Process Mining and Performance Business Rules

Ella Roubtsova[a] and Yoeri Berk
Open University of the Netherlands, Department of Information Science, Heerlen, The Netherlands

Keywords: Process Mining, Performance Business Rules, Interview, Methodology, Application Process of an Employment Agency.

Abstract: Process mining uses event logs to build a business process model, which potentially can be used to better understand the business process performance. However, the understanding of performance almost always demands definitions in form of performance business rules within the business process. This means that process mining should be accompanied with another research method to retrieve the performance business rules of the studied business process. This paper presents an approach showing how the process mining techniques shape the structure of the semi-structured interviews for retrieving performance business rules to be checked on the logged business process. The proposed approach is illustrated with a case study of process mining and performance assessment of an application process in an employment agency.

1 INTRODUCTION

Some modern research methods based on data analysis are not completely self-sufficient when they should serve the needs of business (Williams, 2016),(Dabrowski et al., 2017). One of such business needs is the performance analysis of business processes. Ghasemi and Amyot (2020) in their review of process mining research papers has found that "research about performance indicators measuring goals associated with process mining is sparse"(Ghasemi and Amyot, 2020).

Our research has been inspired by the need for performance analysis at an employment agency. An application process at an employment agency determines a candidate-pool that is proposed by the agency to the client-company, looking for employees. The correspondence of the candidates to the predefined qualities and the speed of delivery of the candidate-pool influence the added value, reputation of the employment agency and the next requests of clients. The move to online job hunting has increased the reach of job postings and thus the number of applications. Each application has own process case, because its applicant may forget to send a necessary document that creates a process loop, or may be rejected at different stages. With this high number of applications and the large variety of process cases, the analyzing process performance is challenging.

Process mining is a technique that is aimed to extract process-related knowledge (e.g., process models) from event logs and exploit it for further analysis (Cheng and Kumar, 2015). Initial application of the process mining tool Disco (DISCO, 2022) to a log of agency (Berk, 2021) reveals, for example, that for one client, 1,472 applications resulted in 224 different process cases. Inspired with this process at an employment agency we have investigated the following research question: *What research method should accompany process mining for performance analysis of a business process with many process cases?*

Section 2 presents the results of our search of an answer in literature. Section 3 describes a structure of a semi-structured interview for gaining performance business rules. *The interview has been designed based on filters chosen in process mining tools.* Section 4 discusses a case study applying a semi-structured interview with the designed structure, retrieving performance business rules and using them for process mining and performance assessment. Section 5 presents conclusions and future work.

2 PERFORMANCE AND PROCESS MINING

The definition of performance found in literature have two parts.

[a] https://orcid.org/0000-0002-4067-3088

Roubtsova, E. and Berk, Y.
Process Mining and Performance Business Rules.
DOI: 10.5220/0011056300003176
In *Proceedings of the 17th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2022)*, pages 387-394
ISBN: 978-989-758-568-5; ISSN: 2184-4895

- First, performance is often defined as the achievement of quantified objectives. It is often called effectiveness, i.e."delivering desired outputs, and even outcomes" (Ghalem et al., 2016, page 7).
- The second definition of performance includes not only a matter of what people achieve, but also how they are achieving it; how they are using the available knowledge and resources; how they are choosing the shortest(cheapest) process path of the business process. The second definition is called efficiency (Ghalem et al., 2016, page 7).

In this work we follow (Ghalem et al., 2016, page 9), who "believe that performance is the combination of these two terms (effectiveness and efficiency), however, and depending on the context in which the performance might be used, other elements can be added to define the term, such as relevance, economy, etc."

If we define an event as a data representation of an occurrence of interest in the real-world business domain (McNeile and Simons, 2006); if we define a business process as a structure of ordered events (Roubtsova and Wiersma, 2018), then

- the effectiveness can be seen as reachability of desired events in the process;
- the efficiency, can be seen as desired subsequences of events or orders of events.

A log of events corresponds to the definition of a process as a a structure of ordered events and the two performance elements can be related to a log. "A log is a textual file containing information about events recorded in the order of happening." (Roubtsova and Wiersma, 2018). Indeed, a log stores the ordered events using the time-stamps. A log describes each event using a log record that contains case designation (process instance), activity(event) label, time stamp, and also resource executing event or used for execution and data elements recorded with the event (e.g., the size of an order, the reason of withdrawal. etc.).

