity in the Dutch Municipality Building Permit Process that has a lower fitness to log.

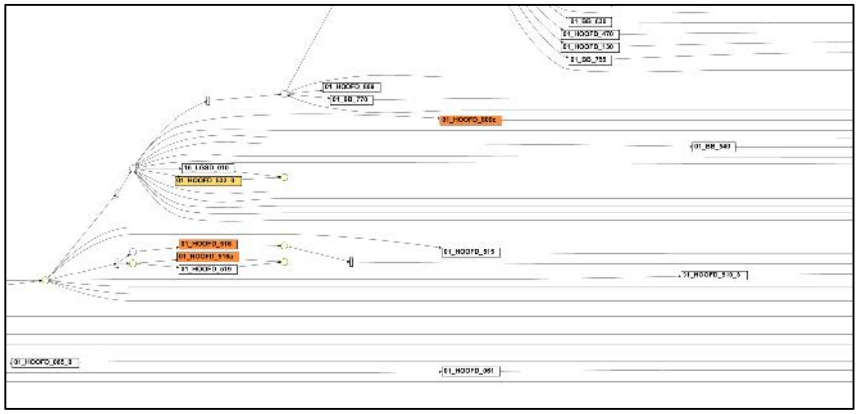


Figure 4.14 Fitness of Dutch Municipality

Next, for performance analysis, since the fitness of the model to the log is relatively high, we will skip the alignment of the log to the model. Just like the previous steps, for performance analysis we also suggest process analyst look into specific problematic activity with poor performance, indicated with darker blue color. In Figure 5.15 below we can see that the time needed to perform the activities in the process varies from as short as in milliseconds, to several days. Another, interesting finding from the performance graph is the most frequent edge that the cases went through (see the thicker dark blue line in the graph) is a loop. This could be interpreted that a large number of activities are not executed sequentially.

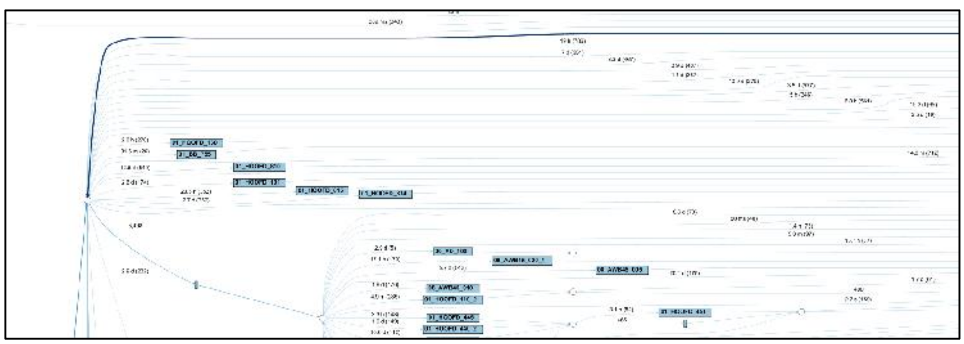


Figure 4.15 Performance of Dutch Municipality

4.3.3. Social Network Analysis

Figure 5.16 below is the social network graph of the 11 resources of the Dutch Municipality case study. From the visualization, we can see that the resource with id 560532 plays a central role in the process, indicated with the many edges directed to the node and the size of the node.

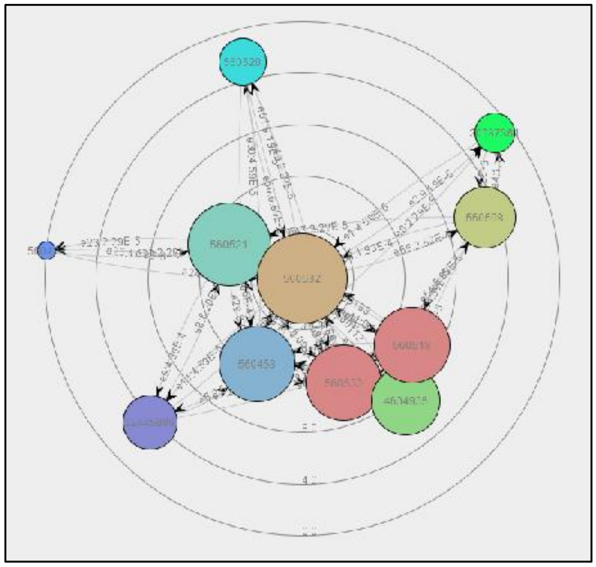
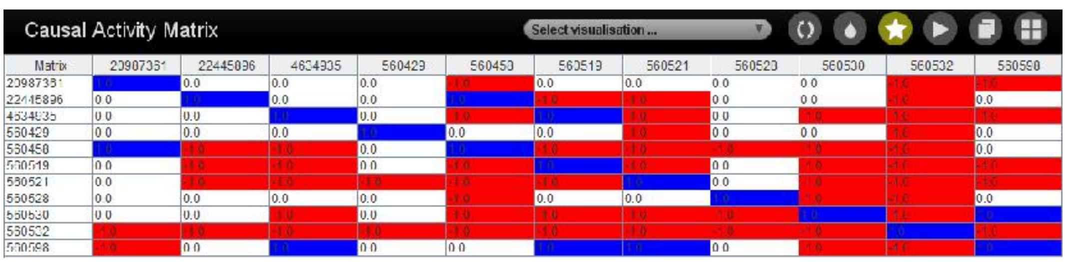


Figure 4.16 Social Network of Dutch Municipality Resources

The problem with the social network graph is that it illustrates only the patterns of connections and not any information about the causality between the resources. When there are a lot of resources within the process, the readability of the graph might be reduced. In that case, the **Discovery of Resources Causal Matrix** can be utilized to look at the relationships between resources. This conditional step will be incorporated into our artifact design.



Causal Activity Matrix

Matrix	20907361	22445896	4634935	560429	560450	560519	560521	560528	560530	560532	560598
20987381	1.0	0.0	0.0	0.0	-1.0	0.0	0.0	0.0	0.0	-1.0	-1.0
22445896	0.0	1.0	0.0	0.0	1.0	-1.0	-1.0	0.0	0.0	-1.0	0.0
4634935	0.0	0.0	1.0	0.0	-1.0	1.0	-1.0	0.0	-1.0	-1.0	-1.0
560429	0.0	0.0	0.0	1.0	0.0	0.0	-1.0	0.0	0.0	-1.0	0.0
560450	1.0	-1.0	-1.0	0.0	1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0
560519	0.0	-1.0	-1.0	0.0	-1.0	1.0	-1.0	0.0	-1.0	-1.0	-1.0
560521	0.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.0	0.0	-1.0	-1.0	-1.0
560528	0.0	0.0	0.0	0.0	-1.0	0.0	0.0	1.0	-1.0	-1.0	0.0
560530	0.0	0.0	-1.0	0.0	-1.0	-1.0	-1.0	-1.0	1.0	-1.0	1.0
560532	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.0	-1.0
560598	-1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	-1.0	-1.0	1.0

Figure 4.17 Causality Matrix of Dutch Municipality Resources

4.3.4. Decision Mining

Oftentimes, if there are formatting errors in the data, then the decision-tree miner will end up running without ends. Unfortunately, this is what happened when we tried to mine the decision tree based on imported the event log for the Dutch Municipality event log. The usability of the plug-in is evaluated further in the evaluation phase.

4.4. Chapter Summary

The demonstration is performed using two different data sets. The Teleclaim.xes is a lasagna process with a clear process model. In contrast, Municipality_1.xes produces a spaghetti-like process model. In such cases, it's recommended to focus on the activity with low performance, e.g., the activity that has a bottleneck or took a longer time to be executed. The approach is presented in the form of white paper[6].

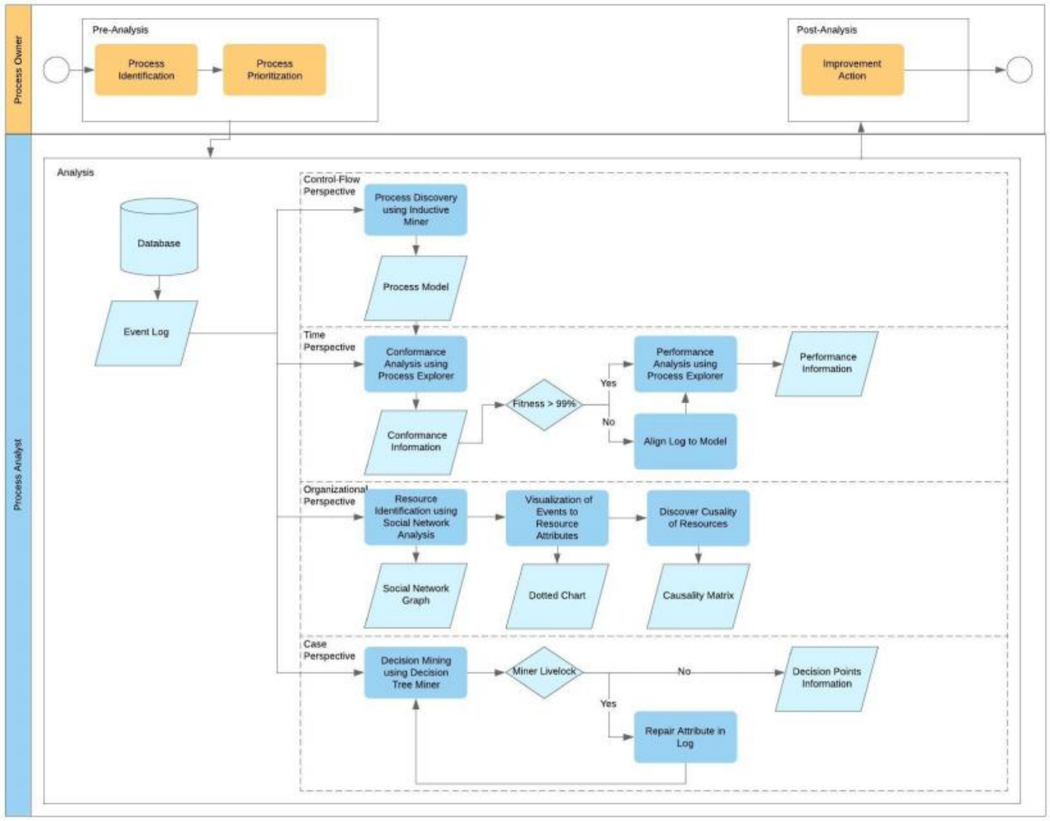


Figure 4.18 Enhancement Approach 2nd Version

The demonstration is also a way to evaluate the applicability of our artifact. The method used in the demonstration is the Case Mechanism Experiment, a test in which the researcher applies stimuli to a model and explains the response in terms of mechanisms internal to the model (Wieringa, 2014). Findings in the demonstration are the input to improve the proposed approach. The figure above is the revision of the artifact after the demonstration phase. This version will be used for the evaluation phase.

[6] See the author ResearchGate page via https://www.researchgate.net/project/Enhancement-in-Process-Mining-A-Literature-Review

5. Evaluation

This chapter is about the evaluation of the proposed approach. It starts with the evaluation of the artifact with an experimental evaluation approach named Technical Action Research. Based on the experiments then the ProM tools user acceptance is evaluated. The last part of this chapter summarizes the evaluation phase of the research

5.1. Technical Action Research

Technical action research is the use of an experimental artifact to help a client and to learn about its effects in practice (Wieringa, 2014). For this evaluation, we created a white paper that contains a guideline for the analysis phase. The paper is presented to two students at the University of Twente who are novice process mining users. After that, we provide them with a set of process analysis tasks. The aim of the task is to see whether a novice process mining user able to extracts valuable information using the provided guideline.

In parallel, another two students who have more experience in using process mining tools, and has completed the process mining course, is also given the same task. The difference is that the latter group perform the task without the guideline. The result of the process analysis between the two groups is compared.

5.1.1. Log Overview

The event log consists of 100 process instance and 3730 events. There are 37 events per case on average and are executed by 11 different actors. The log is about research papers reviewing process that starts when the paper submission sent to three different reviewers. After that, the reviewers are invited to give a feedback report. However, there are occasions where the reviewers do not respond to the invitation. Consequently, the decision to accept or reject the paper cannot be made in the first round of reviews. Then, additional reviewers are invited. This process is repeated until a final decision can be made. Figure 5.1 depicts the process model of the review process.

