UNIVERSITY OF TARTU
Institute of Computer Science
Conversion Master in IT Curriculum

Triin Kraavi

Data-Driven Process Analysis: Case Study of Labour Market Training Using Process Mining

Master's Thesis (15 ECTS)

Supervisor: Fredrik Milani

Tartu 2023

Data-Driven Process Analysis: Case Study of Labour Market Training Using Process Mining

Abstract:

Process mining is a data-driven technique that enables discovering, monitoring, and improving business processes. Process mining uses event log data stored in IT systems during business process execution. The objective of this Master Thesis was to conduct data-driven process analysis to identify improvement opportunities using process mining methodology. Using labour market training data, a case study was conducted in Estonian Unemployment Insurance Fund. Process Mining Project Methodology was used as a framework for this case study. Research questions were raised in the planning phase of the process mining project. The next phases were data extraction, processing, and event log creation. Apromore Enterprise Edition software was used for process mining. After the analysis and validation of results, suggestions for labour market training process improvement and further research were presented.

Keywords:

process mining, process analysis, labour market training, process mining project methodology

CERCS: P170, Computer science, numerical analysis, systems, control

Andmetel põhinev protsessianalüüs: Tööturukoolituse protsessikaeve juhtumiuuring

Lühikokkuvõte:

Protsessikaeve on andmetel põhinev meetodika äriprotsesside kuvamiseks, monitoorimiseks ja parendamiseks. Protsessikaeves kasutatakse sündmuste logi andmeid, mis salvestuvad infosüsteemidesse äriprotsessi läbiviimise käigus. Magistritöö eesmärk oli läbi viia andmetel põhinev protsessianalüüs kasutades protsessikaeve metoodikat. Protsessikaeve juhtumianalüüs viidi läbi Eesti Töötukassa tööturukoolituse andmete põhjal. Juhtumianalüüsi raamistikuks kasutati protsessikaeve projektimetoodikat. Protsessikaeve projekti planeerimise faasis püstitati uurimisküsimused. Sellele järgnes andmete väljavõtmine infosüsteemist, andmete töötlemine ja sündmuste logi loomine. Protsessikaeveks kasutati Apromore Enterprise Edition tarkvara. Tulemuste analüüsi ja valideerimise järel tehti ettepanekud tööturukoolituse protsessi parandamiseks ning edasisteks uurimisteemadeks.

Võtmesõnad:

protsessikaeve, protsessianalüüs, tööturukoolitus, protsessikaeve projektimetoodika

CERCS: P170, Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

Table of Contents

Table of Figures

Acknowledgement

I want to express my most profound appreciation to my supervisor Fredrik Milani, who has been a constant source of inspiration and motivation throughout this journey, from the lectures to this master thesis. I am incredibly grateful for the invaluable guidance, support, and feedback while writing this thesis.

I wish to thank Marlon Dumas and Apromore for providing access to Apromore Enterprise Edition, which enabled me to conduct the analysis needed for this thesis.

And finally, I would like to thank my family, friends, and colleagues from all of my heart for your love, patience and support throughout the years. Reaching this point would not have been possible without you.

1 Introduction

Process mining is a new discipline that academics have researched for over two decades. It has recently become increasingly important and used in different industries. Organisations have understood the benefit they can receive from process mining as it provides a unique view of the processes showing what is happening based on real-life data. Process mining uses data recorded in IT systems during process execution. Each process step is entered into an event log with a unique case ID, activity name, and timestamp. Process mining software visualises the process flows based on the event log data and enables detailed analysis of the result. Process mining tools and methods have been developed over the years. Initially, process mining used historical data to analyse the processes and find improvement opportunities. Now process mining can also be used during process execution to make decisions or by combining machine learning also predict possible future scenarios [1].

Process mining has been used in different industries – there is much academic literature about case studies in healthcare, education, financial auditing, supply chain etc. [1]–[3]. There are also examples of using process mining in the public sector [4]. However, the evidence using process mining in Public Employment Services[1] (PES) is limited. There are a couple of examples – process mining helped to develop a new workflow that enabled to process of a significantly increased number of invoices and to minimise the throughput time of the unemployment benefits handling process [5], [6]. This thesis aims to give a new perspective on process mining in PES by addressing the topic of the Active Labour Market Policies[2] (ALMP) provision process in PES.

The objective of this thesis was to conduct data-driven process analysis to identify improvement opportunities using process mining. Based on labour market training data, the case study was carried out in Estonian Unemployment Insurance Fund. The following research questions were formulated:

RQ1. What does the labour market training process look like using process mining?

RQ2. What kind of improvement opportunities are detected during process analysis?

RQ3. What are the differences between the processes of those who finish and those who quit labour market training?

This thesis focused on four process mining use cases to answer the research questions: process discovery, conformance checking, performance analysis and comparative process mining. Process Mining Project Methodology (PM^2) was used as a framework for this case study to carry out this process mining project and to answer the research questions. Six stages of PM^2 were executed – planning, data extraction, data processing, mining and analysis, evaluation and process improvement and support. Process mining was conducted using Apromore[3] Enterprise Edition software under an Academic license agreement. Process mining results provided new insights – understanding the structure and flexibility of the process and discovering the improvement opportunities and ideas for further research.

The remainder of the thesis has the following structure. Chapter 2 introduces the background of the Estonian Unemployment Fund, labour market training and process mining. The methodology used in this thesis is described in Chapter 3. Chapter 4 presents the results of the

[1] Public employment services are the authorities that connect jobseekers with employers [24]
[2] Active Labour Market Policies are publicly financed measures to help individuals enter the labour market or to prevent already employed individuals from losing their jobs [25]
[3] https://apromore.com/

process mining. Chapter 5 concludes the thesis by presenting a discussion, limitations, and conclusion.

2 Background

Chapter 2 explains the background of the Estonian Unemployment Insurance Fund, labour market training and process mining.

2.1 Estonian Unemployment Insurance Fund

Estonian Unemployment Insurance Fund (EUIF) is an independent public body founded in 2001 to administer unemployment insurance benefits.

EUIF took over the Public Employment Service (PES) responsibilities in Estonia in 2009 and started to provide work-focused counselling and other Active Labour Market Policies (ALMP) to unemployed persons and employers. EUIF received new tasks with Work Ability Reform that began in 2016. EUIF started to assess work ability in cooperation with healthcare service providers, pay work ability allowance and provide specific ALMPs for people with reduced work ability. Unemployment prevention measures for employers and employed persons were added to EUIF's services in 2017. Since 2019 EUIF has been the sole provider of career services in Estonia to all target groups, from school children to employed, unemployed and retired persons[4].

EUIF has tripartite management - the Supervisory Board consists of equal representatives from the government, employers, and employees. Everyday work is directed by four-member Management Board [7].

EUIF has 15 Regional Offices that provide counselling and ALMPs to its clients. Head Office is responsible for ALMP development and centralised payment of benefits and allowances, also providing support (e.g. HR, IT etc.) functions. The number of employees has grown from 20 to almost 1000.

EUIF's mission is to support finding work and employees. Its vision is to be an innovative and efficient organisation that provides the best solutions according to the needs of labour market stakeholders and whose opinions have weight in Estonia and Europe [8].

The development plan states the strategic objectives of EUIF [8]: (1) reducing unemployment and minimising the duration of unemployment, (2) preventing unemployment and supporting employment, (3) increasing labour supply, and (4) developing EUIF as a competence centre. To reach these strategic objectives, EUIF offers various ALMPs to overcome the obstacles that clients have to support entering or staying in the labour market: job search counselling, job mediation, career counselling, labour market training, apprenticeship, work practice, wage subsidy, business start-up subsidy, job club, psychological, debt and addiction counselling, work-related rehabilitation, sheltered employment and working with a support person. In addition, EUIF assesses clients' work ability and pays benefits and allowances: unemployment insurance benefit, redundancy benefit, insolvency benefit, unemployment allowance, workability allowance, and reimbursement of social tax for the employee with a reduced work ability.

EUIF processes are digital and automated as much as possible. Operational processes are supported by 12 information systems that are interlinked and exchange data with 29 different external registries via data exchange layer X-tee[5]. EUIF has fully automated processes

[4] Aas G. Strategic Management of EUIF. Presentation. (04.04.2023)

[5] X-tee is a technological and organisational environment that enables a secure Internet-based data exchange between information systems [26]

for unemployment registration and unemployment insurance benefits application – if all of the necessary data is available in the registries, then the client will receive an automatic decision made by the IT system in seconds.

EMPIS is one of EUIF's central Customer Relation Management systems that is used for maintaining the register of the unemployed, client management (data, contacts, appointments, notifications), ALMP provision management and payment management.

2.2 Labour Market Training

66% of EUIF's clients either have insufficient Estonian language or computer skills or lack professional education. EUIF offers various ALMPs related to skills development to clients: labour market training for the unemployed, reimbursement of the costs of a certificate or attestation, degree study allowance, coaching, labour market training for the employed and training grants for employers. This thesis focuses on labour market training for the unemployed, specifically procured labour market training.

Labour Market Services and Benefits Act states [9] that the goal of labour market training is to obtain or develop professional or other skills which facilitate employment. Labour market training lasts for a period of up to one year.

Labour market training for the unemployed is provided to those seeking work or registered with the EUIF as unemployed, a jobseeker with a notice of dismissal or a jobseeker of retirement age who is not employed [10].

Labour Market Training for the unemployed is one of the highest volumes ALMPs that EUIF provides for its clients, with around 30 000 entrances to service during the last years (see Figure 1). The costs of labour market training for the unemployed were over 25,5 million euros in 2022, which accounted for 36% of the total costs of ALMPs.