On the other hand, a log is the input for process mining techniques. The question is whether there is a process mining type that can sufficiently serve to assessment of effectiveness and efficiency of the mined business process.

The three main types of process mining are process discovery, conformance checking, and process model enhancement (Lamghari et al., 2019).

In process discovery, no prior process-related information is used, no quantified objectives are given. The business process discovered by mining the event log is analyzed to see some degree of effectiveness: What outputs are delivered? It is difficult to analyze the efficiency of the mined process. The analysis may show that some cases of the process have more steps, but it remains an observation as the analyst does not know what is efficient for the mined process.

In conformance checking, an existing process model is compared with an event log of the same process. The results of conformance checking present some descriptive elements of efficiency and effectiveness: what events are often skipped, what resources are scars. If the existing process model is defined by or transformed to business rules that define desired outputs, desired sequences of process steps, the conformance checking can be used for analysis of effectiveness and efficiency. The practical problem is that, the desired process is often presented as a "happy flow", i.e. an ideal process that does not cover exceptions and the business rules are not documented (Berk, 2021). More over, to analyze efficiency, the desired time and resource restrictions have to be added to existing process.

In process model enhancement, there is a current (a-priori) process model (as a goal) and a process mining of an event log. The process mining is used to find possibilities to improve the mined business process. Process model enhancement aims to change or extend the a-priori model (van der Aalst, 2016). Additional requirements are used to extend the model. They may introduce new outputs. i.e. change model effectiveness. They may introduce new possible events, time and resource restrictions, i.e. change the process efficiency. Any model change should have its own goal that can be expressed in terms of performance. This means that the process model enhancement type of process mining is related to process performance. As we see, this process mining type needs some methods that provide the goal, requirements and business rules for process model enhancement.

The concept of an organizational "performance" can be seen as the degree to which the organization meets its objectives (Hong, 2016). The objectives consist of three elements: targets (outcomes) to be reached, ways to get targets (outcomes) and process time (van den Ingh, 2016). In order to measure these elements of performance, business rules and performance indicators are usually defined. Business rules are often related to the execution order of tasks in cases, to the involvement of a role in cases and processes. The performance indicators are related to the business value (costs, time) achieved in cases of the analyzed process.

For example, the Performance Business Rule "*CV Received* comes before *Candidate Proposed*" is about the execution order of the process. ''*CV Received"* and *"Candidate Proposed"* are two events on the business process. Another example, the Performance Business Rule "Every handling of an applica-

tion should take less than 5 days" specifies the desired time of the process "Handling of an application". The time of a specific case of the process "Handling of an application" can exceed 5 days and violate this performance business rule.

There are goal-driven methods for applying process mining to performance. The process mining project methodology PM^2 (van Eck, M. L., Lu, X., Leemans, S. J. J., and van der Aalst, W. M. P., 2015) uses the first iteration of process mining to obtain general insights into the process. After that the goal and business rules are formulated. The PM^2 (van Eck, M. L., Lu, X., Leemans, S. J. J., and van der Aalst, W. M. P., 2015) methodology provides a solution for planning and management of process mining. However, the methodology does not explain how business rules are retrieved from stakeholders and how the process mining filters and techniques are selected.

Therefore, we propose to extend the methodology with semi-structured interviews of process stakeholders. We propose a reusable structure of such an interview influenced by filters of process mining tools that can be mapped on business rules. The interview is aimed to collect the performance business rules that can be directly reformulated into process mining filters.

3 COLLECTING PERFORMANCE BUSINESS RULES

Let a log structure be given. For example Case, Activity, Time Stamp, Resources. A log entry is a record of the log structure filled with values. Let a set of log entries be given. This set of log entries presents a set of cases of a process (workflow).[1]

The filters of the process mining tools support filtering the cases with true (or false) values of specified temporal and boolean expressions. The filters can be classified as

- Followers filters : SELECT CASES such that a temporal relation (an order) on activities(events) is specified with operators ALWAYS, NEXT, IN-FUTURE, ALWAYS UNTIL, EXITS UNTIL.
- End-Activity filters: SELECT CASES that end with specified events or entries with specified attributes.
- Times Frame and Attribute filters: SELECT ENTRIES with specified time frame and/or attribute value.