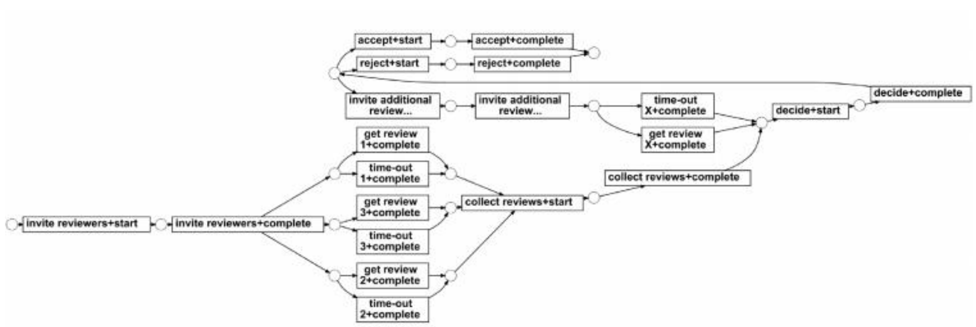


Figure 5.1 Process Model of Reviewing

When the event log is replayed on top of the discovered process model, we can see how long each activity takes on average. As depicted in Figure 5.2, there's a bottleneck in activity "decides." The activity also took the longest time with an average of 114.3 days.

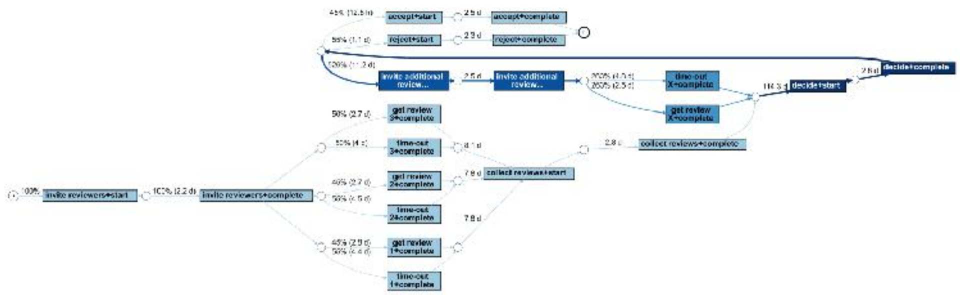


Figure 5.2 Performance of the Review Process

The possible explanation for the bottleneck is because the "decide" activity only relies on one resource. Figure 5.3 shows the dotted chart listing the activity performed by each resource. The figure shows that the "decide" activity performed by a resource named "Will."

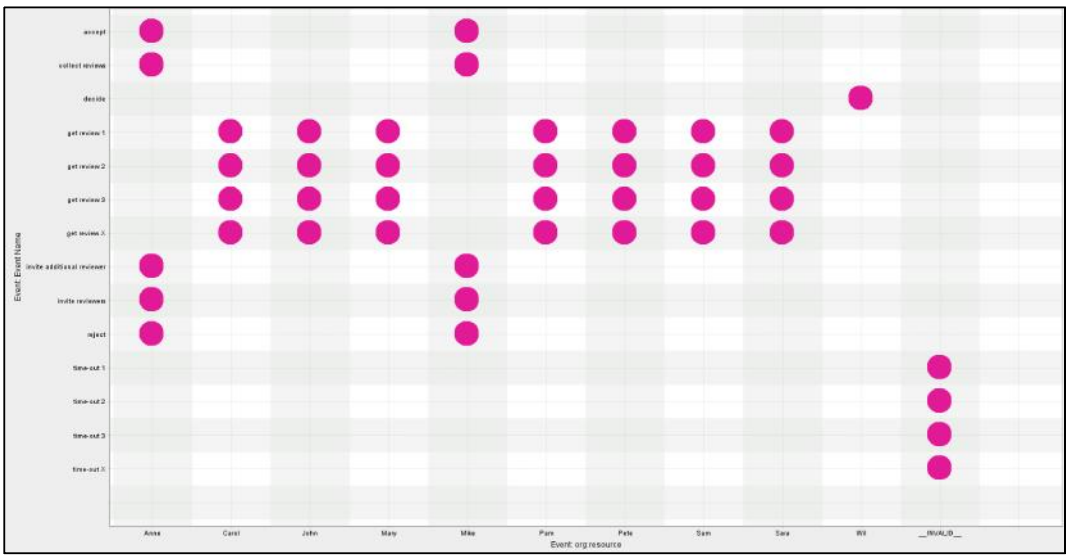


Figure 5.3 Dotted Chart in Review

5.1.2. Evaluation Results

Given the event log, all respondents were asked to answer some questions. The questions can be found in Appendix F. Each question related to a certain process analysis tasks. They completed all the task on average two and a half hours. Below is the list of the task:

- Task 1: Import and inspect the event log.
- Task 2: Explains the business process.
- Task 3: Discover the conformance of the model to the log.
- Task 4: Suggest the activity that needs to be concerned.
- Task 5: Identify the most important resource.
- Task 6: Name the activity performed by the most important resource.
- Task 7: List the resource related to the most important resource.
- Task 8: Find the guard points in the process
- Task 9: Provide process improvement recommendation

In the first task, the respondents were asked about the log properties. All respondents correctly mentioned the number of cases, events, and originators in the process. They also correctly name the minimum, maximum and average number of instance per cases.

In the second task, the respondents were asked to explain briefly in words about the process. All four respondents were able to complete this task by discovering the process model of the process. Respondent 1 and Respondent 2 uses Inductive Miner plug-in as prescribed by the guideline. Respondent 3 also uses the Inductive Miner. When asked why, she/he said that based on her/his experiences, the plug-in delivers a better outcome in comparison to other algorithms. Meanwhile, Respondent 4 uses Alpha Miner plug-in. When asked why, the respondent said that it's because it's most used and popular plug-in process mining. Alpha Miner is popular in process mining practices because it's the first algorithm developed to discover process model. However, as explained in section 3.3.1, the alpha algorithm should not be seen as a very practical mining technique as it has problems with noise, infrequent or incomplete behavior, and complex routing constructs (Aalst, 2016).

In the third task, the respondents were asked to discover the conformance of the model to the log. Respondent 1 and Respondent 2 were able to check the conformance of the model to the log using the Process Explorer as proposed in the guideline. Respondent 3 and 4 also used this plug-in.

In the fourth task, the respondents were asked to recommend an activity that needs to be concerned by a process owner. As depicted in Figure 5.2, there's a bottleneck in "decide" activity, and it took an average 114.3 days to complete. All respondents were able to identify this problem by performing performance analysis using "Multi-Perspective Process Explorer" plug-in in ProM.

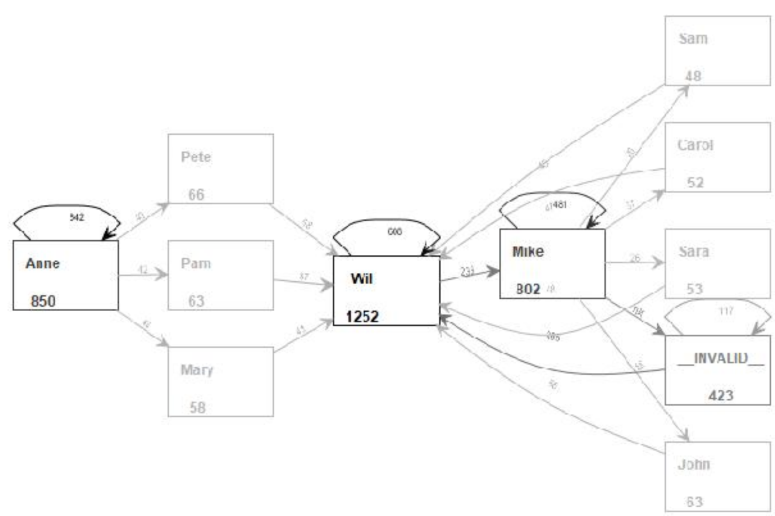


Figure 5.4 Heuristic Net of Reviewing Process Resources

In the fifth task, the respondents were asked to identify the most important resource in the process. There are several ways to identify who is the most important resource. In the proposed guideline, social network analysis and causality matrix between resources are suggested. All four respondent can identify the most important resource. However, Respondent 3 use a rather interesting approach. She/he explains that her/his experience in using process mining mostly focuses on control-flow perspective. The respondent said that she/he never utilizes the social network analysis plug-in and does not understand how to interpret the results. Instead, she/he uses the "Heuristic Miner" plug-in with the resource attribute as the classifier. Figure 5.5 shows the result of the plug-in. It shows that a resource named "Will" has the most frequency of activity.

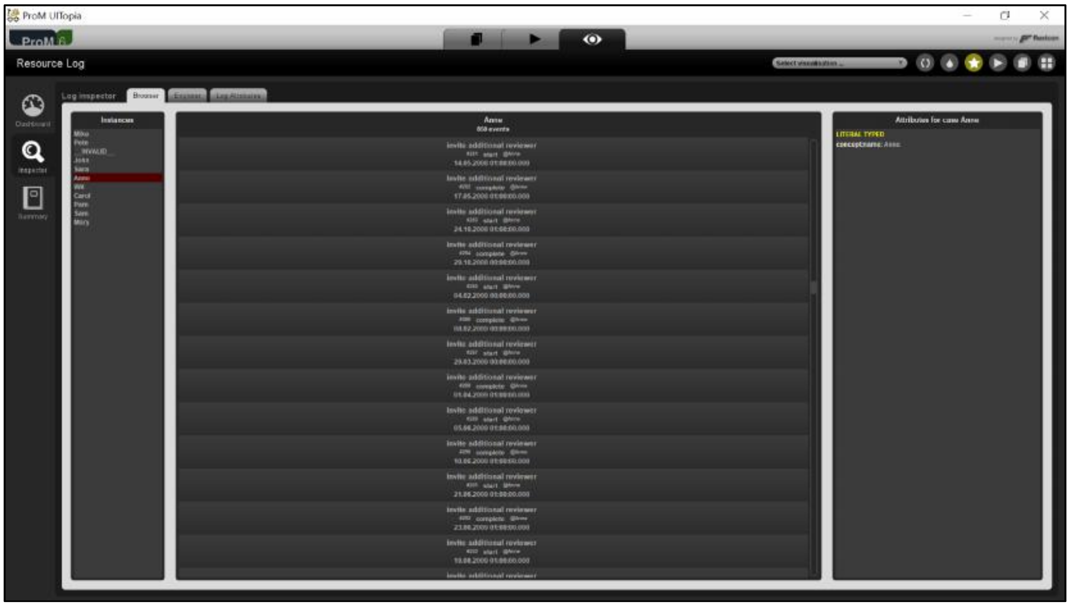


Figure 5.5 Result of Produce Log from Resource of Reviewing Process

In the sixth task, the respondents were asked to name the activity performed by the most important resource. In the proposed guideline, the dotted chart visualization of resource to its activity can be used to perform this task. All four respondents are able to name the "decide" activity as the activity performed by "Will," the most important resource. However, Respondent 3 use a rather interesting approach. Because in Task 4 the respondent found that the bottleneck is in "Decide" activity, and in Task 5 the respondent found that "Will" has the highest frequency of activity, she/he then assumes that resource named "Will" performs the "Decide" activity. Therefore, the identification of the most important resource is based on assumption rather than a conclusive result. This approach is unreliable because it is not applicable in a case where are there are multiple bottlenecks points performed by multiple resources. Meanwhile, Respondent 4 uses "Produce Log from Resource" plug-in. The plug-in lists all activity performed by a selected resource in all cases. This approach is inefficient because it produces redundant information. Moreover, to identify all activity performed by all resources using this approach means that the log needs to be examined manually.

In the seventh task, the respondents were asked to list the resource related to the most important resource. Respondent 1, Respondent 3 and Respondent 4 are able to answer the question related to this task. However, Respondent 2 did not answer the question because she/he uses the Lite version of Prom 6 and it does not support the causality matrix discovery plug-in. ProM Lite uses the user interface of ProM 6 versions but only contains the most typical plugins (Process Mining Group, 2016).

In the eighth task, the respondents were asked to find the guard points in the process, i.e., the activity where determine the path taken by each process instances. Based on the proposed guideline, the "Decision-Tree Miner" can be used. However, in chapter 4, the plug-ins demonstrate some usability problem. The problem is also found in this evaluation phase. Nonetheless, all four respondents were able to identify the decision points by looking at splits in the control-flow process model.