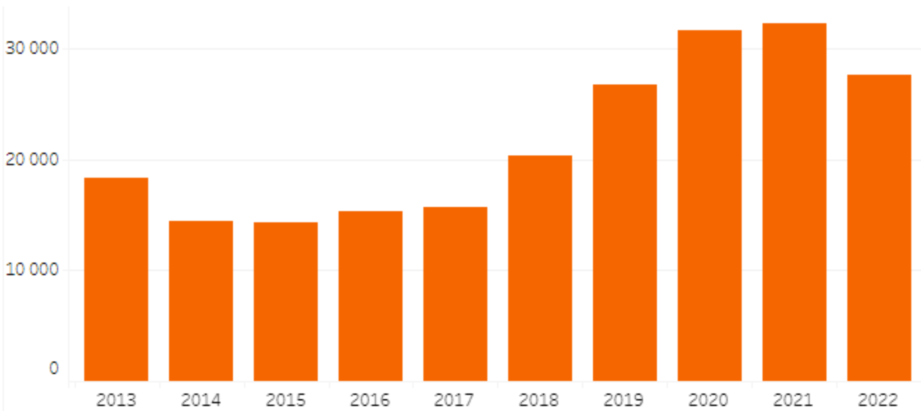

Figure 1. Entrances to labour market training for unemployed 2013-2022
Source: EUIF

Labour market training is provided based on the client's individual and employer's labour needs, by procuring the necessary training or by allowing the client to choose a training that qualified training card partners offer in the open market with a training card[6].

Around one-quarter of labour market trainings for unemployed are procured trainings, and three-quarters are provided by training card. EUIF procures trainings that have continuous high demand – language skills (primarily Estonian language), computer skills (mostly basic computer skills) and entrepreneurship training for clients who want to start their own business. Training provided by training card has a wider variety and corresponds to more specific clients' and labour market needs.

This thesis focuses on procured labour market training process. The process of procured labour market training starts with assessing the training need by the client and EUIF counsellor. After the training need is assessed and agreed upon, the EUIF specialist decides how the training is provided – either procured training or a training card. When procured training is confirmed and the training is procured if necessary, the EUIF service consultant registers the client for the training and the client starts the training. The client is entitled to a grant and commuting and accommodation benefit for attending the training. After the training, the result of the client receiving the necessary skills and knowledge is discussed during counselling sessions. In addition, clients can give feedback to EUIF in the ALMP satisfaction survey.

2.3 Process Mining

Everything an organisation does can be described as business processes that turn inputs into outputs to create value for internal or external customers. Dumas *et al.* [11:6] have defined business processes as "a collection of inter-related events, activities, and decision points that involve several actors and objects, which collectively lead to an outcome that is of value to at least one customer". Organisations continuously work on their business processes to get the maximum results (e.g., faster processes, using fewer resources, creating fewer errors, etc.). Business Process Management (BPM) is an approach used to identify, discover, analyse, redesign, execute and monitor business processes to optimise them [11]. Different process analysis and monitoring methods (e.g. flow and queue analysis, process simulation, process performance dashboards and process mining) all use data to identify patterns, trends, and insights that can be used to improve business processes.

Process mining is a relatively new discipline developed in the late 1990s by Wil van der Aalst [1]. Process mining is a data-driven technique that enables discovering, monitoring, and improving business processes [12]. Process mining bridges data science and process science because it connects traditional model-based process analysis and data-centric analysis techniques (see Figure 2) [13].

[6] Estonian Unemployment Insurance Fund. Rules of procedure. Labour market training (valid from 01.01.2022)

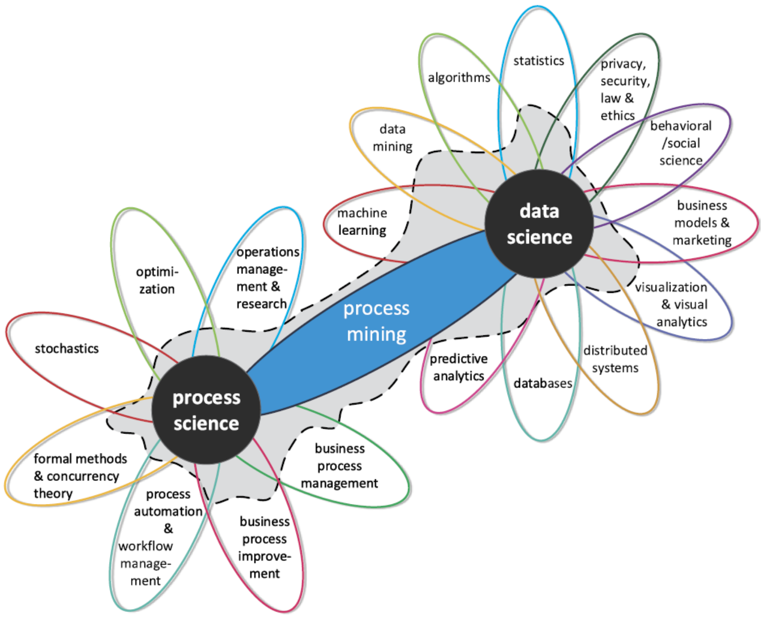

Figure 2. Process mining as the bridge between data science and process science
Source [13]

The prerequisite of process mining is an event log. The event log is created based on the data stored in IT systems during business process execution. Each row of the event log must have a case ID that is used to distinguish different cases, activity name and the timestamp when the activity was performed. The event log can have additional attributes that widen the analysis possibilities (see Table 1). For example, the start and the end timestamps allow analysing the length of each activity performed, and information about the resources allows seeing if there are bottlenecks or ping-pong transfer of tasks between resources etc.

Table 1. Example of an event log

CASE_ID	ACTIVITY	END_TIME	ACTOR	STATUS
2310388	Training need added	01.01.2020 01:00	Counsellor	Deregistration
2310388	Training need evaluated	01.01.2020 02:00	Counsellor	Deregistration
2310388	Training need approved	07.01.2020 03:00	Service consultant	Deregistration
2310388	Cancelling training need due to deregistration	28.06.2020 14:00	System	Deregistration
2310395	Training need added	01.01.2020 01:00	Counsellor	Finished
2310395	Training need evaluated	01.01.2020 02:00	Counsellor	Finished
2310395	Training need approved	01.01.2020 03:00	Service consultant	Finished
2310395	Adding clients to training agreement	06.01.2020 06:00	Service consultant	Finished
2310395	Start of training agreement	19.01.2020 07:00	System	Finished
2310395	End of training	13.02.2020 00:00	System	Finished

Manually created process models are input in the process mining analysis stage.

Numerous process mining software is available (e.g. Apromore, ProM[7], Celonis[8], Disco[9] etc.) [14], [15]. The software inputs the event log and visualises the results as process flows and BPMN[10] models. Filtering, dashboards, simulations, and other features allow detailed process analysis from different angles.

Process mining has different use cases [1]:

1. **Process discovery**
   During process discovery, a process model is created automatically based on the information in the event log.
2. **Conformance checking**
   Conformance checking takes the event log and process model as input, bringing out their differences.
3. **Performance analysis**
   Performance analysis uses an event log and process model as input and uncovers the performance problems in the process – for example, long waiting times, bottlenecks, rework etc.
4. **Comparative Process Mining**
   Comparative process mining takes as an input multiple event logs (for example, from different locations, periods etc.) and shows the differences and commonalities between them and possible factors influencing that.
5. **Predictive Process Mining**
   Predictive process mining combines process mining results with machine learning to predict possible future scenarios.
6. **Action-Oriented Process Mining**
   Action-oriented process mining uses process mining results to make decisions and actions during process execution (for example, if some activity is skipped, the manager is notified) or to discover possibilities for process automation.

Process mining can show what has happened, what is happening now and what will happen next in the business process [1]. This thesis focuses on analysing what has happened in the past, so process discovery, conformance checking, performance analysis and comparative process mining are examined.

[7] https://promtools.org/
[8] https://www.celonis.com/
[9] https://fluxicon.com/disco/
[10] Business Process Model and Notation (BPMN) is a standard visual modeling language for business processes https://www.bpmn.org/

3 Methodology

Chapter 3 describes the methodology used in this thesis.

3.1 Case Study Methodology

A case study is a methodology commonly used in social science research. Case study methodology enables an in-depth exploration of complex phenomena within some specific context [16].

According to Farquhar J.D. [17] a case study in business research means that the phenomenon is examined, and evidence is collected where it occurs – like in a company or an organisation. A case study is usually searching for answers to questions that start with how, who, and why. Different sources of data or different methods are used to collect data in case studies – for example, interviews, internal documentation, reports etc. This ensures that the findings from the case study are strengthened as the evidence is triangulated (investigated from different perspectives).

The case study methodology applies to this thesis because it involves an in-depth investigation of a specific process (labour market training) in a particular context (EUIF). Multiple sources of data (data from IT systems, internal documentation, interviews, and academic literature) are used in this thesis to gain an understanding of the labour market training process, its variants, and to identify opportunities for improvement. The specific process mining methodology that is suitable to carry out this case study is explained next.

3.2 Process Mining Project Methodology

There are several specific methodologies developed over the years to carry out process mining projects – for example, the L * lifecycle model [13], Process mining project methodology (PMPM) [18], Process Mining Project Methodology (PM^2) [19] and Methodological proposal for process mining projects [20].

Process Mining Project Methodology (PM^2) was chosen as the state-of-the-art framework for the labour market training process mining case study. Evidence in the academic literature shows that PM^2 has been successfully applied in several process mining case studies like [19], [21].

PM^2 consists of six stages (see Figure 3) [19]:

1. **Planning**
   The objective of the planning stage is to set up the project and determine the research questions. The first task is to select the business process for process mining. After that, preliminary research questions are identified, and the process mining project team is assembled.
2. **Extraction**
   The objective of the extraction stage is to extract process data. Extraction starts with determining the scope of the data extraction. After that, the data is extracted according to the scope. The third activity is transferring process knowledge, either studying the documentation or process models available or exchanging the knowledge between the project team and domain experts.

3. **Data Processing**
   The objective of the data processing stage is to create event logs. The activities in this stage are creating views, aggregating events, enriching, and filtering logs.
4. **Mining and Analysis**
   The objective of the mining and analysis stage is to apply process mining techniques to answer the research questions. Possible activities in this stage are process discovery, conformance checking, enhancement and process analytics.
5. **Evaluation**
   The objective of the evaluation stage is to relate the findings from the previous stage to improvement ideas that achieve the project's goals. The activities in this stage are first to understand and distinguish interesting findings and refine research questions and, secondly, to verify and validate to ensure the findings are correct.
6. **Process improvement and support**
   The objective of the process improvement and support stage is to use the results of process mining to improve the processes. The activities in this stage are improvement implementation and operational support.