[1]A workflow is a process that starts and ends with the specified start and end activities. A log may contain entries of many workflows (Berk, 2021).

In order to use Process Mining for assessment of business process performance, we need to identify business rules that correspond to filters of process mining tools. We propose a method that combines Process Mining and semi-structured interviews of the reusable structure. The interview structure and, further, the interview questions have been designed around the filters provided by process mining tools.

3.1 Method Combining Interviews and Process Mining

1. Introduction. *"We investigate the use of tools that analyze logs. We do not try to influence the business process of your company."*
   Look at the process identified in the log by the tool."
   (a) *Do you recognize the names of activities(events)?*
   (b) *Rethink last process cases that you handled.*
2. Visualization of the business process using the activities named by the interviewee.
3. Performance Business Rules.
   (a) *What is the goal of a process (a set of cases)?*
   (b) *What is the goal of each process case?*
   (c) *How can you determine whether the goal is achieved?*
   (d) *Are there examples where the internal goal of company deviates from the external goal of the process of clients?*
   (e) *Are there events(activities) that should follow one another? What is the relation of these events(activities)?*
   (f) *Are there mandatory events(activities) in each case? Why are they mandatory?*
   (g) *When can the case be ended? What is the desired outcome of a case?*
   (h) *How is defined who executes a process or a case?*
   (i) *Are there cost-benefit considerations in the choice of an event(activity)?*
   (j) *Are there any obligations about time duration of a case of a business process?*
   (k) *Is the time between events (activities) decisive for the success of the process? If yes, between which events(activities)?*
4. Analysis of answers of interviewees to identify business rules and performance indicators.
5. Process mining for each identified business rule and performance indicator.