Table 5.1 Respondent's Task Completion

Task	Respondent 1	Respondent 2	Respondent 3	Respondent 4
Task 1	✓	✓	✓	✓
Task 2	✓	✓	✓	✗
Task 3	✓	✓	✓	✓
Task 4	✓	✓	✓	✓
Task 5	✓	✓	✓	✓
Task 6	✓	✓	✗	✗
Task 7	✓	✗	✓	✓
Task 8	✗	✗	✗	✗
Task 9	✓	✓	✓	✓

Table 5.1 summarizes the task completion of each respondent. If the respondent does not answer the question related to the task, or they answer with an inefficient and unreliable approach, then the completion will be marked with (✗). As shown in Table 5.1, Respondent 1 and Respondent 2, the two students who have no prior knowledge about ProM were able to perform process analysis in general. The exception is Task 7 by Respondent because of the unavailability of the plug-in.

After they finished the evaluation task, we asked their opinion about the guideline. Both respondents stated that they think that the proposed guideline has helped them in understanding about process mining. They also described that the guidelines and the exercise intrigue their curiosity to learn more about process mining.

Another interesting finding in this evaluation phase is that, even though Respondent 3 and Respondent 4 has more experiences in using process mining tools, they unable to discovers organizational perspectives efficiently. They both states that this might be due to the focus on control-flow perspectives in current process mining techniques. Another possible explanation about the use of inefficient approach by the more experienced users is because of the "Paradox of Active User" effect. Fu and Gray (2004) explain that for people who expect an outcome of a system, have the intention to adopt the technology for its offered benefits, but it decreases their motivation to spend the time just learning about it. This bias can lead to the use of "preferred procedure" instead of a more efficient procedure. A preferred procedure is typically a well-practiced, generic procedure that is applicable either within the same task environment in different contexts or across different task environments (Fu & Gray, 2004). The approach of Respondent 4 to solve Task 6 is the example of the use of the preferred procedure instead of an efficient procedure. When we asked why, she/he claimed that even though she/he knows that there are other plug-ins to perform the task such as Social Network Analysis, she/he prefers to use a more the less efficient but familiar approach instead of learning a new one.

5.2. User Acceptance Test

After the evaluation through the experiment, the two respondents who are novice process mining tools were asked about their opinion regarding ProM. We use the Technology Acceptance by Davis et al. (1989). The model describes that for potential users, perceived usefulness and perceived ease of use are the determining variables that can affect the user's intention to adopt the technology. Perceived usefulness is the degree to which a person believes that using a particular system would enhance his or her job performance. Perceived ease of use is the degree to which a person believes that using a particular system would be free of effort (Davis, 1989). After years of research and a large multitude of studies investigating TAM, it is believed that that perceived usefulness is a very influential belief and that perceived ease of use is an antecedent of perceived usefulness, and an important determinant of use (Benbasat & Barki, 2007).

Table 5.2 User Acceptance Test toward ProM

Construct	Respondent 1	Respondent 2	Average
Perceived Usefulness 1	1	4	3
Perceived Usefulness 2	4	4	4
Perceived Usefulness 3	4	4	4
Perceived Usefulness 4	3	4	3.5
Ease of Use 1	4	4	4
Ease of Use 2	4	4	4
Ease of Use 3	3	4	3.5
Intention to Use 1	5	4	4.5
Intention to Use 2	1	4	2
Intention to Use 3	1	4	2

The respondents filled out a questionnaire about the ProM tools. The questionnaire uses a five-point Likert scale. The scale is sorted from biggest number to lowest in this order: "strongly agree," "agree," "neutral," "disagree," and "strongly disagree."

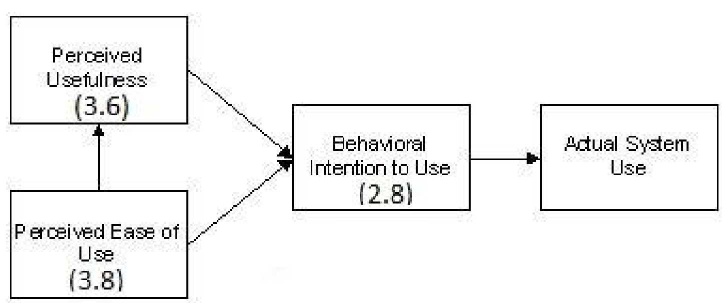


Figure 5.6 ProM User Acceptance Result

First, for perceived usefulness, we asked whether using ProM enables respondents to know about the control-flow, organizational, time perspectives and case perspectives respectively. Then, in the perceived ease of use, we asked whether the respondent's experience using ProM during TAR was easy. After that, we asked about the respondent's perceived ease of use in learning and using ProM in the future. Lastly, both respondents express that they have a rather high intention to use ProM but do not predict nor plan to use it in the near future. Figure 5.6 depicts the result of user acceptance test. The number below each construct represents the average of the result.

5.3. Chapter Summary

We evaluate the using three approaches namely Technical Action Research (TAR) and User Acceptance Test (UAT). The TAR shows that using the guideline, the novice process mining users are able to do a process analysis of an example event log. The UAT shows that after their interaction with ProM in TAR experiment, the novice process mining users have an average score of 3.6 out of 5 for perceived usefulness, and 3.6 out of 5 for perceived ease of use. In the intention to use, the respondent has the intention to use ProM but does not predict nor plan to use it in the near future. This results in an average score of 2.8 out of 5 for the intention to adopt.

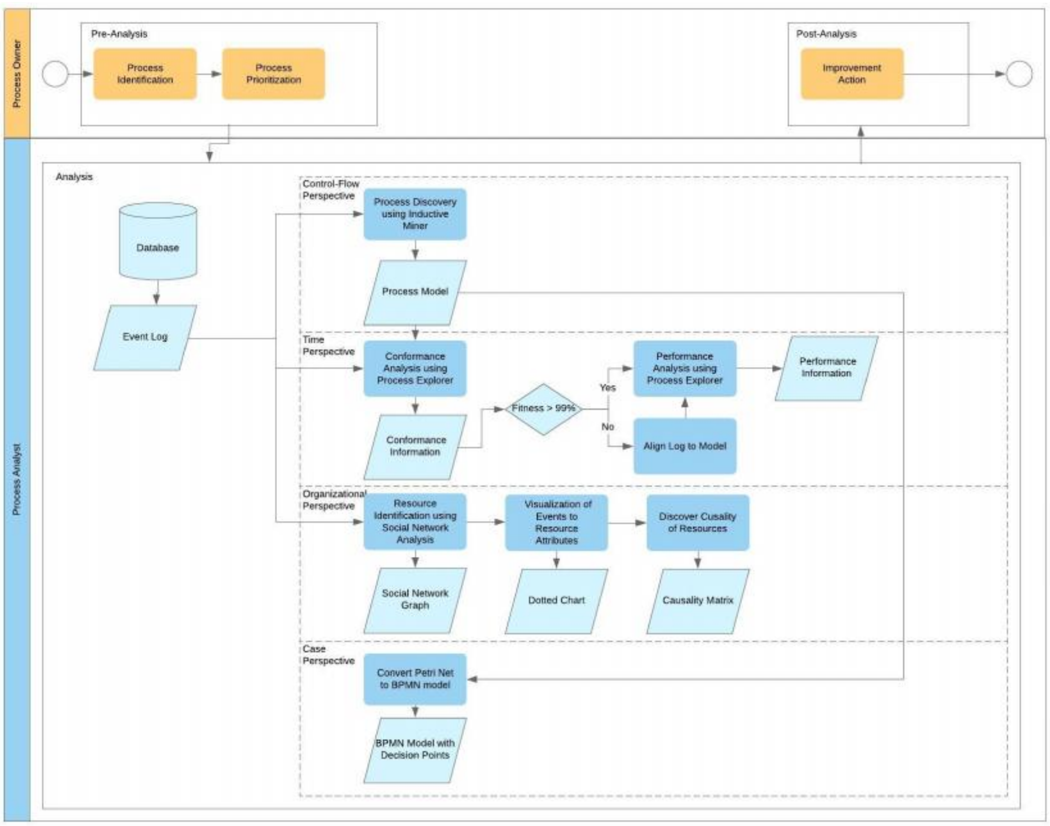


Figure 5.7 Enhancement Approach Final Version

Also, in the experimental evaluation, we encountered a usability problem with the Decision-Tree Miner. In fact, all the respondents prefer to identify decision points in the process from the control-flow perspective. Therefore, for the case perspective, we replaced the Decision-Tree Miner plug-in with "Convert Petri Net to BPMN" plug-in. This change depicted in the figure above.

6. Discussion and Future Work

The main goal of this study is to produce an artifact that can guides enhancement activity in process mining, through the exploration of the current state of the art of enhancement practices in the process mining field. Below are the answers to the research questions constructed to achieve the research's goal:

> **RQ1: What is the current state-of-the-art of enhancement practices in process mining?**
>
> Our systematic literature reviews show that process mining is adopted in various industries. Within the scope of enhancement activity, organizational perspectives are largely discussed. On the other hand, there's a few discussion and reliable support for case perspectives. In term of the tools used in practices, ProM and Disco used the most. Both tools support the analysis of XES event log file. However, Disco only focuses on the control-flow perspective. At the other hand, ProM also supports the analysis for Organizational, Time and Case perspectives. ProM can produce a process model in different notations, but the most used in practices are Petri Net and BPMN. However, conformance and performance analysis for BPMN is not supported in ProM.
>
> **RQ2: How can we assist process mining users in conducting enhancement activity using process mining techniques?**
>
> To assist the process mining user in conducting the enhancement activity, first, the profile of the users is identified. Process owner and process analyst are crucial stakeholders in process analysis using process mining techniques. For them, a three-phased enhancement approach is presented. In the pre-analysis, process owner determines which process needs to be analyzed through process identification and prioritization. After that, in the analysis phase, process analyst discovers the control-flow, time, organizational and case perspectives of the process. Based on the analysis results, the decision of process improvement is left to the process owner.
>
> **RQ3: Does the proposed artifact applicable in the real-life case?**
>
> After the initial design of the approach is created, the author applied it to two real-life data sets. The case experiment mechanism shows that the approach can be implemented for both processes with some additional and conditional steps. The additional steps then incorporated to the second version of the artifact. After that, the artifact is evaluated by two novice process mining users by using it to perform process analysis tasks. The evaluation shows that the discovery of case perspectives using decision-tree miner is not practical. Therefore, in the final version of the artifact, we replace it with the conversion of Petri Net to BPMN.

The objective of the artifact is to demonstrate a successful interaction with the process mining tools so that it can increase the perceived ease of use and self-efficacy of novice users towards process mining tools. The higher perceived ease of use and self-efficacy leads to higher perceived usefulness of the tools. This is important because, for people who do not have any prior knowledge in process mining, the currently available way to structurally know about the process mining topic would be to either read the 400-pages of process mining books or following the four-weeks online tutorial. Both by the founder and leading process mining research community of TU Eindhoven. Besides that, ProM provides a getting started tutorial for installation on their website[7]. However, a large part of it does not specifically address practical steps. For example, they explain the definition of process discovery but does not explains what are the plug-ins that can be used to discover the process model.

In addition, there's no standardized and centralized plug-ins catalog that explains the demonstration of plug-ins usability and their expected outcome in ProM. This might create confusion since some of the plug-ins only available on a certain version of the tools. For example, the popular plug-in for discovering organizational perspective in ProM 5.2 is the "Organization Miner," but it is not available in the ProM 6. Unfortunately, this kind of information can only be obtained by exploring process mining forums. Moreover, process mining forum is a closed community. Even though the discussion threads can be accessed by the public, to submit a question or to start a discussion about a certain topic, one needs to request to be invited as the member of the forum.

If potential process mining users want to explore approaches available in process mining through the scientific literature, then they will be faced by a large variety of practices. They are different in terms of tools, process modeling language, event log mechanism, and plug-ins. However, some tools provide more feature than another, some modeling language is more suitable for process mining, and some event log mechanism is preferred than the others within the process mining community. By understanding the state-of-the-art and the differences between the approach, we can find some practices that can be adapted to delivers a holistic process analysis from four process perspectives.

We proposed an approach to conduct enhancement in process mining based on the current state-of-the-art of the field. Then, a "guideline" consists of the elaboration, and the demonstration of the approach is presented in the form a white paper. The experimental evaluation using TAR shows that two respondents who are novice process mining users were able to perform process analysis with the assistance of the guideline. After that, we asked two respondents about their opinion of the guideline. The respondents stated that they think that the proposed guideline has helped them in understanding about process mining. They also described that the guidelines and the exercise intrigue their curiosity to learn more about process mining. Furthermore, we evaluate the respondent's user acceptance towards ProM.

6.1. Our Contribution

This study makes several contributions. This thesis provides contributions to both practices and theory. In summary, our contributions are the following:

1. **Presentation of the current state-of-the-art of process mining fields.** Besides providing the foundation for our research, our scoping study highlight where are gaps in current process mining practice. For example, from the literature review, we know that the case perspective is widely discussed. Moreover, through the constructions of literature graph in Appendix C and the table summary of the literature in Appendix B, we help process mining practitioners to explore the available studies within the scope of enhancement.