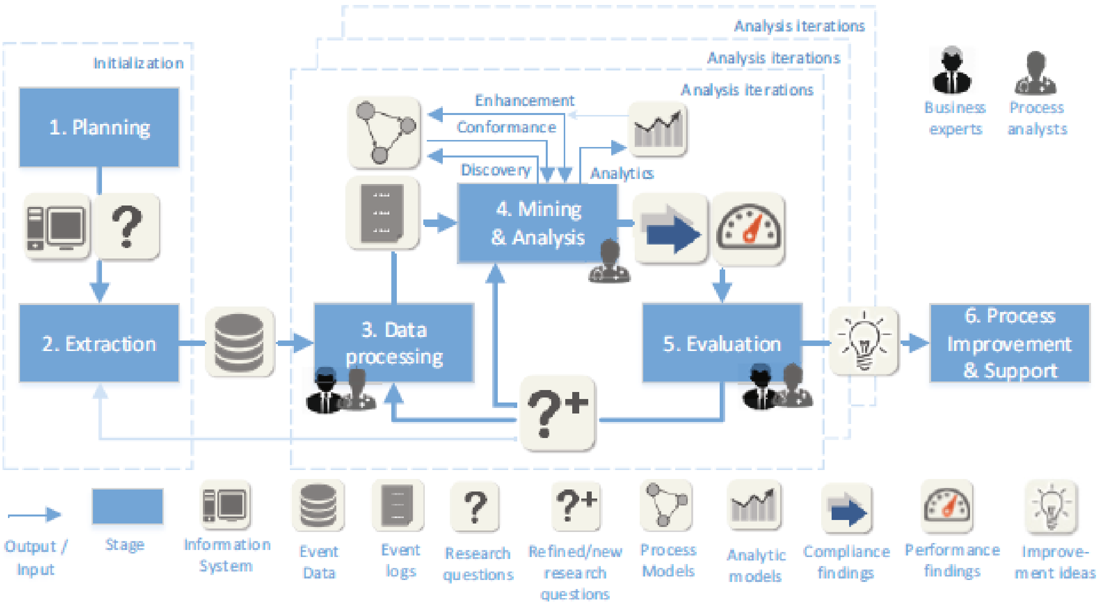

Figure 3. Overview of the PM^2 methodology
Source [19]

Slight changes were made in the PM^2 methodology for this case study:

1. "Composing project team" was skipped in the Planning stage as this was a one-person project conducted by the author.
2. "Aggregating events" in the Data processing stage was not applicable.
3. There are only improvement suggestions in the Process improvement and support stage, as no improvements have not implemented yet.

3.3 Planning

Chapter 3.3 corresponds to step "1. Planning" in Figure 3.

3.3.1 Selecting Business Processes

As this case study aimed to give value to EUIF, different departments were approached to get input on which process could be analysed using process mining. As a result, procured labour market training process was selected as this process is planned to be reviewed and renewed in the nearest future, and the results of process mining can be used as one of the inputs. There is also a problem that many clients (25% in 2022) quit and do not finish procured labour market training.

EUIF's research study commission granted the permission to carry out this case study and use EUIF's data necessary for this process mining project.

3.3.2 Identifying Research Questions

Based on the input from domain experts, three research questions were posed:

> **RQ1.** What does the procured labour market training process look like using process mining?
>
> **RQ2.** What kind of improvement opportunities are detected during process analysis?
>
> **RQ3.** What are the differences between the processes of those who finish and those who quit procured labour market training?

As this is the first time a process mining project is carried out on procured labour market training, the results of RQ1 set the scene and create an understanding of what the process looks like using real-life data. In addition, the results will be used as input for RQ2 and RQ3. Procured labour market training process is planned to be reviewed and improved in EUIF. Therefore, the results of RQ2 will give additional information for that. RQ3 addresses the problem that procured labour market training has for many years – many clients quit the training. Comparing the processes of finished and quit training might give insights that can be used to tackle this problem and improve the effectiveness of labour market training.

These research questions allow us to implement four out of six process mining techniques:

- RQ1 - process discovery.
- RQ2 - conformance checking and performance analysis.
- RQ 3 - comparative process mining.

Predictive process mining and action-oriented process mining are out of the scope of this case study.

3.4 Extraction

Chapter 3.4 corresponds to step "2. Extraction" in Figure 3.

3.4.1 Determining Scope

The search for necessary data for process mining started from test-EMPIS database schemas. Test-EMPIS has the same functionality as EMPIS, but the data is masked out. Data masking substitutes realistic but false data for original data so testing, training, development, or support teams can work with a dataset without putting real data at risk [22]. DBeaver[11] software was used to access test-EMPIS database tables.

[11] DBeaver is free and open-source universal database tool. https://dbeaver.io/

Seven tables were identified where the necessary data was located:

1. EMP_MAKSE
2. ISIK_TEENUS
3. KOKKULEPE
4. KOOLITUSVAJADUS
5. LEPING
6. OSAVOTUGRAAFIK
7. TEENUS_VAJADUS

All the data attributes from the tables were extracted, limiting the time starting from 2020. This proved beneficial as missing data in the event log was discovered in the project's next steps, and additional data was extracted from already available tables with necessary attributes.

3.4.2 Extracting Event Data

After access to live EMPIS database schemas was granted, Talend[12] software was used to extract data from the live EMPIS database to the local database.

A separate job for data extraction was created for each table. Figure 4 shows an example of a Talend job.

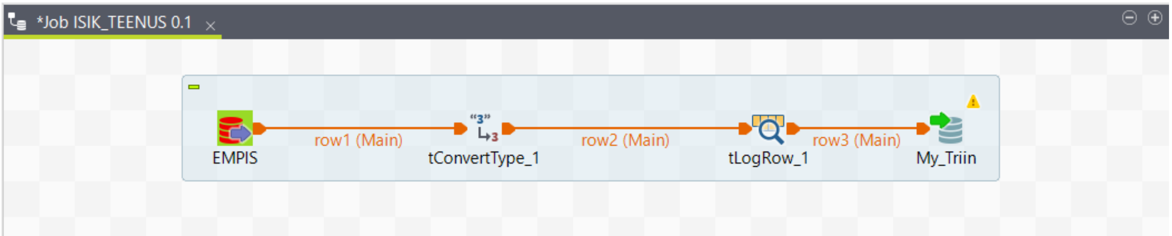

Figure 4. Visualisation of Talend job extracting data from ISIK_TEENUS table

Extracted tables varied in size significantly – the smallest table had 3 349 rows, largest table 1 496 891 rows. The number of table attributes also varied – from 20 to 73. As the amount of data was very large and running jobs took much time, originally used SQLite[13] was replaced with MySQL[14] to enhance performance.

3.4.3 Transferring Process Knowledge

EMPIS user manual for labour market training[15] was used to understand the process steps and manually create a more detailed BPMN process model of labour market training. Three one-hour meetings were conducted with two labour market training chief specialists to validate and update the process model and process steps before and during the data extraction stage.

[12] Talend is an open-source data integration tool. https://www.talend.com/
[13] SQLite is SQL database engine. https://www.sqlite.org/index.html
[14] MySQL is a relational database management system. https://www.mysql.com/
[15] Estonian Unemployment Insurance Fund. EMPIS user manual. Labour market training. (24.04.2015)

3.5 Data Processing

Chapter 3.5 corresponds to step “3. Data Processing” in Figure 3.

3.5.1 Creating Views

Example cases were checked manually before creating the event log to see if the process step data results were logical and corresponded to the information gained from the before-mentioned analysis. It was discovered that due to unclear table attribute names, two process steps (“Training need evaluated” and “Training need approved”) were incorrectly identified and in the wrong order. As a result, the name of an activity and process step order was corrected.

A SQL[16] script was written to create a new EVENT_LOG table with the columns CASE_ID, ACTIVITY, END_TIME and ACTOR. A separate SQL script was created to add each process step to the EVENT_LOG table (see Figure 5).

```
INSERT INTO EVENT_LOG (CASE_ID, ACTIVITY, END_TIME, ACTOR)
SELECT
     tv.ID AS CASE_ID,
    'Training need added' AS ACTIVITY,
    tv.LISAMISE_KP AS END_TIME,
    'Counsellor' as ACTOR
FROM teenus_vajadus tv
INNER JOIN koolitusvajadus k ON tv.ID =k.TEENUS_VAJADUS_ID
WHERE tv.TEENUS_KOOD = 'KOOLITUS' AND
k.KOOLITUS_VIIS_KOOD = 'TELLITUD_KOOLITUSED'
;
```

Figure 5. Example of SQL script for process step “Training need added”

Training need ID (tv.ID) was used in each script as a case identifier, and different tables were joined to extract only procured labour market training data. Indexes were added to table columns to ease and quicken the joining of the tables.

3.5.2 Enriching Logs

Adding ACTOR column

As hundreds of employees are implementing the procured labour market training process in EUIF, using unique resource identifiers for each was not feasible. Instead, an ACTOR column was created where the role of the process step implementor (e.g. counsellor, client, service consultant etc.) was identified.

Modifying the timestamp

Unfortunately, EMPIS does not record the exact time when the activity happens, and all activities are recorded to occur at midnight (00:00:00). This causes problems in process mining when activities occur on the same date – activities are shown as parallel activities

[16] SQL is a standard language for accessing and manipulating databases.

instead of consecutive activities. It was necessary to manipulate the time for each process step to overcome this problem and present the results logically (see Figure 6).

```
UPDATE EVENT_LOG
SET END_TIME = DATE_FORMAT(END_TIME, '%Y-%m-%d 01:00:00')
where ACTIVITY = 'Training need added'
;

UPDATE EVENT_LOG
SET END_TIME = DATE_FORMAT(END_TIME, '%Y-%m-%d 02:00:00')
where ACTIVITY = 'Training need evaluated'
;
```

Figure 6. Example of the timestamp modification

The order of the activities was based on the manually created process model created during the data extraction stage.

Adding STATUS column

An additional column, STATUS, was added to make variant analysis easier for comparing the processes of those clients who quit and finished the labour market training (see Figure 7).

```
update EVENT_LOG el
inner join isik_teenus_koolitus itk ON el.CASE_ID = itk.TEENUS_VAJADUS_ID
set STATUS = 'Finished'
where itk.SUUNAMISE_TULEMUS_KOOD  = 'LOPETAS'
;

update EVENT_LOG el
inner join isik_teenus_koolitus itk on el.CASE_ID =itk.TEENUS_VAJADUS_ID
set STATUS = 'Quitted'
where itk.SUUNAMISE_TULEMUS_KOOD  = 'KATKESTAS'
;
```

Figure 7. Example of adding a STATUS column

Adding activity "Cancelling training need"

During process mining analysis, an anomaly was detected – there were thousands of cases where training need was evaluated and approved, but the clients had not received the training. After analysis, it was discovered that there is one activity, "Cancelling training need", missing from the EVENT_LOG table. Missing activity data was extracted from the EMPIS tables and added to the EVENT_LOG table and process mining analysis was done over again.