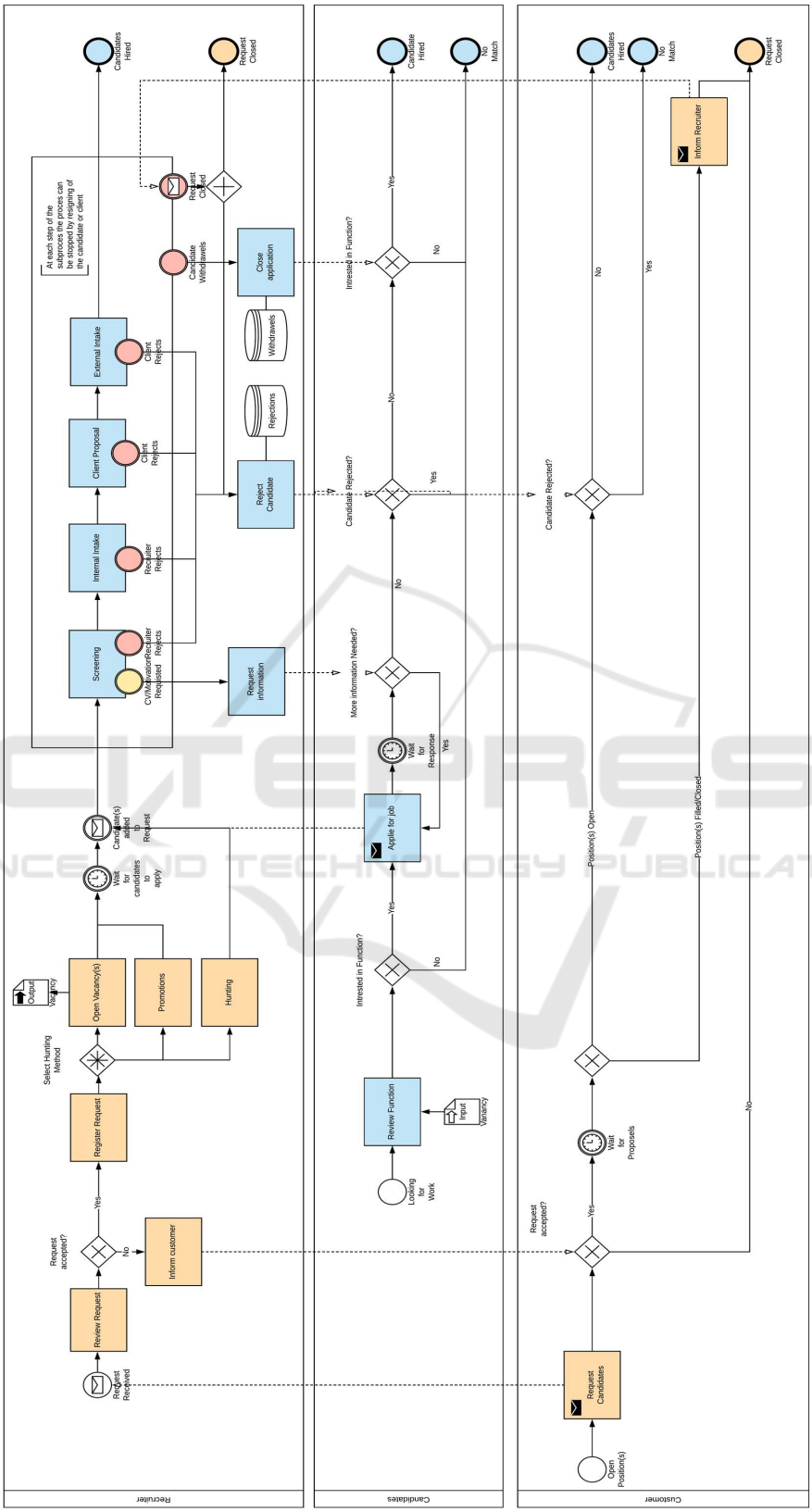

Figure 1: Business process of a recruiting agency.

4 CASE STUDY: EMPLOYMENT AGENCY

The scope of the log extraction was limited to all requests received by an Employment Agency and finished during the period from 01-10-2020 to 31-12-2020. This period was selected to rule out seasonal market influences. The extracted log can be found in (Berk and Roubtsova, 2021).

4.1 Introduction of Interviews

The interviews have been conducted with four recruiters. On the basis of the answers, a business process has been depicted in BPML (Figure 1). The activities mentioned by interviewees have been used for the process description.

4.2 Visualization of the Business Process

Figure 1 shows that a client-organization, that hires, submits a request for a candidate (Request Candidate) that is logged in the requests. The recruiters use this request to track demand and handle applications.

A request is reviewed (Review Request), than it is registered and sent to one of the hunting methods: Open Vacancy, Promotions leaflets and Newspapers, Database search, Use of machine learning tools. Positions are open and Applications are accepted until the customer withdraws, the customer fills the position, a competitor fills the position, or, in the best-case scenario, the agency fills the position. The request is then closed and the process ends (Berk, 2021).

If an Application has been accepted, it is screened on required parts (CV, Motivation). If application is complete, its applicant goes through an Internal Intake by the recruiter, Client Proposal, and External Intake by the hiring client. There may be no match if the recruiter or client finds the candidate not suitable, depending on where in the procedure the candidate is. In addition, in all of these procedures, the candidate may lose interest in the function or the request may become closed due to reasons outlined in the request process. In the best-case scenario, the candidate passes through all activities and has an opportunity to start within the function he/she desired. If a candidate is hired, the Request Candidate Process is closed.

The Application Process (Figure 1) can be divided into the main process of handling a request of the client and the sub-process for applications, where each request can have a number of applications. The processes have to be analyzed together as the Request Process has data needed for the Application Process. Therefore, the Request Process activities (Open, Close, Connect Vacancy) have been added to each application log. Thus, applications within one request have the activity "Open Request" with the same time stamp. This makes it impossible to count the number of Open Request activities, but makes it possible to compare the time stamps within the requests with the time stamps of the applications. The data transformations have been done within Alteryx-tool (Alteryx, 2021) and resulted in 326,198 records with 1,420 requests and 48,307 applications. The fields in the dataset are presented in Figure 2.