[7] see http://www.promtools.org/doku.php?id=gettingstarted:start

2. **Evaluation of the most used practices in process mining through experiments.**
   Through our Case Mechanism Experiment, we contribute to the evaluates the existing process mining plug-ins, especially the plug-ins that are used in our artifact. For example, even though social network analysis is widely used for the discovery of organizational perspectives, we found that the social network analysis alone is not sufficient for linking the resource attribute to the activity they perform.
3. **The creation of an approach to conduct enhancement activity in process mining.**
   Process mining is presented as a revolutionary instrument to improve processes, but very little attention is given on how to execute a complete project (Driessen, 2013). We contribute to present a process analysis of the four process perspectives namely control-flow, organizational, time and case perspective by creating a guideline. The guideline is a white-paper version of the approach and its demonstration. Therefore, it presents a practical process analysis support for process analysts.
4. **Linking the organizational aspects to the technical part of the process mining project.**
   A survey about the state-of-the-art of data science and machine learning with more than 16,000 responses reveals that four out of seven obstacles data scientists face at work are non-technical including (Kaggle, 2017). There is a demand for linking the technical part of data science into the organizational context. We integrate the two part of process analysis by involving the process owner to the pre-analysis and post analysis.

6.2. Limitations and Future Work

Conducting research in a relatively young field like process mining is challenging. The first challenge lies in the limitation of the available literature. A lot of practical information is available in the process mining forums instead of scientific literature. This informations are valuable but are not eligible to be cited in a scientific study like this thesis.

Also, there is a limitation in ProM tools. The development of ProM tools depends on the initiative of the researcher within the process mining community. Moreover, there's a tendency of research focus in a certain spectrum of process mining. Also, because ProM tools is not a commercial tool, there's a lack of quality assurance for the tools. Some underperforming plug-ins are still available and widely used in practices even though it was proven unreliable by some studies. This limitation gives motivation for future research or practices in process mining community to consolidates the practical knowledge about process mining practices.

We faced a challenge in finding respondents in the evaluation. The respondents are the students who has completed or is enrolled in the data science course at the University of Twente. Unfortunately, there are only a few students choose the process mining track.

Therefore, we're left with four respondents. For future work, an empirical study to check the generalizability of the proposed artifact is necessary.

Lastly, the demonstration of the proposed approach in this study uses freely available event log provided by the Process Mining Community. Thus, during the demonstration phase, the author of this thesis plays a role as a process analysis and not as a process owner. Therefore, there is a lack of understanding about the business context of the process. In the future, case studies in which the proposed approach is implemented by a certain organization might provide more insight about the usability of the proposed approach.

References

1. Adriansyah, A. and J. C. A. M. Buijs (2013). Mining process performance from event logs. Lecture Notes in Business Information Processing. 132 LNBIP: 217-218.
2. Anuwatvisit, S., et al. (2012). Bottleneck mining and Petri Net simulation in education situations. International Conference on ICT and Knowledge Engineering.
3. Appice, A., et al. (2016). Discovering and tracking organizational structures in event logs. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9607: 46-60
4. Appice, A. and D. Malerba (2016). A Co-Training Strategy for Multiple View Clustering in Process Mining. IEEE Transactions on Services Computing 9(6): 832-845.
5. Armas Cervantes, A., et al. (2017). Interactive and incremental business process model repair. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10573 LNCS: 53-74.
6. Ballambettu, N. P., et al. (2017). Analyzing process variants to understand differences in key performance indices. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10253 LNCS: 298-313.
7. Basile, F., et al. (2015). Model repair of Time Petri Nets with temporal anomalies. IFAC-PapersOnLine 28(7): 85-90.
8. Bozkaya, M., et al. (2009). Process diagnostics: A method based on process mining. Proceedings International Conference on Information, Process, and Knowledge Management, eKNOW 2009.
9. Burattin, A., et al. (2013). Business models enhancement through discovery of roles. 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM).
10. De Leoni, M., et al. (2016). A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. Information Systems 56: 235-257.
11. Djedovic, A., et al. (2016). Model business process improvement by statistical analysis of the users' conduct in the process. 2016 International Multidisciplinary Conference on Computer and Energy Science, SpliTech 2016.
12. Fahland, D. and W. M. P. Van Der Aalst (2015). Model repair - Aligning process models to reality. Information Systems 47: 220-243.
13. Ferreira, D. R. and C. Alves (2012). Discovering user communities in large event logs. Lecture Notes in Business Information Processing. 99
14. Gupta, M. (2014). Nirikshan: process mining software repositories to identify inefficiencies, imperfections, and enhance existing process capabilities. Companion Proceedings of the 36th International Conference on Software Engineering. Hyderabad, India, ACM: 658-661.
15. Gupta, M. and A. Sureka (2014). Process Cube for Software Defect Resolution. 2014 21st Asia-Pacific Software Engineering Conference.
16. Hidayat, B. N. A., et al. (2016). Process model extension using heuristics miner: (Case study: Incident management of Volvo IT Belgium). 2016 International Conference on Computational Intelligence and Cybernetics.
17. Kim, K., et al. (2012). Discovery of information diffusion process in social networks. IEICE Transactions on Information and Systems E95-D(5): 1539-1542.
18. Mannhardt, F., et al. (2015). Extending process logs with events from supplementary sources. Lecture Notes in Business Information Processing. 202: 235-247.
19. Mannhardt, F., et al. (2016). Measuring the precision of multi-perspective process models. Lecture Notes in Business Information Processing. 256: 113-125.
20. Mans, R. S., et al. (2012). Mining processes in dentistry. IHI'12 - Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium.
21. Mans, R. S., et al. (2008). Process mining in healthcare - A case study. HEALTHINF 2008 - 1st International Conference on Health Informatics, Proceedings.

22. Mitsyuk, A. A., et al. (2017). Process model repair by detecting unfitting fragments? CEUR Workshop Proceedings.
23. Polyvyanyy, A., et al. (2016). Impact-driven process model repair. ACM Transactions on Software Engineering and Methodology 25(4).
24. Rogge-Solti, A., et al. (2013). Repairing event logs using timed process models. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 8186 LNCS: 705-708.
25. Rouibah, K., et al. (2007). Combining workflow and PDM based on the workflow management coalition and STEP standards: The case of axalant. International Journal of Computer Integrated Manufacturing 20(8): 811-827.
26. Rozinat, A., et al. (2009). Discovering simulation models. Information Systems 34(3): 305-327.
27. Schönig, S., et al. (2015). Mining the organisational perspective in agile business processes. Lecture Notes in Business Information Processing. 214: 37-52.
28. Schönig, S., et al. (2016). Discovery of multi-perspective declarative process models. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9936 LNCS: 87-103.
29. Sturm, C., et al. (2017). Distributed multi-perspective declare discovery. CEUR Workshop Proceedings.
30. Swennen, M., et al. (2016). Capturing resource behaviour from event logs. CEUR Workshop Proceedings.
31. Turi, A., et al. (2008). Distributed discovery of multi-level approximate process patterns. SEBD 2008 - Proceedings of the 16th Italian Symposium on Advanced Database Systems.
32. van der Aalst, W. M. P., et al. (2007). Business process mining: An industrial application.
33. Wang, J., et al. (2016). Mining organizational behaviors in collaborative logistics chain: An empirical study in a port. 2016 International Conference on Logistics, Informatics and Service Sciences (LISS).
34. Werner, M. (2013). Colored Petri Nets for integrating the data perspective in process audits. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 8217 LNCS: 387-394.
35. Wongvigran, S. and W. Premchaiswadi (2015). Analysis of call-center operational data using role hierarchy miner. International Conference on ICT and Knowledge Engineering.
36. Zhang, X., et al. (2018). An Approach for Repairing Process Models Based on Logic Petri Nets. IEEE Access 6: 29926-29939.
37. Zhao, W., et al. (2016). An entropy-based clustering ensemble method to support resource allocation in business process management. Knowledge and Information Systems 48(2): 305-330.
38. Stroinski, A., et al. (2017). A distributed discovery of communicating resource systems models. IEEE Transactions on Services Computing: 1-1.
39. van Eck, M. L., et al. (2016). Discovering and exploring state-based models for multi-perspective processes. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).
40. Chamorro, M. and S. Maturana (2017). Method for Applying Process Mining to the Distribution of Non-alcoholic Beverages. Proceedings - International Conference of the Chilean Computer Science Society, SCCC.
41. Schönig, S., et al. (2016). A framework for efficiently mining the organisational perspective of business processes. CEUR Workshop Proceedings.
42. Ikeda, M., et al. (2014). Formal concept analysis for process enhancement based on a pair of perspectives. CEUR Workshop Proceedings.
43. Schonig, S., et al. (2016). Efficient and customisable declarative process mining with SQL. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9694: 290-305.
44. Aalst, W. v. d. (2016). *Process Mining Data Science in Action.* Springer.

45. Aalst, W. v. d., & Weijters. (2004). Process mining: A research agenda.
46. Harris, H., Vaisman, M., & Murphy, S. (2013). Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work. In *Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work*: O'Reilly.
47. Rao, S. S. (Producer). (1998). Diaper-beer syndrome. Retrieved from https://www.forbes.com/forbes/1998/0406/6107128a.html#2ec360f86260
48. Aalst, Netjes, & Reijers (2007). Supporting the Full BPM Life-Cycle Using Process Mining and Intelligent Redesign. *Contemporary Issues in Database Design and Information Systems Development*. doi:10.4018/9781599042893.ch004. Retrieved from http://wwwis.win.tue.nl/~wvdaalst/publications/p381.pdf
49. Rosemann, M., & Brocke, J. V. (2015). Business Process Management. *Handbook on Business Process Management 1*. Retrieved from https://www.researchgate.net/publication/313993293_Business_Process_Management
50. Wieringa, R. J. (2014). Design Science Methodology for Information System and Software Engineering. doi:10.1007/978-3-662-43839-8__1
51. Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (n.d.). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems /
52. Goedertier, Stijn & Vanthienen, Jan. (2007). Declarative Process Modeling with Business Vocabulary and Business Rules. 603-612. 10.1007/978-3-540-76888-3_83. Retrieved from https://www.researchgate.net/publication/225152137_Declarative_Process_Modeling_with_Business_Vocabulary_and_Business_Rules/citations
53. Davis, R., & Brabander, E. (2007). *ARIS Design Platform: Getting Started with BPM.* Springer.
54. Object Management Group. (n.d.). BPMN Specification - Business Process Model and Notation. Retrieved from http://www.bpmn.org/
55. Van Der Aalst, W. (2013). Business Process Management: A Comprehensive Survey. ISRN Software Engineering,2013, 1-37. doi:10.1155/2013/507984. Retrived from http://downloads.hindawi.com/journals/isrn.software.engineering/2013/507984.pdf
56. Burattin, A. (2015). Process Mining. Process Mining Techniques in Business Environments Lecture Notes in Business Information Processing, 33-47. doi:10.1007/978-3-319-17482-2_5
57. Fluxicon. (n.d.). Products. Retrieved from https://fluxicon.com/disco/
58. Process Mining Group. (2014). ProM Tools. Retrieved from http://www.promtools.org/doku.php
59. van der Aalst, W., et al. (2012). Process Mining Manifesto, Berlin, Heidelberg, Springer Berlin Heidelberg.
60. Song, M., & Aalst, W. M. (2008). Towards comprehensive support for organizational mining. Decision Support Systems, 46(1), 300-317. doi:10.1016/j.dss.2008.07.002
61. Dumas, M., Rosa, M. L., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of business process management.* Berlin, Germany: Springer.
62. Xiao, Y., & Watson, M. (2017). Guidance on Conducting a Systematic Literature Review. Journal of Planning Education and Research. doi:10.1177/0739456x17723971
63. Aalst, W. M., Hee, K. M., Werf, J. M., & Verdonk, M. (2010). Auditing 2.0: Using Process Mining to Support Tomorrows Auditor. *Computer,43*(3), 90-93. doi:10.1109/mc.2010.61
64. Tversky, A. (1972). Choice by elimination. *Journal of Mathematical Psychology,9*(4), 341-367. doi:10.1016/0022-2496(72)90011-9
65. Recker, J., Indulska, M., Rosemann, M., & Green, P. (2010). The ontological deficiencies of process modeling in practice. European Journal of Information Systems, 19(5), 501-525. doi:10.1057/ejis.2010.38
66. Devi, Kumudavalli, Sudhamani. (2017). An Informative and Comparative Study of Process Mining Tools. International Journal of Scientific & Engineering.
67. "Enhance". (n.d.). In *Merriam-Webster.com*. Merriam-Webster Dictionary. Retrieved in October 2018, from https://www.merriam-webster.com/dictionary/enhance