3.5.3 Filtering Logs

Filtering logs were carried out multiple times during the case study analysis. For example, two logs (who finished and quit the training) were filtered out during comparative process analysis. To make a comparison of two variants more comparable, activities related to participation sheets and client payments were filtered out.

Two activities extracted from different tables duplicated the same information – "Start of training agreement" and "Start of training". As a result, the "Start of training" activity was filtered out in the analysis.

3.6 Mining and Analysis

Chapter 3.6 corresponds to step "4. Mining and Analysis" in Figure 3.

Process mining was conducted using Apromore Enterprise Edition under an Academic license agreement. It was chosen because of our previous experience with the software during academic studies, and it was possible to use the Apromore Enterprise Edition free of charge for education and research purposes.

Event log data was inserted into Apromore process mining software for automatic process discovery. The results of automatic process discovery were compared with the manually created labour market training process model during conformance checking. Business rules fulfilment was checked during performance analysis. In addition, analysis templates for identifying improvement opportunities developed by Danylyshyn N. [23] were used. When using analysis templates, the following steps were performed:

1. Checking the minimum data needed for using the template.
2. If existing event log data corresponded to the minimum data needed, guidelines on identifying the improvement opportunity were followed step by step.
3. Output was described.
4. Results and redesign possibilities were analysed.

As the event log of this case study contains only the end time and does not contain resource information, only part of the templates was applicable:

- Ping-pong behaviour
- Rework
- Knock-out
- Highest waiting time in the business process
- Cases with the highest overall waiting times

Two event logs were created to conduct comparative process mining – cases where the training was finished and cases where the training was quit. BPMN models and active cases of these two event logs were compared over time. Also, the analysis based on the length of the training was carried out.

The results of the mining and analysis stage are presented in Chapter 4 Results.

3.7 Evaluation

Chapter 3.7 corresponds to step "5. Evaluation" in Figure 3.

Evaluation of the results was conducted first by the author. Based on that, missing data was discovered, and as a result, the process moved back to the data extraction and processing stage to correct the event log. After that, the mining and analysis were performed again. Results were presented and discussed with domain experts (skills development services' development manager, labour market training service manager and two labour market training chief specialists) during the 1,5 h meeting.

The results of the evaluation are presented in Chapter 5 Discussion.

3.8 Process Improvement and Support

Chapter 3.8 corresponds to step "6. Process improvement and support" in Figure 3.

Several process improvements and further research suggestions were made based on the analysis described in Chapter 4 and the evaluation discussion with domain experts. There has not been enough time to implement any improvement suggestions yet.

The process improvement and further research suggestions are presented in Chapter 5 Discussion.

3.9 Using artificial intelligence assistance

Artificial intelligence tool Grammarly[17] was used to proofreading this thesis.

[17] https://app.grammarly.com/

4 Results

Chapter 4 presents process mining results and analysis. RQ1 is analysed in Chapter 4.1 Process Discovery. RQ2 is analysed in Chapter 4.2 Conformance Checking and Chapter 4.3 Performance Analysis. RQ3 is explored in Chapter 4.4 Comparative Process Mining.

4.1 Process Discovery

Final EVENT_LOG table had 553 037 rows with a total of 44 042 cases and 18 different process steps:

1. Training need added.
2. Training need evaluated.
3. Training need approved.
4. Training agreement approved.
5. Training agreement signed.
6. Adding clients to training agreement.
7. Start of training agreement.
8. Receiving participation sheet.
9. Registering participation sheet.
10. Payment of commuting and accommodation benefit.
11. Payment of grant.
12. Quitting participation in training.
13. Cancelling training need.
14. Cancelling training need due to deregistration.
15. Client does not start training.
16. End of training.
17. End of training: cancelling agreement.
18. End of training: early completion.

The process had 4 827 different case variants. The ten most frequent case variants cover 51,7% of all cases (22 778 cases). 2 807 least frequent variants have each only one case. A large number of case variants indicates that there is much flexibility in the real-life process. Process design allows to perform different activities in a different order (e.g. training agreement can be signed already before the client's training need is approved) or there are a different number of repetitive activities in different cases (e.g. depending on the length of the training there can be 1 to 10 participation sheets) that also add variability of case variants. However, some cases have an illogical sequence of activities (e.g. training need is cancelled before it is evaluated and approved) and could be a task for further analysis.

The least frequent case variants were filtered out for automatic process discovery, leaving 41 235 cases with 2 020 case variants. Each node (activity) is presented as a rectangle. Arcs between nodes represent the relationships between them.

Figure 8 shows the automatically discovered procured labour market training process with 100% nodes and 10% arcs. The process starts with three activities related to the training need ("Training need added", "Training need evaluated", "Training need approved"). Then the process diverges in three directions. Two activities ("Cancelling training need" and "Cancelling training need due to deregistration") lead straight to the end of the process. The third branch leads to activities related to training agreements ("Training agreement approved", "Training agreement signed", "Adding clients to the training agreement", "Start of training agreement”). After that, the process diverges again in three directions. Two activities (“Client does not start the training” and “End of the training: cancelling agreement”) lead straight to the end of the process. The third activity (“Quitting participation in training”) leads to activities related to participation sheets and payments (“Receiving participation sheet”, “Registering participation sheet”, “Payment of commuting and accommodation benefit”, “Payment of grant”). At the end of the figure are two activities indicating the end of training (“End of training”, “End of training: early completion”). Arcs lead back from the end of training activities to the “Receiving the participation sheet” activity because the participation sheet is sent to EUIF each month, and payments are paid according to that. That is why “Payment of grant” is shown in the figure as one of the final activities of the process.

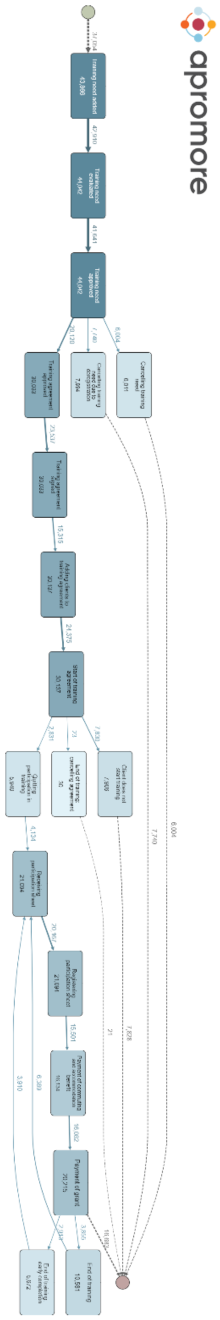

Figure 8. Labour market training process – 100% nodes, 10% arcs

When the percentage of arcs is increased, the repeating process steps and the flexibility of the process become more visible. Figure 9 shows that, in reality, the order of the activities is not as straightforward as presented in Figure 8. For example, the training agreement may have been approved and signed before the clients' training need was added, evaluated, and approved. The ping-pong behaviour between activities "Adding clients to training agreement" and "Training agreement signed" is also visible. This is caused by the fact that if the training agreement has been approved, it is possible to add clients to the training agreement before or after it has been signed. By increasing the percentage of arcs, repetitive activities become visible (activities related to the participation sheets and payments performed each month).

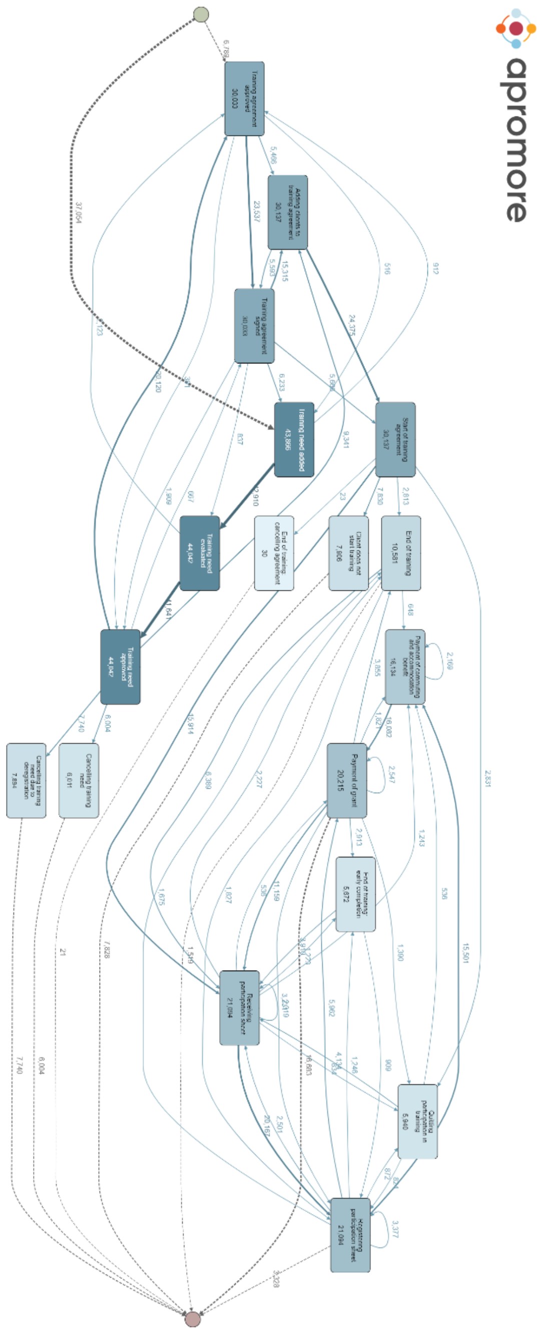

Figure 9. Labour market training process – 100% nodes, 50% arcs

By increasing the percentage of arcs to 100%, the automatically discovered process map becomes quite fuzzy and not easily comprehensible (see Figure 10). However, it enables us to detect some abnormal activities that could be further investigated and what did not appear in Figures 8 and 9. For example, there are 6 cases where the activity “Cancelling training need” is followed by the activity “Training need approved”. There are also cases where the activity “Cancelling training need due to deregistration” is before activities “Training need added” or “Training need approved”. This last anomaly is explained by the rule that if the client is not eligible to be registered as unemployed anymore, the deregistration is fixed by the date when the right to be registered as unemployed was not valid anymore. This deregistration date can also be in the past.