4.3 Performance Business Rules

In order to identify business rules in answers of interviewees, the analyst should know the activities of the logged business process. Analyzing the answers of the interview questions, the analyst identifies pairs of following activities in one case and uses them to filter cases.

For example,
Question: What is the goal of the process?
Answer: "You want that the person you propose to be accepted."

Business rule: A business rule identifies a case that has event [Proposed] and "in future" event [Accepted]. If a case contains [Proposed] and gets "in future" event [Rejected], this case is filtered as a case violating the business rule. The answers of interviewees and their analysis are available in (Berk, 2021).

For each business rule, a new process mining process is started. Each business rule aims at a specific problem, e.g., the relationship between vacancies and exit reasons. Because of the space limitation, we present process mining for four business rules of 10 identified business rules.

Business Rule 1: *A vacancy should attract candidates who meet the demands requested by the client.* The agency uses vacancies to attract and select candidates and fill in the open positions for the client. Each application of a candidate is screened; this takes time. The application and the job vacancy description are used to make an assumption whether or not the applied candidate is suitable for the job. If this assumption is incorrect, it could be caused by the quality of the job vacancy description. Not every rejection in the screening stage is based on requirements for the function. There are other exit reasons: hard requirements, availability, accessibility etc. These exist reasons should be specified by business rules.

The business rule 1 combines

- *Trace: [Vacancy opened,New,Rejected] and*

Field	Description
Request_ID	Anonymised record count for each new request
Request_received_at	Recruiter register when the request is received
Req.created_at	System register when recruiter creates request
Request_candidate_starts_working_at	Recruiter register when client wants candidates to start
Request_closed_at	Recruiter register when reqeust ends
Request_Client_ID	Anonymised record count for each new client
Request_Unit_ID	Anonymised record count for each new unit
Request_User_ID	Anonymised record count for each new recruiter
Application_ID	Anonymised record count for each new application
Application_candidate_ID	Anonymised record count for each new candidate
Application_log_ID	Anonymised record count for each new status in application
Application_log_created_at	Date and time from the start of each new status
Application_log_closed_at	Date and time from the end of each new status
Application_log_status_name	Name of status
Application_log_status_bucket_name	Grouping of statuses
Application_log.rejection_reason	Reason for rejecting candidate, empty when not rejected
Application_log.rejection_subreason	Subreason for rejecting candidate, empty when not rejected
Application_log.withdrawal_reason	Reason for withdrawal from candidate, empty when not withdrawn

Figure 2: Fields of the log of the analyzed employment agency.

- *Source: Rejected reason = "hard_requirements", "availability" or "accessibility".*

The process map for all 1015 applications in the Unit A of the agency was made in Disco (DISCO, 2022)(Figure 3). This map shows 100% of the activities and 25% of the paths. 372 (marked in Figure 3) go directly from New to Rejected; 275 of these applications were identified by this filter.

The summary results (Table 1) show that for the entire agency, approximately 1 of 4 applications are from candidates who do not meet the stated requirements. The conclusion is: "the requirements should be made clearer by the job vacancy description."

Business Rule 2: Select cases with withdrawal reasons that should be clear from the job vacancy description: "commuting distance", "different function expectations", "hours at customer", "salary", "travel expenses", "want no shifts".

The amount of cases with these withdrawal reasons is 1%. This shows that the job vacancy description is clear about these mentioned points.

Business Rule 6: *A recruiter should react within* 24 *hours after a candidate's application.*

The 24-hour reaction time on an application is a standard by the agency to ensure that candidates are not waiting in the process and are quickly open for new opportunities. In addition, candidates could find a job via other agencies; thus, the faster the agency responds, the less opportunity there is for other agencies to get in contact with the candidates. When a candidate applies for a job, the application is automatically added to the request, which is registered as "New" in the log. The candidate then goes to screening, which results in either "Reject Candidate" or "Plan Intake" in the log, which is shown in Figure 1.

The choice between "Reject Candidate" and "Plan Intake" is not made mandatory in the system; therefore, in this business rule, we look at any activity that occurs after "New", which results in this business rule that identifies processes that do not meet the 24-hour reaction time.