68. Fluxicon Process Laboratories. (2009). Extensible Event Stream Standard Definition. Retrieved in January 2019, from http://www.xes-standard.org/_media/xes/xes_standard_proposal.pdf
69. Dijkman, Vanderfeesten & Reijers. (2016). Business process architectures: overview, comparison and framework, Enterprise Information Systems, 10:2, 129-158, DOI: 10.1080/17517575.2014.928951
70. Dijkman, Vanderfeesten & Reijers. (2011). The road to a business process architecture: an overview of approaches and their use. BETA Working Paper Series, WP 350. Eindhoven University of Technology, Eindhoven
71. Remco M. Dijkman. (n.d.). Designing a Process Architecture - A Concrete Approach. Technical report, TU Eindhoven
72. Sabbagh. (2015). Business Process Architectures: Concepts, Formalism, and Analysis. Hasso Plattner Institute, University of Postdam.
73. Harmon, Paul. (2014). Business Process Change, Third Edition: A Business Process Management Guide for Managers and Process Professionals, Morgan Kaufmann Publishers Inc.
74. Rojas, Munoz-Gama, Sepúlveda, Capurro. (2016). Process mining in healthcare: A literature review. Journal of Biomedical Informatics 61: 224-236.
75. Fluxicon. (2010). ProM Tips - Which Mining Algorithm Should You Use? - Flux Capacitor. Retrieved from https://fluxicon.com/blog/2010/10/prom-tips-mining-algorithm/
76. Ministerie van Volkshuisvesting Ruimtelijke Ordening en Milieu. (n.d.). Bouwregelgeving. Retrieved from https://www.rijksoverheid.nl/onderwerpen/bouwregelgeving
77. Buijs. (n.d.). Introduction to Process Mining with ProM. Retrieved from https://www.futurelearn.com/courses/process-mining/0/steps/15643
78. Ube van der Ham. (2015). Benchmarking of Five Dutch Municipalities with Process Mining Techniques Reveals Opportunities for Improvement. Fifth International Business Process Intelligence Challenge (BPIC'15).
79. Process Mining Group. (2016). ProM Lite 1.1. Retrieved from http://www.promtools.org/doku.php?id=promlite11
80. Robledo, P. (2018). Process Mining plays an essential role in Digital Transformation. Retrieved from https://medium.com/@pedrorobledobpm/process-mining-plays-an-essential-role-in-digital-transformation-384839236bbe
81. Kosorukoff, A., & Passmore, D. L. (2011). Social network analysis: Theory and applications. S.L.: Passmore, D. L.
82. Kaggle. (2017). The State of ML and Data Science 2017. Retrieved fromhttps://www.kaggle.com/surveys/2017
83. Aalst, van der, W. M. P., Reijers, H. A., & Song, M. S. (2005). Discovering social networks from event logs. Computer Supported Cooperative Work, 14(6), 549-593. DOI: 10.1007/s10606-005-9005-9
84. M. De Leoni & W.M.P. van der Aalst. (2013). Data-Aware Process Mining: Discovering Decisions in Processes Using Alignments. In S.Y. Shin and J.C. Maldonado, editors, ACM Symposiumon Applied Computing (SAC 2013), pages 1454–1461.
85. Rozinat and W.M.P. van der Aalst. (2008). Conformance Checking of Processes Based on Monitoring Real Behavior. Information Systems, 33(1):64–95.
86. Rokach & Lior. (2014). Data mining with decision trees: theory and applications - 2nd edition. World Scientific Publishing.
87. Karwehl. (2018). Use case based introduction to process mining and current tools. Hamburg University of Applied Science.
88. Kebede. (2015). Comparative Evaluation of Process Mining Tools. University of Tartu.
89. Mannhardt, Leoni, & Reijers. (2015). The Multi-Perspective Process Explorer. 13th International Workshops on Business Process Management Workshops (BPM 2015).
90. Rozinat, & Van Der Aalst. (2006). Decision Mining in ProM. Lecture Notes in Computer Science Business Process Management,420-425. doi:10.1007/11841760_33

91. Massimiliano. (2015). Decision Miner in 6.5. Retrieved from https://www.win.tue.nl/promforum/discussion/507/decision-miner-in-6-5
92. Peeters. (2016). Decision Mining in ProM: From Log Files to DMN Decision Logic. Universiteit Gent. Retrieved from http://lib.ugent.be/fulltxt/RUG01/002/273/963/RUG01-002273963_2016_0001_AC.pdf
93. Scott. (2012). What is Social network analysis? London: Bloomsbury Academic.
94. Kalenkova, Aalst, Lamazova & Rubin. (2017). Process mining using BPMN: relating event logs and process models. Software & Systems Modeling 16(4): 1019-1048.
95. Fu & Gray. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. Cognitive Science, 28(6), 901-935. doi:10.1016/j.cogsci.2004.03.005

Appendix A

Letter of Invitation

Dear Sir/Madam,

My name is Fitri Almira Yasmin and I'm a master student in the Business Information Technology with the specialization of IT Innovation and Management at the University of Twente located in Enschede, The Netherland. My area of expertise including, but not limited to, the Process Mining.

Process Mining is a novel data analysis technique that is not yet largely adopted in South-East Asia. Process mining is one of the skills that help to analyze, visualize and optimize big datasets with the focus on improvement of the end-to-end processes by interplaying the event data and process models. It offers more broad improvement possibilities to your company in term of organizational structure, decision-making, organizational performance and workflow.

As I'm currently in the process of planning my Master Research in the aforementioned field. I would like to offer to do a collaboration with your company. Through process mining techniques, I could identify the current business process within the scope of your company or more specific context such as the process related to a specific program such as seller performance derived from their event log and based on that, I could provide the business enhancement solutions and/or suggestions to your company.

I would like to mention that my skills are not limited to only process mining, you can find more details regarding my research area in the attached document. I will also enclose the example of process mining research that my colleague and I previously conducted, to give you an idea about the topics.

Thank you for your time and consideration, if you have any further questions, please feel free to contact me.

Regards,

Fitri Almira Yasmin | Business Information Technology |
University of Twente | Faculty of Electrical Engineering, Mathematics and Computer Science |
T: +31633376015 | fitrialmirayasmin@student.utwente.nl |
https://www.linkedin.com/in/fitri-almira-yasmin-350384102

Appendix B

Table of Literature Summary

No	Citation	Goal	Results	Datasets		Techniques	Tools	Plug-ins	Enhancement Definition	Notes
				Process Owner	**Quantity**					
1	Adriansyah, A. and J. C. A. M. Buijs (2013). Mining process performance from event logs. Lecture Notes in Business Information Processing. 132 LNBIP: 217-218.	To find an optimal alignment between the process model and the traces; that matches as many movement of process model with movement of the traces manually	Ability to pinpoint bottlenecks and automated activity major deviations can be fixed manually	Dutch Municipality	13,087 Traces 262,200 Events	Repair	ProM 6.1	Log Filtering α-algorithm Heuristics miner ILP-Miner Replay for Conformance and Performance Checking*	Enhancement of a process model uses the recorded behavior to project information, such as performance or decision information, on the process model.	A proposed framework is a repair techniques
2	Anuwatvisit, S., et al. (2012). Bottleneck mining and Petri Net simulation in education situations. International Conference on ICT and Knowledge Engineering.	To know whether the model and the log conform to each other; To detect violations and ensure transparency	A broad overview of the process in the information system within a short period of time	A University in Thailand	308 Cases	Extension: Organizational Perspective	ProM ProM Import Framework	ProM Analysis (Dotted Chart) Log Filtering 80% Process Discovery* Conformance Checker*	NA	Include Role Analysis about the division of work within departments of the organization, whether people across departments execute the system to make their work easier.
3	Appice, A., et al. (2016). Discovering and tracking organizational structures in event logs. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9607: 46-60	Focuses on the idea of performing organizational mining of event logs via social network mining.	Detect how organization structured in business process and discover communities overtime	Dutch Financial Institute (BPI 2012) Dutch Academic Hospital	13,087 Traces 262,200 Events 69 Resources 1,143 Traces 150,291 Events	Extension: Organizational Perspective	NA	NA	NA	The proposed TOSTracker (Time-evolving Organizational Structure Tracker)
4	Appice, A. and D. Malerba (2016). A Co-Training Strategy for Multiple View Clustering in Process Mining. IEEE Transactions on Services Computing 9(6): 832-845.	To present a multiple view clustering based on co-training strategy	Traces clustered have high compactness compared to alternatives; it's also minimize problem of spagehetti-like models	Volvo IT Belgium - cProblem - incident - oProblem Loan Dutch Academic Hospital (BPI 2011) ISBPM Photo Repair Review	 1,487 Traces 6,660 Events 7,554 Traces 65,533 Events 819 Traces 351 Events 475 Traces 2,440 Events 1,143 Traces 150,291 Events 2,000 Traces 28,534 Events 100 Traces 40,995 Events 1,104 Traces 236,360 Events 3,512 Traces 46,138 Events	Repair	ProM	Alpha++ mining algorithm Petri-Net Complexity Analysis plug-in Conformance checker plug-in	NA	It splits up an event log into clusters of similar traces, so as to handle variability in the recorded behavior and facilitate process model discovery.

5	Armas Cervantes, A., et al. (2017). Interactive and incremental business process model repair. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10573 LNCS: 53-74.	Advocates for an interactive and incremental approach to process model repair, where differences between the model and the log are visually displayed to the user, and the user repairs each difference manually based on the provided visual guidance.	Higher F-score and higher structural similarity between observed and discovered model; Provide visual guidance for repair	Managing Road Traffic Fines in Italy	150,370 Traces561,470 Events	Repair	Apromore Onle Process Analytics Platform (Compare Plug-ins)BIMP Simulator (Generate Event Log)	Compare Plug-ins	NA	It compared event structure of the log and that of the model
6	Ballambettu, N. P., et al. (2017). Analyzing process variants to understand differences in key performance indices. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10253 LNCS: 298-313.	Identify process variants and key differeces among them annotated into a porcess map	Provide insight at various level regarding process variants	Service Delivery Organization	NA	Extension: Time Perspective	ProM	Heuristics miner plug-in Analyze process variants (their own)	NA	Include time perspective such as flow time, activity execution times, etc as determining process variances and then implement it as plug-ins
7	Basile, F., et al. (2015). Model repair of Time Petri Nets with temporal anomalies. IFAC-PapersOnLine 28(7): 85-90.	The model repair for timed DESs that can exhibit temporal anomalies	Mixed-Integer Linear Programming Approach without modifying the structure of the Petri Net nominal model	NA	NA	Repair	Cplex (Optimization tools)	NA	An existing process model is extended or improved using information about the actual process recorded in some event log	Modify firing interval of the nominal model
8	Bozkaya, M., et al. (2009). Process diagnostics: A method based on process mining. Proceedings - International Conference on Information, Process, and Knowledge Management, eKNOW 2009.	Propose a methodology to perform process diagnostic based on process mining	A broad overview of the process supported by information system that can be performed in short period of time	Dutch Governmental Organization	60 Activity 83,611 Cases 276,333 Events 60 Originators	Extension: Organizational Perspective	ProM	Fuzzy Miner Performance Sequence Analysis plug-in Dotted chart analysis Performance analysis Organizational Miner Social network analysis	NA	Include organizational perspective
9	Burattin, A., et al. (2013). Business models enhancement through discovery of roles. 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM).	Proposes an approach to enhance a business process model with information on roles. (BPMN)	Measures of handover of roles are defined and employed; and an approach to automatically extract only the significant partitioning is shown too	NA	NA	Extension: Organizational Perspective	ProM	PLG (Artificial process originator) ProM (Adding originators)	Enrich a process model, given as input, with new information extracted starting from an input log.	The identification of roles is based on the detection of handover of roles.
10	De Leoni, M., et al. (2016). A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. Information Systems 56: 235-257.	Propose framework that unifies a number of approaches for correlation analysis proposed in literature	FeaturePrediction plug-ins in ProM that created a decision or a regression tree that can be used to split event logs into cluster of traces with similar behabior	UWV (Dutch governmental institute in charge of supplying benefits)	NA	Extension: Time, Organization, and Cost Perspective	ProM Weka	Inductive Miner FeaturePrediction (their own plug-in)	NA	Analysis use case using various process characteristics that focus on data-flow, time, organization and conformance perspective