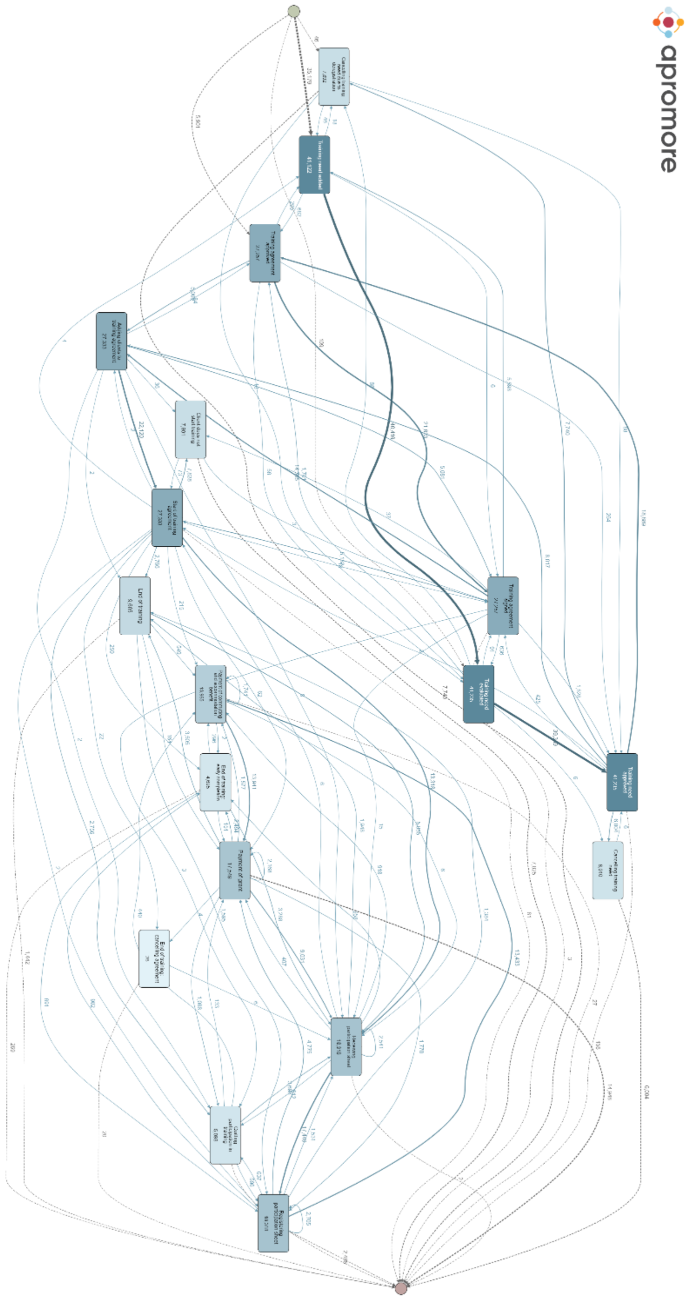

Figure 10. Labour market training process – 100% nodes, 100% arcs

The average case duration is 3,79 months, the median is 3,2 months, the minimum is 12 hours, and the maximum is 2,87 years.

Figure 11 shows the results analysing the status of the cases.

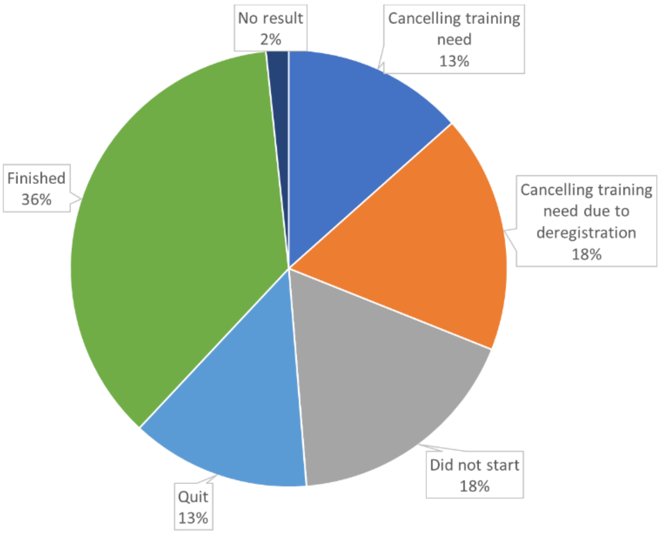

Figure 11. Distribution of cases based on status

31% of training needs are cancelled (13% “Cancelling training need”, 18% “Cancelling training need due to deregistration”). 18% of clients are added to the training but do not start the training. 50% of the cases reach the training, of which 73% finish and 27% quit. 2% of training needs are open and have no result.

4.2 Conformance Checking

Process discovery results (see Figure 8) and manually created process model (see Figure 12) were used for conformance checking.

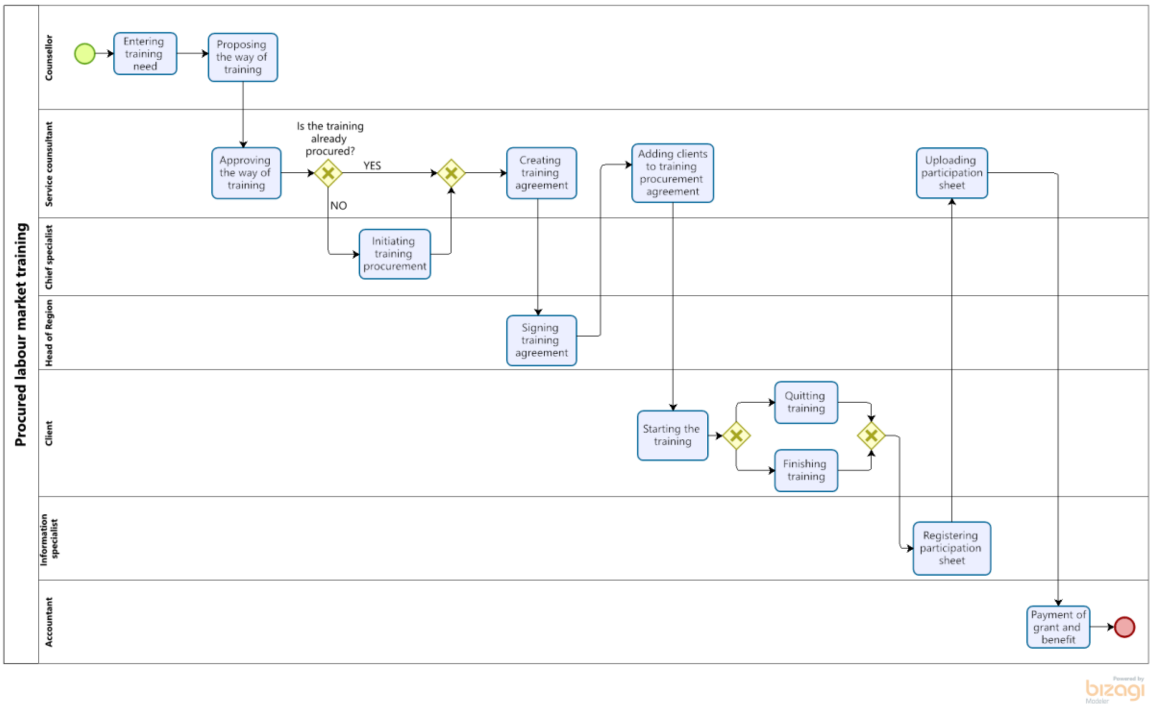

Figure 12. Manually created procured labour market training process model

The manually created process model is similar to the discovered process model using process mining. The manual process model has 13 activities, discovered process model has 18 activities.

As the manually created process model describes the process flow where clients start the training, some of the activities ("Cancelling training need", "Cancelling training need due to deregistration", and "Client does not start training") are missing compared with discovered process model (see Figure 13).

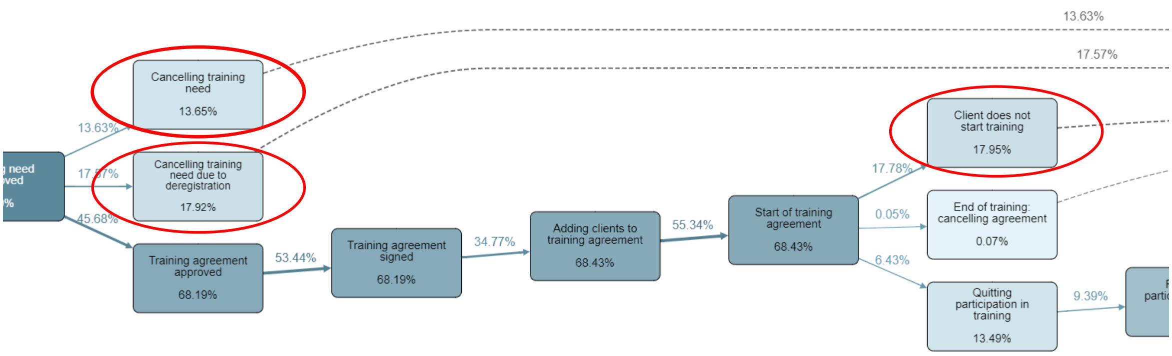

Figure 13. Activities missing in the manual process model (detail)

As the relative case frequency of these activities is 14% to 18%, these activities could be added to the manual process model to give an adequate picture of the process.

The manually created process model has an activity “Initiating training procurement”, missing from the automatically discovered model (see Figure 14). This activity was decided to exclude from the process mining because procured training contracts are usually signed for 1,5-3 years. This would have caused disorder in automatic process discovery visualisation when some of the activities have happened so far in the past.