The business rule combines the trace that starts with *New* and follows by any other event (activity) $*$:

Trace: [New,] Time between events New and the next following event is:* > 24 *hours.*

Based on the data for the entire agency (Table 1), the reaction time is above 24 hours for approximately 1 of 3 applications. Thus, for these cases, the standard of reaction has not being met. The validity of the standard of reaction should be discussed in agency.

Business Rule 7: The agency works to prevent the cases when each "Proposed" candidate is "Rejected" by the client.

The cases with a fragment

Trace[Proposed, Rejected]

violate the business rule 7.

47% of the candidates proposed by the agency are rejected by the client. This is a warning result that agency should discuss and understand.

Table 1: Percentage of cases violating Business Rules.

N	Business Rule	% of	Agency
1	Each vacancy should attract candidates who meet the requested demands from the client	Total cases	27%
2	Each vacancy fully described the employment conditions	Total cases	1%
3	Each vacancy is directly opened after receiving a request	Total cases	15%
4	Each vacancy is closed in time to get a chance to every applied candidate	Total cases	44%
5	The number of cases with intake is limited	Cases with intake	38%
6	The recruiter reacts within 24 hours after receiving of each application.	Total cases	33%
7	Each candidate proposed to the client meets requested requirements	Cases with Proposed	47%
8	Proposed candidate had an internal intake	Cases with Proposed	49%
9	Candidate turns up at his appointment	Total cases	0%
10	Proposed candidate fits within the team of the client	Cases with Intake s	10%

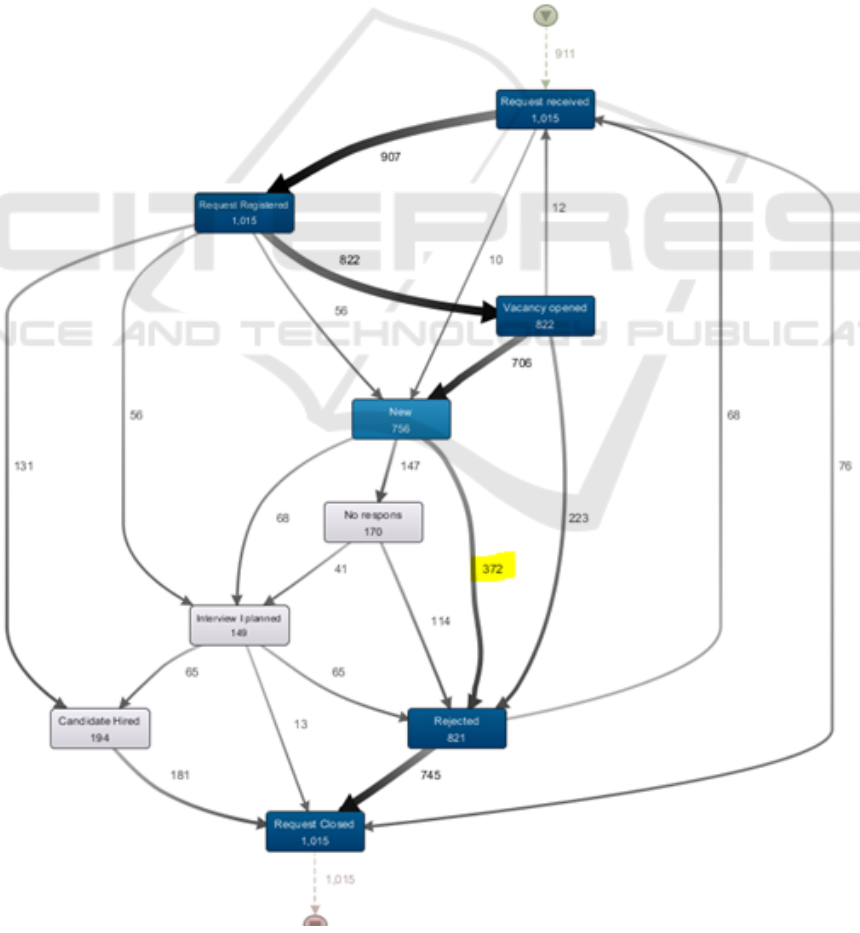

Figure 3: Disco Map for Business Rule 1.