11	Djedovic, A., et al. (2016). Model business process improvement by statistical analysis of the users' conduct in the process. 2016 International Multidisciplinary Conference on Computer and Energy Science, SpliTech 2016.	Introduce a process discovery method that combines an organizational perspective with probabilistic approach (BPMN)	Identification of more qualified user; Identification of complex steps in the process	Bank in Bosnia and Herzegovina	NA	Extension: Organizational Perspective	BPMN 2.0Oracle BPM Suite	NA	Enhancement. It takes an event log and process model and extend or improve the model using the observed events.	Statistical analysis of users conduct in the process to improve business process model
12	Fahland, D. and W. M. P. Van Der Aalst (2015). Model repair - Aligning process models to reality. Information Systems 47: 220-243.	Investigate the problem of repairing a process model w.r.t a log such that the resulting model can replay to the log and is as similar as possible to the original model	The repaired model allows to understand how the original model deviated and had to be changed to achieve conformance to the log.	Dutch Municipality	NA	Repair: Alignement	ProM	Filter Log Using Simple Heuristics, ActiTrac Trace Alignment (with Guide Tree) Filter Out Unmapped Event Classes Construct Log From Alignment Align Log to Model Repair Model (find loops) Repair Model (find subprocesses) Repair Model (remove unused parts) Align Log And Model for Repair (global costs)	NA	The paper itself defines the repair techniques
13	Ferreira, D. R. and C. Alves (2012). Discovering user communities in large event logs. Lecture Notes in Business Information Processing. 99	Describe how to use hierarchical clustering together with the concept of modularity to analyze social networks obtained from large event logs	Discover community structure; An effective means to discover user groups based on the actual user interactions (working together instead of task similarity metrics)	Hospital of Sao Sebastiao in Santa da Feira, Portugal	1,836 Traces 11,506 Events 4,851 Traces 22,803 Events 78,623 Traces 536,735 Events	Extension: Organizational Perspective	ProM	Hierarchical Clustering with Modularity *their own plug-in	NA	The proposed approach is part of mining organizational perspective
14	Gupta, M. (2014). Nirikshan: process mining software repositories to identify inefficiencies, imperfections, and enhance existing process capabilities. Companion Proceedings of the 36th International Conference on Software Engineering. Hyderabad, India, ACM: 658-661.	Propose a research framework 'Nirikshan' to process mine the data of software repositories from multiple perspectives like process, organizational, data and time	Framework on process model event logs mined from software repositories for discovering runtime, inefficiencies and inconsistencies; a recommendation system for a better workload management	FLOSS (software repositories of large open source)	NA	Extension: Organizational Perspective	ProM Disco	Dotted chart visualization Fuzzy Miner Discover runtime process model Performance analysis Conformance verification Social network analysis	NA	Analysis from organizational perspective to understand team compatibility, perform delta analysis between the process adopted by long term contributors (LTC) and novice (specifically for open source), identify specialists and generalists.

15	Gupta, M. and A. Sureka (2014). Process Cube for Software Defect Resolution. 2014 21st Asia-Pacific Software Engineering Conference.	Process cube facilitates process mining from multiple-dimensions and enables comparison of process mining results across various dimensions.	A framework to mine repositories using process cube	Downloaded from Google ITS, Rietveld PCR and subersion VCS of Google Chromium Browser Project	NA	Extension: Time and Organizational perspective.	ProM and the Disco SQL	Discover PM in Disco (Fuzzy Miner)	NA	Every cells of the process cube is defined by metrics from Process Mining perspective like time, control-flow, conformance and organizational
16	Hidayat, B. N. A., et al. (2016). Process model extension using heuristics miner: (Case study: Incident management of Volvo IT Belgium). 2016 International Conference on Computational Intelligence and Cybernetics.	Proves the usability of heuristics miner in discovering process model of the real case study and additional information could then be used to extend process model for further analysis	A process model with precision, recall and f-measure > 96%; reflect on the quality of the incident management process and improve their decision making in resource allocation and process improvement, among others.	Incident Management of Volvo IT Belgium	149 Cases2771 Events	Extension: Time and Organizational perspectives	ProM	Heuristics MinerConformance CheckerBottleneck analysisRole analysis of originatorsTime – Impact analysis	Enhancement is the third type of process mining focuses on developing and improving the existing process model using additional information about the process recorded in the event log	The use of several organizational mining techniques
17	Kim, K., et al. (2012). Discovery of information diffusion process in social networks. IEICE Transactions on Information and Systems E95-D(5): 1539-1542.	Introduce approaches to discovering information diffusion process in social networks based on process mining. (The proposed approach will be made available online)		Tumblr	73 posts 2446 published times 433 different users	Extension: Organizational	ProM 5.2	Fuzzy Miner social network analysis α-algorithm	NA	Social Network Analysis and Community Recognition
18	Mannhardt, F., et al. (2015). Extending process logs with events from supplementary sources. Lecture Notes in Business Information Processing. 202: 235-247.	Present a method to extend an incomplete main event log with events from supplementary data sources, even though the latter lack references to the cases recorded in the main event log	LogEnhancement package in ProM to extend process logs with events from supplementary sources that cannot be trivially linked to specific traces in a main event log	Managing Road Traffic Fines in Italy	145,800 Traces 543,583 Events	Extension: Time, Resource, and Data Perspectives	ProM	LogEnhancement package *their own plug-in	NA	Correlate unlinked events from supplementary data sources (control-flow, time, resource, and data perspectives), i.e., events that have been recorded by additional systems, to the traces of the main event log.
19	Mannhardt, F., et al. (2016). Measuring the precision of multi-perspective process models. Lecture Notes in Business Information Processing. 256: 113-125.	Extends existing precision measures to incorporate the other perspectives (Data, Resource and Time)	Multi-perspective precision measure as a plug-in in the ProM framework	Managing Road Traffic Fines in Italy	11 different activities 150,000 traces 550,000 events	Extension: Data, Time and Resource Perspective	ProM	Multi-perspective precision measure *their own plug-in		AMBIGU
20	Mans, R. S., et al. (2012). Mining processes in dentistry. IHI'12 - Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium.	To demonstrate the usefulness of process mining for the domain of dentistry	Insight on the control-flow, organizational, and performance perspectives. (General and quick but the repeatability is questioned)	Medium-sized private dental practice in Netherland	55 patients 1542 events	Extension: Organizational Perspective	ProM 5.2 and ProM 6	Heuristics mining algorithm Conformance checker The Social Network miner Performance Analysis with Petri Nets	A-priori model is extended with a new aspect or perspective, i.e., the goal is not to check conformance but to enrich the model with the data from the event log. (EXTENSION)	Also focussed on the discovery of organizational aspects (resource perspective), as well as the discovery of performance related information (performance perspective)

21	Mans, R. S., et al. (2008). Process mining in healthcare - A case study. HEALTHINF 2008 - 1st International Conference on Health Informatics, Proceedings.	Focus on the applicability of process mining in the healthcare domain	Obtaining insights into the careflow by looking at the control-flow, organizational and performance perspective (Understandable but have room for improvement)	Dutch Academic Hospital (Gynecology)	627 Patients 376 Different event names	Extension: Organizational Perspective	ProM	Heuristic Miner Trace Clustering plug-in The Social Network Miner Dotted chart	A-priori model is extended with a new aspect or perspective, i.e., the goal is not to check conformance but to enrich the model with the data from the event log. (EXTENSION)	The use of social network miner
22	Mitsyuk, A. A., et al. (2017). Process model repair by detecting unfitting fragments? CEUR Workshop Proceedings.	Repair model by finding the non-conforming fragments in a model and replace them with conforming ones	Repairing model that allows to keep the initial model structure as much as possible (Without real case and siginificantly reduce precision)	NA	NA	Repair	Gena (Event Log Generation)ProM	ILP (Integer Linear Programming)-based algorithmInductive MinerAlignment-based fitness evaluation functionCalculate Graph Edit Distance Similarity	NA	Repair proposed
23	Polyvyanyy, A., et al. (2016). Impact-driven process model repair. ACM Transactions on Software Engineering and Methodology 25(4).	Propose different alternative approaches; Assigns predefined costs to repair actions (allowing inserting or skipping of activities).	Repair by checking all possible combinations of repair actions, i.e., inserting and skipping activities, we can always find an "optimal" repaired model within the range set by the available repair resources	Dutch Financial Institute (BPI 2012) Volvo IT Belgium (BPI 2013)	NA	Repair	ProM 6.3	Inductive miner Conformance Checker	NA	Propose various approach to repair
24	Rogge-Solti, A., et al. (2013). Repairing event logs using timed process models. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 8186 LNCS: 705-708.	Facilitate analysis of incomplete logs by combining stochastic Petri Nets, alignments, and Bayesian networks.	Method to repair timed event logs in order to make them available (Not yet validated in real life case)	NA	NA	Repair	NA	NA	NA	The proposed approach is a repair technique
25	Rouibah, K., et al. (2007). Combining workflow and PDM based on the workflow management coalition and STEP standards: The case of axalant. International Journal of Computer Integrated Manufacturing 20(8): 811-827.	To achieve inter-company workflow coordination, ease inter-company engineering change management, facilitate system integration and ease configuration management	Allows easy modifications to combine tasks, modify and reuse processes, and rearrange resources allocation to tasks, and thus give users more flexibility to deal with ad-hoc changes that may occur in their business (Case study with specific technology thus limited)	NA	NA	Extension: Organizational Perspective	ProView	NA	Enhancement includes the extension of the data model, the modification of the corresponding software, the modification of the user interface, which helps to generate graphical process definitions (EXTENSION)	Includes the used of enhancement of organizational structure through the usage of roles as a resource for process activities

26	Rozinat, A., et al. (2009). Discovering simulation models. Information Systems 34(3): 305-327.	Use a combination of process mining techniques to discover multiple perspectives (namely, the control-flow, data, performance, and resource perspective) of the process from historic data, and we integrate them into a comprehensive simulation model	analyze the process, e.g., evaluate the performance of different alternative designs; "what if" analysis, i.e., it allows to "look into the future" under certain assumptions	Dutch Municipality	363 Cases 1817 Events 570 Cases 6616 Events	Extension: Data and Resource Perspective	ProM	CPN Tools 2.0 Export plug-in a-algorithm Decision Miner (Decision point analysis) Performance analysis with Petri Net Organizational Miner	NA	The proposed plug-ins incorporate various perspective
27	Schönig, S., et al. (2015). Mining the organisational perspective in agile business processes. Lecture Notes in Business Information Processing. 214: 37-52.	Propose a process mining approach to discover resource-aware declarative process models (DECLARATIVE)	The approach implemented in DpilMiner to extract of complex rules for resource assignment that integrate the control-flow and organisational perspectives	A university business trip management system	2104 Events	Extension: Organizational Perspective	ProM DpilMiner	DeclareMiner	Enhancements aim to improve the mining performance as well as the readability of discovered models (DeclareMiner)	Technique for mining organizational perspective in declarative proces model
28	Schönig, S., et al. (2016). Discovery of multi-perspective declarative process models. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9936 LNCS: 87-103.	Present a full-fledged approach for the discovery of multi-perspective declarative process models from event logs that allows the user to discover Declarative models taking into consideration all the information an event log can provide. (DECLARATIVE)	MP-Declare allows for the acquisition of knowledge that goes beyond the classical declarative mining, which is focused only on the behavioral perspective in the vast majority of cases.	Managing Road Traffic Fines in ItalyDutch Municipalities	150,370 Traces561,470 Events	Extension: Data, Time and Resources perspective	SQL-based	NA	NA	Various perspective mining in Declarative process model
29	Sturm, C., et al. (2017). Distributed multi-perspective declare discovery. CEUR Workshop Proceedings.	Present a fast distributed approach and software prototype to discover multi-perspective declarative models out of event logs, based upon parallel computing	An efficient distributed mining framework for discovering MPDECLARE models that leverages latest big data analysis technology and builds upon the distributed processing method MapReduce.	Managing Road Traffic Fines in Italy Dutch Municipalities Dutch Hospital	NA	Extensions: Resource and Time perspective	MP-DECLARE	MapReduce Miner	NA	Various perspective mining in Declarative process model with the additional distributed processing
30	Swennen, M., et al. (2016). Capturing resource behaviour from event logs. CEUR Workshop Proceedings.	Extends the process metrics framework towards the resource perspective.	Unbiased, algorithm-agnostic information retrieved from event logs, extended with resouce perspective to retrieve useful insights in resource behaviour.	Hospital (Name Unknown, Not a real life case)	5.167 activity	Extension: Resource Perspective	R-package edeaR	NA	NA	Extending the original model of process metrics discovery by adding resource perspective