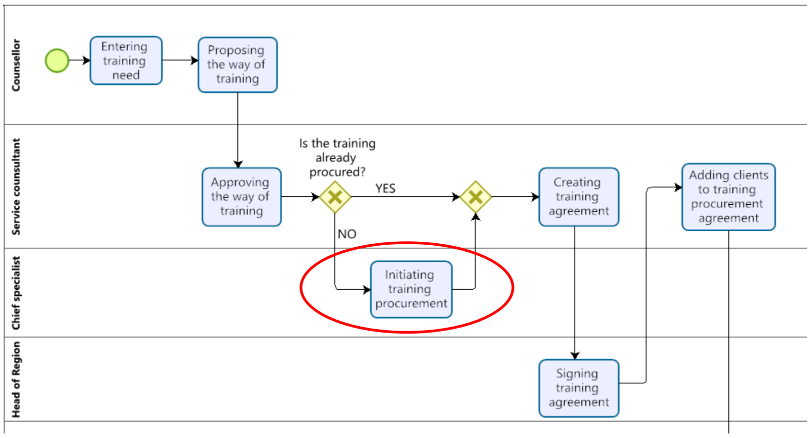

Figure 14. Activity missing in the discovered process model (detail)

In addition, some activities’ detail level is greater in the discovered process model, and the manual process model has aggregated activities:

1. “Payment of grant” and “Payment of commuting and accommodation benefit” are separate activities in the discovered process model. In the manual process model, they are aggregated into one activity.
2. Discovered process model has three different types of ends of training. The manual process model has aggregated them into one activity.

Aggregated activities in the manual process model are justified as it keeps the model as simple as possible.

4.3 Performance Analysis

Training that lasts over a year

According to Labour Market Services and Benefits Act [9], labour market training lasts up to one year. Eleven cases were detected in process mining where the length of the training from the beginning of the training agreement until the end is more than one year (see Table 2). These cases need further analysis.

Table 2. List of cases where training lasts over a year

Case ID	Activity instances	Duration	Case variant ID
2581115	21	1.95 yrs	2
2645662	17	1.37 yrs	1
2663813	21	1.49 yrs	2
2665446	17	1.26 yrs	1
2687435	17	1.14 yrs	1
2689309	17	1.12 yrs	1
2693254	17	1.10 yrs	1
2693291	17	1.10 yrs	1
2700706	21	1.28 yrs	5
2701773	21	1.27 yrs	4
2703632	23	1.26 yrs	3

Ping-pong behaviour

Ping-pong behaviour occurs when a case circulates from one activity to another. Some ping-pong behaviours are detected in the process mining (see Figure 15), but it is due to the flexibility of the process – it is allowed to carry out some process steps in random order. For example, when the training agreement is approved, it is possible to add clients to it before or after the Head of the Region has signed the agreement. However, during process discovery, an anomaly was detected that needs further analysis - there are 6 cases where the activity "Cancelling training need" is before the activity "Training need approved".

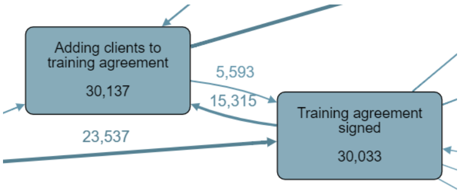

Figure 15. Example of ping-pong behaviour in the process (detail)

Rework

Rework is a situation where one activity is repeated several times. Rework can be seen with activities related to the participation sheets and payments (see Figure 16). That is also part of the normal process because during the period of training, each month new participation sheet is sent to EUIF, and monthly payments to clients are made based on that.

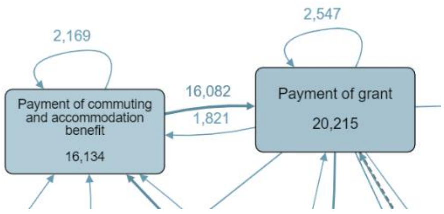

Figure 16. Example of rework in the process (detail)

Knock-out

Knock-out is a situation where an activity that is not the final activity leads directly to the end, skipping one or more activities in the process. There are 747 cases where the activity "Training need approved" and 8 cases where the activity "Start of training agreement" leads straight to the end (see Figure 17).

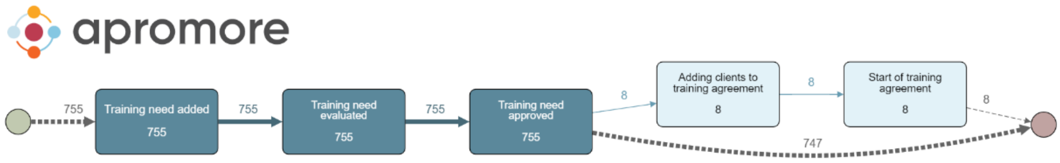

Figure 17. Example of knock-outs in the process

Most of the "Training need approved" cases are from the second half of 2022, which is expected because training has not started, and clients have not been added to the agreement yet (see Figure 18). But 53 cases need further analysis from January 2020 – June 2022 where the training need has been detected and approved, but the client has not received training.

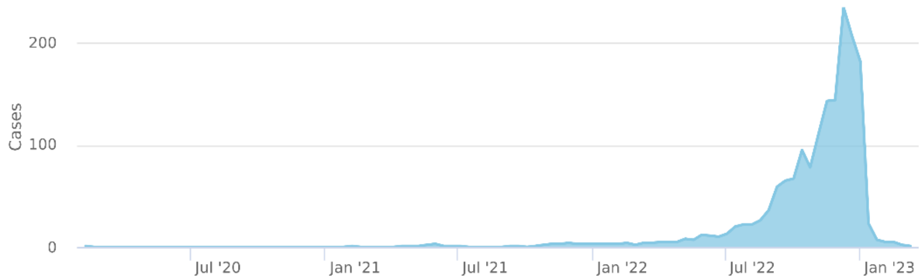

Figure 18. Active cases over time of knock-out "Training need approved"

"Start of training agreement" cases need further analysis if data is missing in the event log because all eight trainings started in 2020.

Highest waiting time in the business process

Highest waiting time in the business process identifies the highest average waiting time between two activities. The highest average waiting time (1,59 months) is between activities "Training need approved" and "Training agreement approved" (see Figure 19). Considering that the training has a fixed time when they start and sometimes procurement also takes time, the average waiting time of 1,59 months between approving training need and training agreement is not a critical issue and does not need improvement.

Figure 19. Highest waiting time in the business process (detail)

Cases with the highest overall waiting times

Cases with the highest waiting times identify the cases with the highest overall waiting time and the highest waiting time between activities for each case.

Sixty-one cases last over 1,5 years. Half of the cases (31) have only four activities and end with cancelling the training need (see Figure 20).

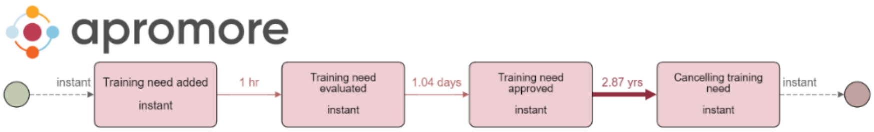

Figure 20. Example of cases with the highest waiting time (4 activities)

For the rest of the cases, it is visible that the longest waiting time between activities is at the beginning of the process, where training need is identified (see Figure 21).

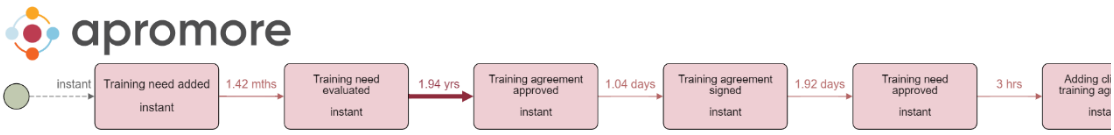

Figure 21. Example of cases with the highest waiting time (more than 4 activities) (detail).

When comparing cases that last up to one year (see Figure 22) and that last more than 1,5 years (see Figure 23), the different average waiting times at the beginning of the process are clearly visible.

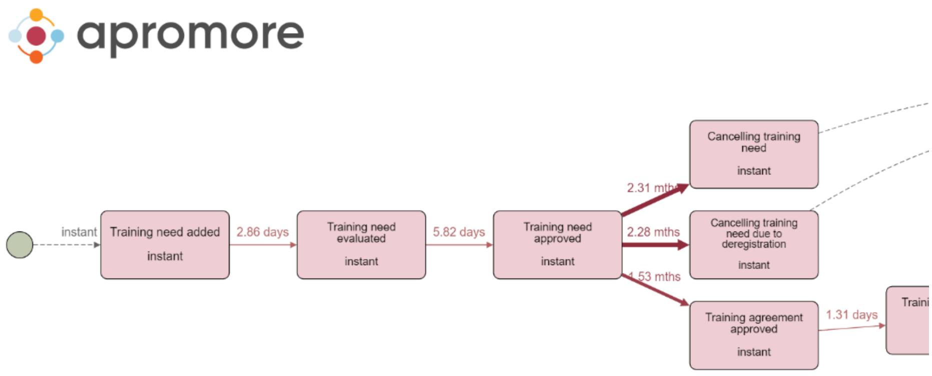

Figure 22. Cases with the highest waiting times (cases that last up to one year) (detail)

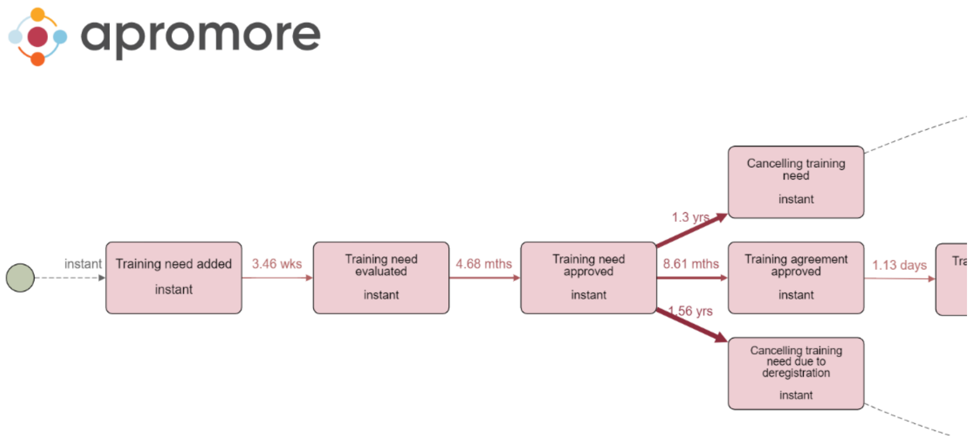

Figure 23. Cases with the highest waiting times (cases that last more than 1,5 years) (detail)

All 61 cases need further analysis of the root cause for so long waiting time between activities and how this could be minimised in the future.