4.4 Summary of the Application of the Proposed Method

The semi-structured interviews have retrieved 10 business rules that have been mapped to process mining filters. In order to answer the interview questions, the interviewees should have knowledge of the detailed steps of the process. Thus, most useful answers have come from the operational employees. Some answers have not mentioned the activities of the business process.

Table 1 presents the results of Process Mining for analysis of violation of 10 business rules. Disco (DISCO, 2022) filters based on the business rules have been used to separate wanted and unwanted process cases. The percentage of wanted cases to all cases gives an information about performance of the process. The agency has identified the process fragments for which improvement can be made.

Process discovery has been used to observe how the performance business rules affect the process map. Each of the unwanted processes can also be analyzed separately to identify reasons for deviation from the expected process.

5 CONCLUSIONS

The results reported in this paper provide an evidence that process mining directs business to well understood and defined performance analysis, to formulation of goals and business rules.

We have proposed a method that uses filters available in process mining tools to shape the semi-structured interviews. The interviews are aimed to identify the performance business rules for these types of filters. Filters are related to the business process performance and are used to understand and formalize the desired outcomes and desired process sequences, their time and resources restrictions. As we identify performance business rules, our research combines elements of goal-oriented methods with process mining techniques.

In this paper the method has been used for performance analysis of a business process at an employment agency. In near future, several projects will reuse the combination of process mining and semi-structured interviews for identification of performance business rules in different business domains. The aim of this replication experiment is to generalize the types of performance business rules applicable for process mining.

REFERENCES

Alteryx (2021). Simply Powerful. Extraordinarily Easy. https://www.alteryx.com/.

Berk, Y. (2021). A process mining method for analysing performance-based business rules of the application process at an employment agency, https://research.ou.nl/ws/portalfiles/portal/41056781.

Berk, Y. and Roubtsova, E. (2021). Data from employment recruiting agency. https://doi.org/10.17026/dans-zjv-87yz.

Cheng, H.-J. and Kumar, A. (2015). Process mining on noisy logs—can log sanitization help to improve performance? *Decision Support Systems*, 79:138–149.

Dabrowski, J., Kifetew, F. M., Muñante, D., Letier, E., Siena, A., and Susi, A. (2017). Discovering requirements through goal-driven process mining. In *25th International Requirements Engineering Conference Workshops (REW)*, pages 199–203. IEEE.

DISCO (2022). Flexicon DISCO, http://fluxicon.com/disco/.

Ghalem, Â., Chafik, O., Chroqui, R., and El Alami, S. (2016). Performance: A concept to define. *La performance: Un concept à definer*.

Ghasemi, M. and Amyot, D. (2020). From event logs to goals: a systematic literature review of goal-oriented process mining. *Requirements Engineering*, 25(1):67–93.

Hong, T. T. B. (2016). Process Mining-driven Performance Analysis in Manufacturing Process: Cost and Quality Perspective, http://unist.dcollection.net.

Lamghari, Z., Radgui, M., Saidi, R., and Rahmani, M. D. (2019). Passage challenges from data-intensive system to knowledge-intensive system related to process mining field. In *ArabWIC 6th Annual International Conference Research Track*, pages 1–6.

McNeile, A. and Simons, N. (2006). Protocol modelling: a modelling approach that supports reusable behavioural abstractions. *Software & Systems Modeling*, 5(1):91–107.

Roubtsova, E. and Wiersma, N. (2018). Involvement of business roles in auditing with process mining. In *International Conference on Evaluation of Novel Approaches to Software Engineering*, pages 24–44. Springer.

van den Ingh, L. (2016). Evaluating business process performance based on process mining, https://pure.tue.nl/ws/portalfiles/portal/46934802/846648-1.pdf.

van der Aalst, W. (2016). Data science in action. In *Process mining*, pages 3–23. Springer.

van Eck, M. L., Lu, X., Leemans, S. J. J., and van der Aalst, W. M. P. (2015). PM2: a process mining project methodology. In *Lecture Notes in Computer Science*, volume 9097, pages 297–313. Springer.

Williams, S. (2016). *Business intelligence strategy and big data analytics: a general management perspective*. Morgan Kaufmann.