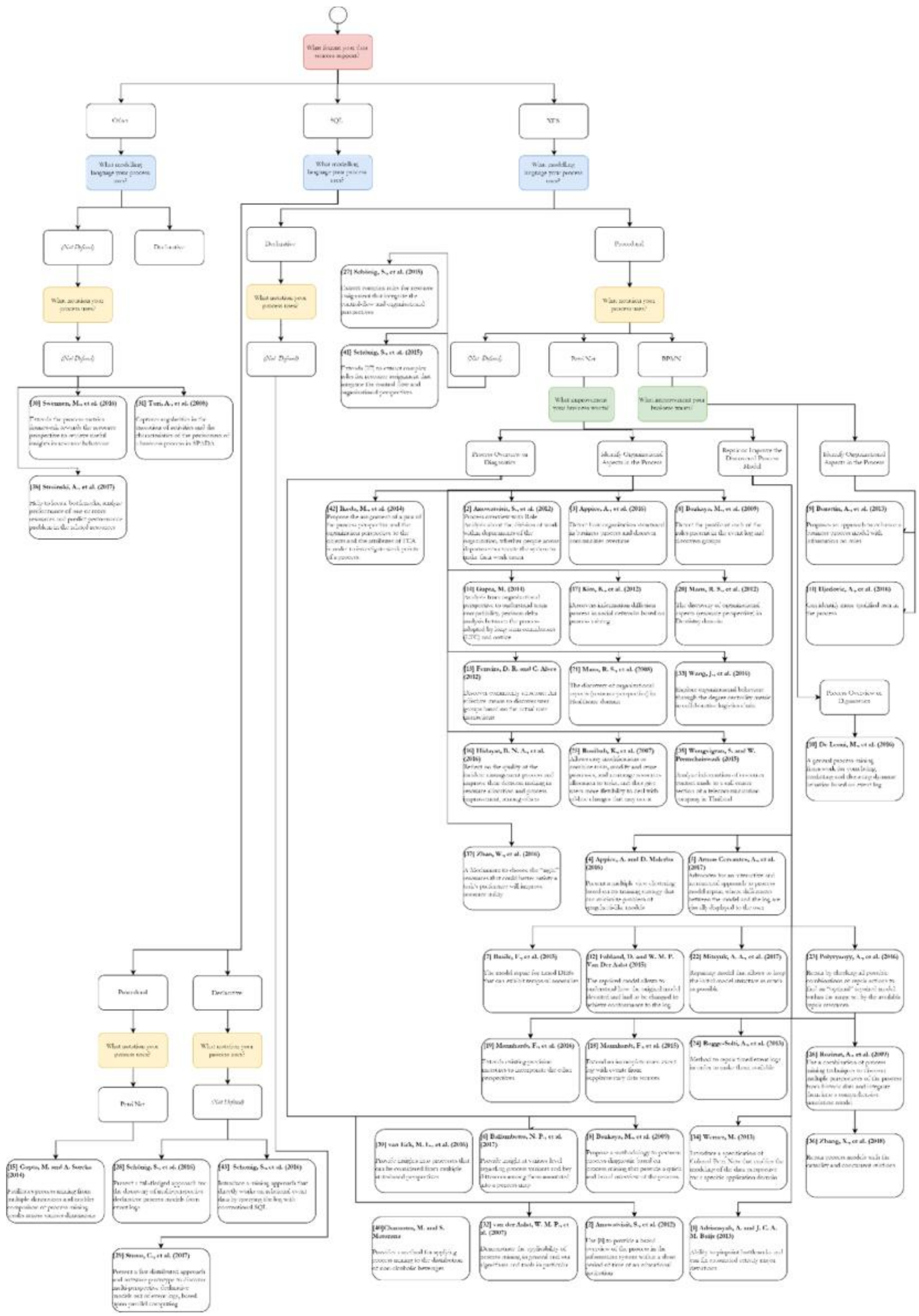
31	Turi, A., et al. (2008). Distributed discovery of multi-level approximate process patterns. SEBD 2008 - Proceedings of the 16th Italian Symposium on Advanced Database Systems.	The goal is to discover a set of process patterns frequently occurring in the log.	human interpretable patterns which captures regularities in the execution of activities and the characteristics of the performers of a business process	THINK3	21,256 Events	Extension	SPADA	NA		
32	van der Aalst, W. M. P., et al. (2007). Business process mining: An industrial application.	To demonstrate the applicability of process mining in general and our algorithms and tools in particular	The discovery of the main flow in the invoice handling process; the identification of places in the process where the circling of work is undesirable.	Dutch National Public Works Department	147,579 Events	Extension: Organizational and Case Perspective	ProM MiSoN (ProM Framework)	Heuristic Miner	NA	Analyzed the processing of invoices sent by the various subcontractors and suppliers from three different perspectives: the process perspective, the organizational perspective, and the case perspective.
33	Wang, J., et al. (2016). Mining organizational behaviors in collaborative logistics chain: An empirical study in a port. 2016 International Conference on Logistics, Informatics and Service Sciences (LISS).	Investigates the organizational behaviors in collaborative logistics chains from a process perspective using a synergy of process mining and social network analysis technique	To exploreorganizational behaviors through the degree centrality metric in collaborative logistics chain	Guangzhou Port inSouth China	5040 Cases 628,657 Events	Extension: Organizational Perspective	ProM	Heuristics Miner algorithmOrganization minerSocial Network Analysis		The organizational perpsective put into a study case
34	Werner, M. (2013). Colored Petri Nets for integrating the data perspective in process audits. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 8217 LNCS: 387-394.	Introduce a specification of Colored Petri Nets that enables the modeling of the data perspective for a specific application domain	Colored Petri Nets that allows the modeling of the control flow and data perspective simultaneously	SAP ERP	100,000 process instances	Extension: Data Perspective	ProM	Financial Process Mining (FPM) algorithm	NA	
35	Wongvigran, S. and W. Premchaiswadi (2015). Analysis of call-center operational data using role hierarchy miner. International Conference on ICT and Knowledge Engineering.	Analyze information of customer contact made to a call center section of a telecommunication company in Thailand	Identify the processes that are redundant/ excessive and the processes that get the most complaints within the study case	Telecommincation Company in Thailand	NA	Extension: Organizational Mining	Disco Fluxicon ProM 5.2	Role Hierarchy Miner	NA	
36	Zhang, X., et al. (2018). An Approach for Repairing Process Models Based on Logic Petri Nets. IEEE Access 6: 29926-29939.	Propose a repair approach based on an extended Petri Net named logic Petri Net	The proposed approaches can repair process models with the causality and concurrent relations	Hospital	1208 Traces 1598 Traces 1800 Traces	Repair	ProM 6.6	NA	The process enhancement is to extend or improve an existing process model with the generated event logs.	

37	Zhao, W., et al. (2016). An entropy-based clustering ensemble method to support resource allocation in business process management. Knowledge and Information Systems 48(2): 305-330.	Proposes a novel mechanism in which resource allocation is considered as a multi-criteria decision problem and solved by a new entropy-based clustering ensemble approach	Choosing the "right" resources that could better satisfy a task's preference will improve resource utility	Chinese hospital	11, 490 Cases 80, 430 Events	Extension: Resource Perspective	NA	NA	NA	
38	Stroinski, A., et al. (2017). A distributed discovery of communicating resource systems models. IEEE Transactions on Services Computing: 1-1.	Present a new distributed algorithm, dRMA, discovering process models from CRS event logs	The ability to discover a real model based on the historical execution of the system; Real models help to locate bottlenecks and analyze performance of one or more resources. In such cases, resource hierarchy allows us to predict, e.g. that if a performance problem occurs in some resource, then it will also occur in the sub-resources	Virtualized Environment	NA	Extension: Resource Perspective	NA	NA	NA	
39	van Eck, M. L., et al. (2016). Discovering and exploring state-based models for multi-perspective processes. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).	Present an approach to provide insights into processes that can be considered from multiple state-based perspectives	The tool highlights interesting relations between states and transitions graphically and quantifies them in terms of support, confidence and lift	Loan Application Process (BPI Challenge 2012)Phillips	262.200 Events358 Instance8369 State Entries	Extension: Case Perspective	ProM 6	CSMMiner	NA	The state is one of properties of cases, hence case perspective
40	Chamorro, M. and S. Maturana (2017). Method for Applying Process Mining to the Distribution of Non-alcoholic Beverages. Proceedings - International Conference of the Chilean Computer Science Society, SCCC.	Provides a method for applying process mining to the distribution of non-alcoholic beverages, aiming to increase the quality of service delivered to customers by making the process more transparent.	Guides the proposed the application of much of the knowledge generated in PM into a specific industry	Non-alcoholic beverage bottling companies in Chile	1000 cases	Extension:	ProM Disco	ProM 5.2	NA	Performance Analysis
41	Schönig, S., et al. (2016). A framework for efficiently mining the organisational perspective of business processes. CEUR Workshop Proceedings.	Develop an efficient and effective process mining framework that provides extensive support for the discovery of patterns related to resource assignment.	Declarative process mining approach for the organisational perspective, which supports all the creation patterns as well as what we called 25 cross-organisational patterns, which discover how the involvement of resources influences the control-flow of the process.	A university business trip management system	NA	Extension: Resource Perspective	ProM The DpilMiner	DeclareMiner	NA	The extended version of [27]

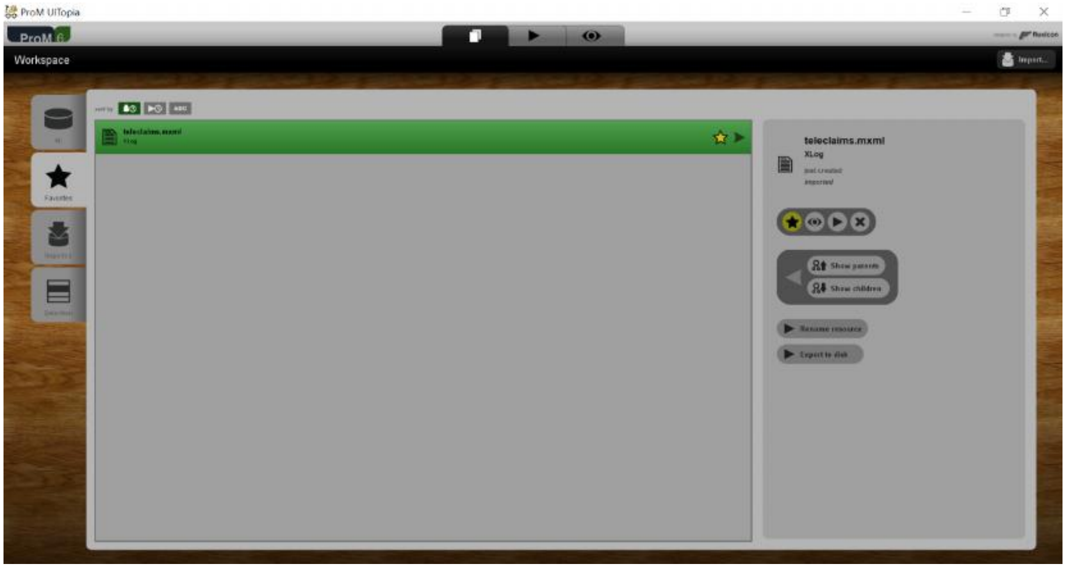
42	Ikeda, M., et al. (2014). Formal concept analysis for process enhancement based on a pair of perspectives. CEUR Workshop Proceedings.	Propose the assignment of a pair of the process perspective and the organization perspective to the objects and the attributes of FCA in order to investigate weak points of a process.	Bridging the fact that models focusing on one perspective are apt to neglect other perspectives, in this case, for process and resource perspective	NA	NA	Extension: Resource Perspective	NA	NA	NA	Combining the duality of the perspective, both process and resource. Bridging the gap of the current available approach.
43	Schonig, S., et al. (2016). Efficient and customisable declarative process mining with SQL. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9694: 290-305.	Introduce a mining approach that directly works on relational event data by querying the log with conventional SQL.	Customisation capabilities to discover commonly used process constraints by means of conventional SQL queries.	NA	NA	Extension: Resource Perspective	SQL-based	NA	NA	Declarative process

Appendix C

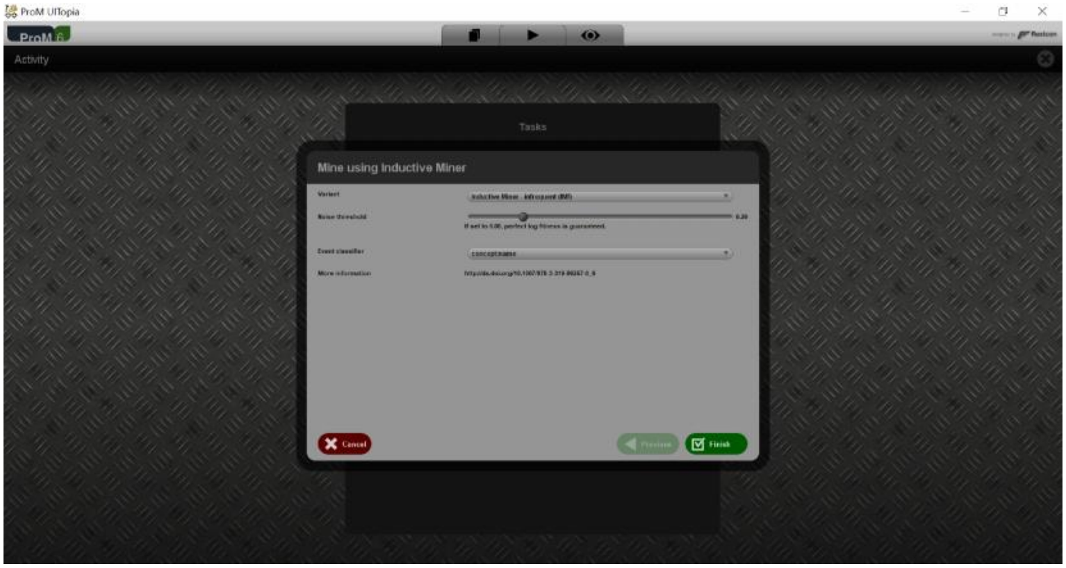
Literature Classification Graph (*please see [this link](#) for the actual size of the graph*).