4.4 Comparative Process Mining

In comparative process mining (also called variant analysis), the processes of training that are finished and training that is quit are compared. Activities related to processing participation sheets and payments are filtered out to make the two variants more comparable. Average and median case durations (3,6 and 3 months respectively) of quit training are about one month shorter than of finished training, which is also expected as training is quit before the end. Average training lasts for 2,3 months, and when the client quits the training, it happens on average 1,3 months after the start of the training agreement. Comparing active cases of two variants over time, it is visible that there was a higher proportion of quit cases between August 2020 to September 2021 (see Figure 24).

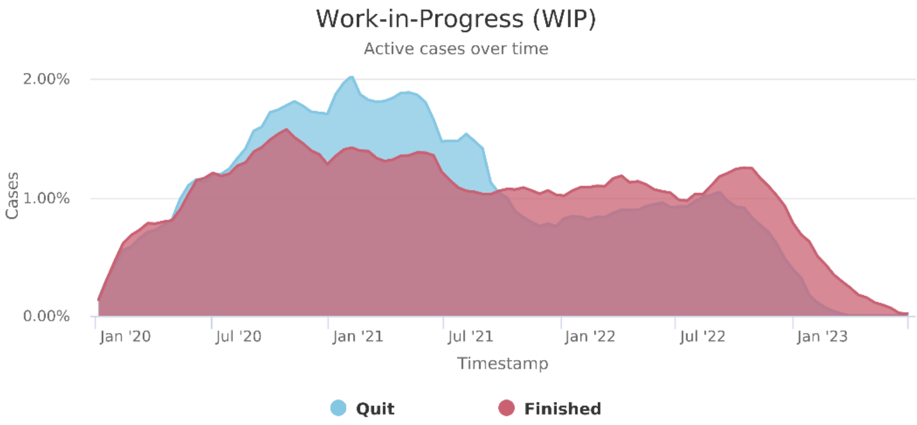

Figure 24. Quit and finished training active cases over time

When comparing BPMN models of two variants (10% arcs) that Apromore creates automatically, the processes are the same for both variants. Only the final activities (end of training and quitting training) are different, which is expected (see Figures 25 and 26).

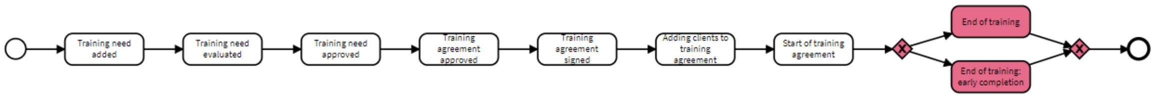

Figure 25. BPMN model of finished trainings (10% arcs)

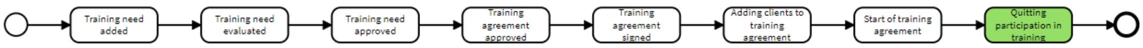

Figure 26. BPMN model of quit trainings (10% arcs)

When increasing the percentage of arcs to 50%, the changes in the order of activities become visible (see Figures 27 and 28). Finished trainings have activities related to training need before the training agreement approval activity, and quit trainings have the opposite order. Although this kind of flexibility is allowed in the process, this finding could be further analysed if this impacts clients quitting training.

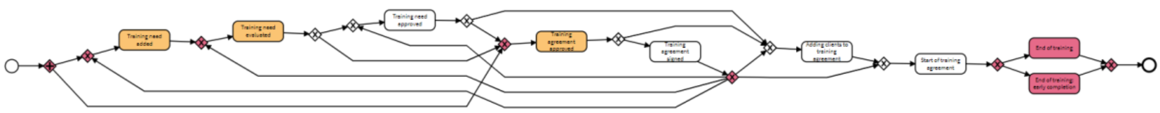

Figure 27. BPMN model of finished trainings (50% arcs)

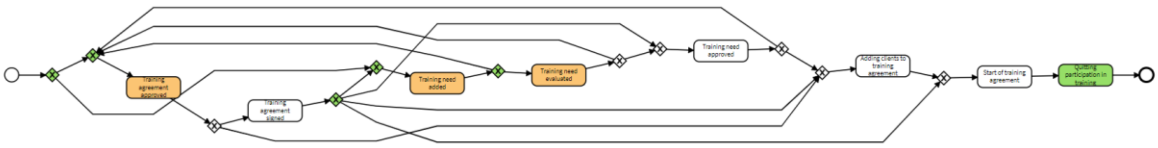

Figure 28. BPMN model of quit trainings (50% arcs)

Significant differences can be detected when comparing two variants based on the overall length of the training (see Figure 29).

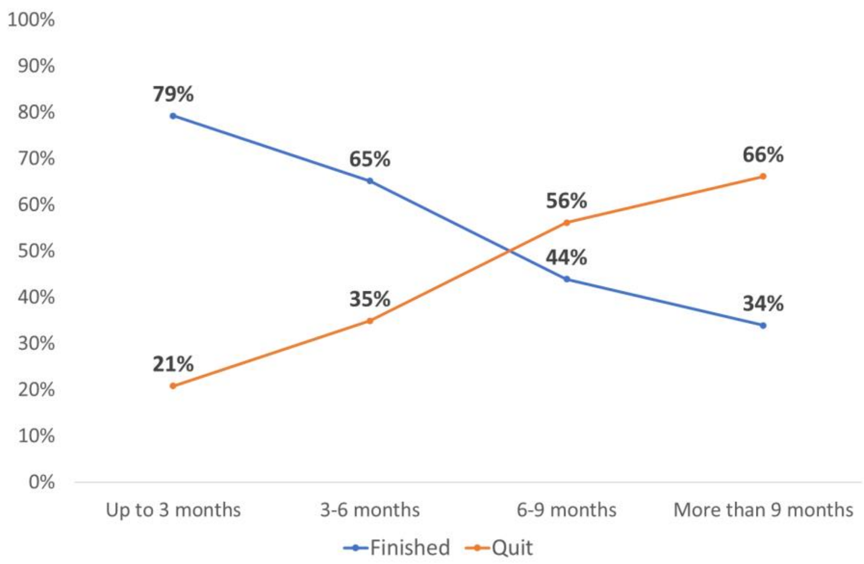

Figure 29. Proportion of clients who finish and quit training based on the length of the training

Longer training has a higher proportion of clients who quit the training, which indicates a correlation between the length of the training and quitting the training.

5 Discussion

Chapter 5 presents the discussion, limitations, and conclusion of this thesis.

The process mining results described in Chapter 4 were presented for validation and discussion with domain experts. According to domain experts, many unusual results (e.g. violation of the rule of maximum length of training, a larger proportion of quit trainings from August 2020 to September 2021) can be explained by Covid-19 restrictions in Estonia. The training was longer because they were put on hold. Face-to-face training access was restricted, and participation requirements (e.g. mandatory vaccination certificate) caused clients to quit training. Checking the data confirmed domain experts' hypothesis that these cases happened during Covid-19 restrictions. The analysis results also raised questions. For example, why some training needs are open for such a long time and what is the real root cause of why clients quit procured training?

This thesis sought answers to three research questions.

The first RQ aimed at discovering the labour market training process using process mining. The results of automatic process discovery are presented in Chapter 4.1 Process Discovery (see figures 8, 9, 10). The results show considerable flexibility and, therefore, significant variability in the process. Only 50% of clients reach the procured labour market training whose training need has been approved.

The second RQ aimed at discovering improvement opportunities in the process. The results of conformance checking showed three activities related to cancelling training needs and not starting the training that existed in the automatically discovered process model but were missing in the manually created process model. As the relative case frequency of these activities was 14-18%, they could be added to the manually created process model to reflect the reality of the process more precisely. Performance analysis showed that there were cases where the training need had been approved a long time ago, but the client had not received the training. These clients were still unemployed because the training need was active as these cases did not have activity "Cancelling training need due to deregistration". As the goal of ALMPs is to provide clients services that help them overcome employment obstacles, this issue needs further analysis to understand what are the reasons why the client has not been provided training. Another improvement opportunity is cases with long overall waiting times (over 1,5 years). Analysis showed that after the client's training need was approved, there were cases when they wait for the next activity for a very long time – the most extreme example was waiting 2,87 years until the training need was cancelled. Both performance analysis improvement opportunities indicate a bottleneck after the "Training need approved" activity that needs to be tackled to improve the process.

The third RQ aimed to understand the differences between finished and quit processes. Comparative process mining showed that on the higher level, the processes of two variants were the same: one process ended with quitting, and the second with finishing the training. When increasing the percentage of arcs between nodes, a different sequence of some activities became visible – finished training had activities related to training need before the training agreement approval activity, and quit training had the opposite order. Although this different sequence of activities does not violate the business rules, this could be analysed more if this is one factor that impacts the fact that training is quit. Significant differences were found comparing the training length and the training result (finished or quit). Longer training had a higher proportion of clients who quit the training.

The improvement opportunities detected during this process mining case study are described next. The improvement suggestions have not been implemented yet.

Firstly, the IT system database tables' attributes should be documented systematically during IT system development. This will help to make searching for the correct data needed for process mining easier and less time-consuming. Also, the exact time when the activity was performed should be recorded in the database. This will give an overview of the order in which the activities are performed on the same day. The activity's start time should be recorded in the database where possible. This will enable us to analyse the activities' length, which was impossible in this case study.

Secondly, the process of labour market training could be improved by adding additional automatic tasks for counsellors in EMPIS when the training need has been open too long – either to cancel the training need when it is not relevant anymore, search for other training possibilities for the client or to offer alternative ALMPs instead. This would eliminate the problem that training need is open too long without providing the training or the long waiting time after the training need approval prolongs the overall waiting time.

Thirdly, as the proportion of clients who quit the training increases with the length of the training, shorter training or training composed of different smaller modules could be organised.

Fourthly, a couple of process mining projects have been carried out in EUIF before, but they could be a regular exercise performed in EUIF to enhance data-driven decisions. Process mining could be used as one of the inputs when process review and improvement are planned. As the visual analytics platform Tableau[18] is used in EUIF daily, the ProcessMining[19] extension for Tableau could be tested for the following process mining projects.