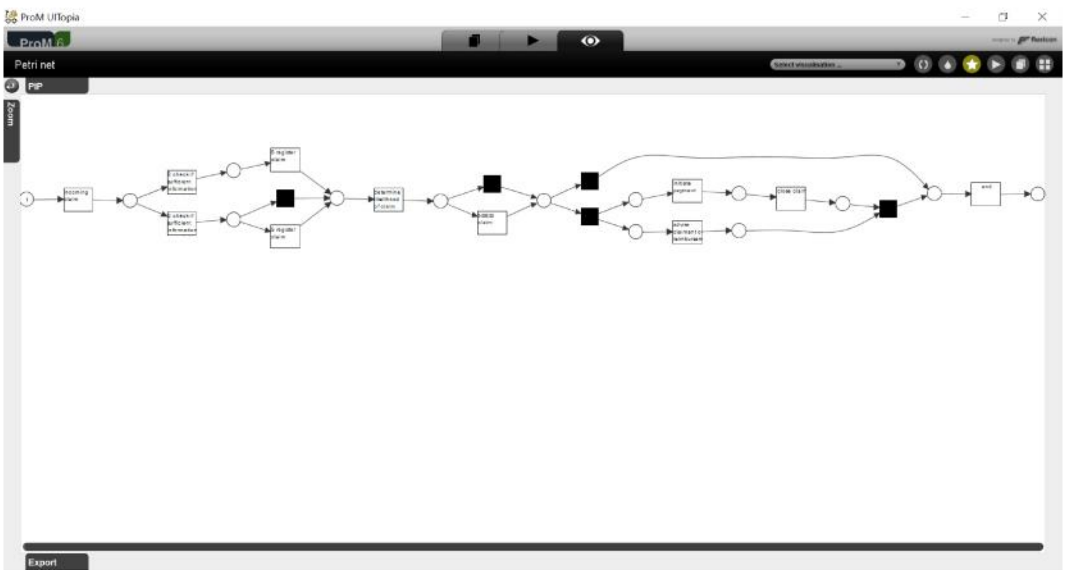


Appendix D


ProM UITopia
Workspace
teleclaims.mxml
XLog
Show parents
Show children


ProM UITopia
Actions
Input
Actions
Output
Mine Petri net with Inductive Miner
Mine process tree with Inductive Miner
Mine with Inductive visual Miner
Reset
Start
Plugin action info

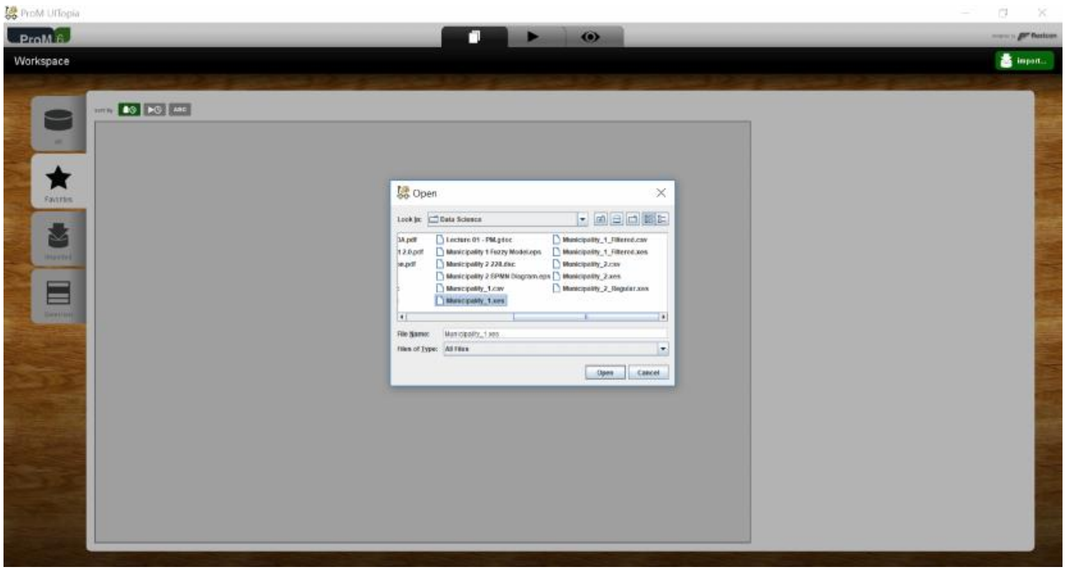

ProM UITopia
Activity
Tasks
Mine using Inductive Miner
Cancel
Finish

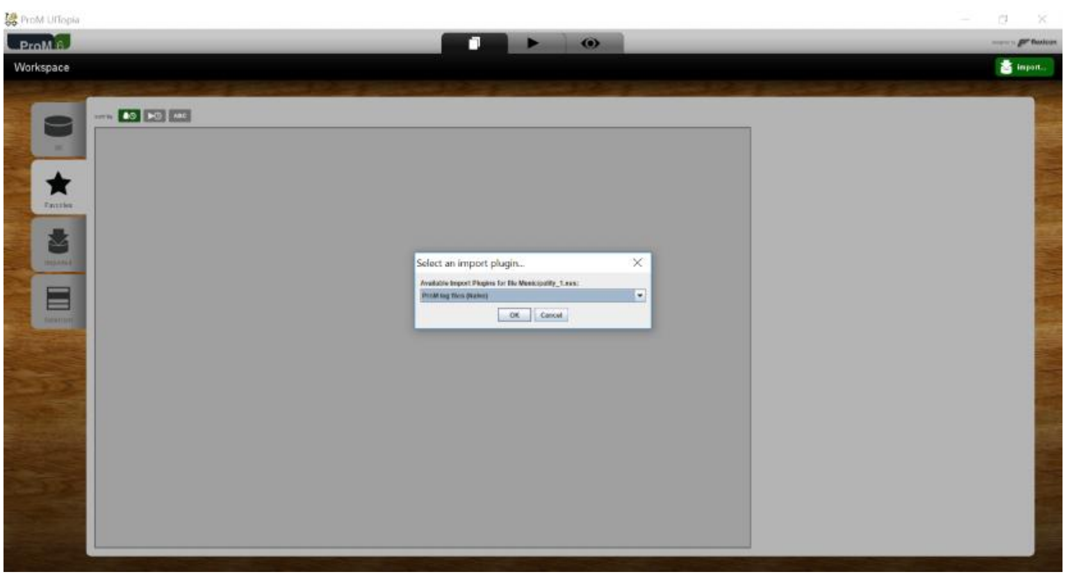

ProM UITopia
Petri net
PIP
Zoom
Export

Appendix E

Dutch Municipality Case Study Implementation in ProM

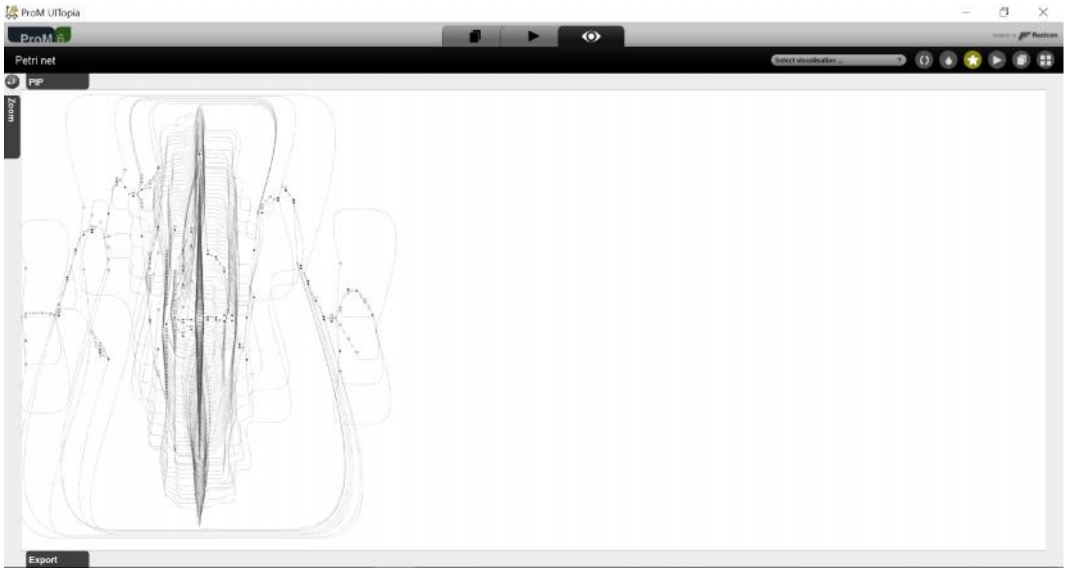
Process Discovery from BPI15_2.xes using Inductive Miner






ProM UITopia
Actions
Input
Output
Mine Petri net with Inductive Miner
Mine process tree with Inductive Miner
Mine stochastic Petri net from log
Mine with Inductive visual Miner
Visualise deviations on process tree
Petri net
Initial marking
Final marking
Reset
Start
Plugin action info
Mine Petri net with Inductive Miner


ProM UITopia
Activity
Tasks
Mine using Inductive Miner
Cancel
Finish


ProM UITopia
Petri net
PIP
Zoom
Export

Appendix F

In this exercise, please answer the questions below based on the analysis of the reviewing.xes business process. Please go to this link to download the event log for the exercise.

1. Event Log Overview

How many cases are in the reviewing process?

How many events are in the reviewing process?

How many events per case in average?

What is the maximum number of event per case?

What is the minimum number of event per case?

How many originators/resources recorded in the event log?

2. Process Discovery

Please explain briefly the process of reviewing based on the discovered process model:

3. Conformance and Performance Analysis

What is the percentage of log that conform to the discovered process model i.e., fitness?

Please list the event that needs to be concerned by the process owner e.g., events that cause bottleneck or took the longest time. Please includes the reason why:

4. Social Network Analysis

Please name the most important resource and explain why:

Please list the activities that are performed by the most important resource:

Please list the resources that are working with the most important resources. Please mention its causality degree to the most important resource:

5. Decision Mining

Please list the guards point in the process:

6. Recommendation

Please give an improvement recommendation for the reviewing:

Appendix G

UAT Questionnaire

General Question

What is your field of study?

How long have you been studying about process mining?

Are you familiar with the topic of data analytics?

Have you ever used data analytics techniques in your work or study?

Usefulness

ProM useful to the discovery of the activity ordering in the business process

Using ProM enables me to know about the relations between resources or actors in the business process

ProM helps me to know about the performance of the business process

Using ProM let me know about decision points in the business process

I can provide an improvement recommendation for the business process based on the analysis features provided by ProM

Ease of Use

My interaction with the ProM for future process analysis would be clear and understandable

It would be easy for me to become skillful at understanding the process mining techniques using ProM

I found ProM is easy to use

Intention

I intend to use the ProM for process analysis in the future

I predict i will use ProM in the future for process analysis

I plan to use ProM in the future for process analysis