Two further research suggestions are proposed next. Firstly, only half of the clients with training need reach the training. Further analysis is needed to understand what happens to clients who had their training needs cancelled or did not start the training. As they had obstacles returning to the labour market and needed skills enhancement, did they go to training using a training card, or did they use other ALMPs provided by EUIF? Cases, where no additional services were provided must be analysed as this might indicate that clients do not receive the help needed. The analysis can give input to evaluating the quality of counselling. Secondly, comparative process mining gave two ideas (order of activities and length of training) what could influence quitting the training, but further analysis could provide more information. For example, a descriptive analysis could be carried out comparing the results based on the topics of training, service providers, training channels etc. In addition, qualitative research on clients' learning motivation and reasons for quitting could be carried out. This might give additional information that can be used to minimise the problem of clients quitting the training.

5.1 Limitations

This case study had the following limitations. There were two occurrences of limited data. Firstly, only the end time of the activity is stored in EMPIS, so it was not possible to visualise and analyse the length of the activity, only the time between activities. This limits the ability to draw conclusions from the data, as some are missing. Secondly, another problem was that the end time of the activity was midnight (00:00:00) for all activities. This caused

[18] https://www.tableau.com/
[19] https://exchange.tableau.com/en-gb/products/166

difficulties in understanding and visualising the sequence of the activities. Activities performed on the same date were shown as parallel activities, although performed one after another. To fix this problem, the time data of activities was manipulated to correspond to the normal sequence of activities according to the manually created process model. Unfortunately, this way, it was impossible to detect the abnormal sequence of activities that occur on the same day, which can lead to wrong conclusions.

As the EMPIS's database table attributes are poorly documented, extracting and using the wrong data for process mining is possible, which can cause data validity limitations. Numerous manual checks of case activities were performed, and the results were validated with domain experts to minimise this risk.

The interviews were held with domain experts from the Head Office responsible for developing the labour market training service. Therefore, only a partial view was considered. For example, the view of counsellors and service consultants from Regional Offices was missing, which can cause the limitation of perspective. In the future, all actors in the process could be included to give input during the process mining planning stage and evaluate the results.

There is also a generalisability limitation – the results of this process mining case study cannot be generalised to other ALMPs that EUIF offers. Different ALMPs have been developed at different times by different people, and the rules and processes of ALMPs are pretty different. Separate process mining case studies must be conducted for different ALMPs in the future.

5.2 Conclusion

The objective of this thesis was to carry out data-driven process analysis using process mining methodology. Process mining is a data-driven technique that uses event log data from IT systems to discover, monitor, and improve business processes. This case study followed the Process Mining Project Methodology (PM^2).

The project started by defining the process that needed analysis. Different departments were approached and procured labour market training was selected for the case study. After that, the search for the data in the IT system databases started. Before data extraction, the data planned to extract was reviewed together with the domain experts. The event log was composed based on the data extracted and processed. Each row in the event log had a case ID, name of the activity, timestamp when the activity was performed, information about the role who performed the activity and the status of the activity. The event log had data from the last three years and had 553 037 rows with a total of 44 042 cases and 18 different activities.

Three research questions were raised for this process mining case study. First RQ aimed to discover the process using process mining. The analysis showed that the discovered process model is more detailed than the manually created one. As there are many case variants (4 827), the sequence of process activities is very flexible. The second RQ aimed to discover improvement opportunities in the process. Performance analysis indicated that in some cases, there is a bottleneck after the activity "Training need approved" that causes long waiting times and what needs to be tackled to improve the process. The third RQ aimed to understand the differences between finished and quit processes. Comparative process mining brought out two aspects (order of some activities and length of training) that are different when comparing processes when the client finishes or quits the training.

These research questions enabled the implementation of four different process mining use cases – process discovery (RQ1), conformance checking (RQ2), performance analysis (RQ2) and comparative process mining (RQ3).

Following improvement suggestions were made. To make process mining easier, more efficient, and more informative, we suggest systematically documenting the database tables' attributes description, recording the exact time when the activity was performed and adding the activity's start time in the database where possible. Automatic tasks to counsellors could be added to EMPIS when the training need has been open for too long, and shorter training could be offered to prevent quitting the training. In addition, process mining could in incorporated into EUIF's process analysis and improvement regularly.

This case study had limited data, data validation, perspective, and generalisability limitations that were mitigated when possible and proposals were made that can be taken into consideration when starting new process mining projects.

Based on the results of this case study, two further research suggestions were made. Firstly, to analyse what happens with these clients whose training need is cancelled, or they don't start the training – do they get help and services needed to overcome the obstacles entering the labour market? Secondly, the root cause of why clients quit training needs further analysis – either detailed descriptive analytics or qualitative research where the client's learning motivation and reasons for quitting are studied.

To summarise – this process mining case study gave insights into what is happening in the procured labour market training process and uncovered new information that can be used and analysed further when improving the procured labour market training process in EUIF. This thesis also gives additional input to the academic literature by providing a case study of process mining in Public Employment Services.

6 References

[1] W. M. P. van der Aalst and J. Carmona, Eds., *Process Mining Handbook,* vol. 448. in Lecture Notes in Business Information Processing, vol. 448. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-031-08848-3.

[2] H. Alqaheri and M. Panda, "An Education Process Mining Framework: Unveiling Meaningful Information for Understanding Students' Learning Behavior and Improving Teaching Quality," *Information (Switzerland)*, vol. 13, no. 1, Jan. 2022, doi: 10.3390/info13010029.

[3] T. Reil, E. Groher, and P. Siegfried, "Process Mining in Supply Chain Management," vol. 12, pp. 1–13, 2022, [Online]. Available: https://www.researchgate.net/publication/359704405_Process_Mining_in_Supply_Chain_Management

[4] "Process Mining In Government — Flux Capacitor." https://fluxicon.com/blog/2022/04/process-mining-government/ (accessed May 07, 2023).

[5] "Swedish Public Employment Service Leverages Process Mining and Automation | UiPath." https://www.uipath.com/resources/automation-case-studies/swedish-public-employment-service-leverages-process-mining-and-automation (accessed Apr. 30, 2023).

[6] M. Dees, M. de Leoni, and F. Mannhardt, "Enhancing process models to improve business performance: A methodology and case studies," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10573 LNCS, pp. 232–251, 2017, doi: 10.1007/978-3-319-69462-7_15.

[7] "Eesti Töötukassa põhikiri–Riigi Teataja." https://www.riigiteataja.ee/akt/131052017012 (accessed Mar. 28, 2023).

[8] "Development Plan | Töötukassa." https://www.tootukassa.ee/en/footer/estonian-unemployment-insurance-fund/development-plan (accessed Apr. 07, 2023).

[9] "Labour Market Services and Benefits Act–Riigi Teataja." https://www.riigiteataja.ee/en/eli/510032022003/consolide (accessed Mar. 28, 2023).

[10] "Training | Töötukassa." https://www.tootukassa.ee/en/services/career-and-training/training (accessed Mar. 28, 2023).

[11] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018. doi: 10.1007/978-3-662-56509-4.

[12] W. Van Der Aalst *et al.*, "Process mining manifesto," *Lecture Notes in Business Information Processing*, vol. 99 LNBIP, no. PART 1, pp. 169–194, 2012, doi: 10.1007/978-3-642-28108-2_19.

[13] W. M. P. van der Aalst, *Process Mining : Data Science in Action*, 2nd ed. Berlin, Heidelberg : Springer, 2016.

[14] O. Loyola-González, "Process mining: software comparison, trends, and challenges," *International Journal of Data Science and Analytics*. Springer Science and Business Media Deutschland GmbH, 2022. doi: 10.1007/s41060-022-00379-0.

[15] "2023 Gartner MQ Process Mining Tools report | Software AG." https://www.softwareag.com/en_corporate/platform/aris/process-mining-tools-gartner-report.html (accessed May 07, 2023).

[16] Y. Rashid, A. Rashid, M. A. Warraich, S. Sameen Sabir, and A. Waseem, "Case Study Method: A Step-by-Step Guide for Business Researchers," *Int J Qual Methods*, vol. 18:160940691986242, doi: 10.1177/1609406919862424.

[17] J. Farquhar, *Case Study Research for Business*. London: SAGE Publications Ltd, 2012. doi: 10.4135/9781446287910.

[18] T. H. C. Van Der Heijden, "Process Mining Project Methodology: Developing a General Approach to Apply Process Mining in Practice," Master's thesis, Eindhoven University of Technology, 2012. [Online]. Available: https://pure.tue.nl/ws/files/46910196/739765-1.pdf

[19] M. L. Van Eck, X. Lu, S. J. J. Leemans, and W. M. P. Van Der Aalst, "PM2: A process mining project methodology," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9097, pp. 297–313, 2015, doi: 10.1007/978-3-319-19069-3_19.

[20] S. Aguirre, C. Parra, and M. Sepúlveda, "Methodological proposal for process mining projects," *International Journal of Business Process Integration and Management*, vol. 8, no. 2, pp. 102–113, 2017, doi: 10.1504/IJBPIM.2017.083793.

[21] R. Hobeck, L. Pufahl, I. Weber, and E. Preprint, "Process Mining on Curriculum-based Study Data: A Case Study at a German University," 2022. [Online]. Available: https://www.researchgate.net/publication/364671374

[22] "What is Data Obfuscation? Definition and Techniques | Talend." https://www.talend.com/resources/data-obfuscation/ (accessed Apr. 16, 2023).

[23] N. Danylyshyn, "Analysis templates for identifying improvement opportunities using Apromore," Master's thesis, University of Tartu, Tartu, 2022. [Online]. Available: https://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=75330&year=2022

[24] "Public employment services - Employment, Social Affairs & Inclusion - European Commission." https://ec.europa.eu/social/main.jsp?catId=105&langId=en (accessed Apr. 30, 2023).

[25] "Active Labour Market Policy (ALMP) - European Centre for Social Welfare Policy and Research." https://www.euro.centre.org/domains/active-labour-market-policy (accessed Apr. 24, 2023).

[26] "Data exchange layer X-tee | RIA." https://ria.ee/en/state-information-system/data-exchange-platforms/data-exchange-layer-x-tee (accessed Apr. 12, 2023).

Appendix

I. Licence

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Triin Kraavi,

1. grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

Data-Driven Process Analysis: Case Study of Labour Market Training Using Process Mining

supervised by Fredrik Milani.

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Triin Kraavi

09/05/2023