

Process Mining



Institute for Business Informatics (IWI)
at the German Research Center for Artificial Intelligence (DFKI)

Winter Term 2021/2022

Prof. Dr. Peter Fettke
M.Sc. Alexander Beuther | M.Sc. Peter Pfeiffer

Case Study – Information and Description

Issue:	24.01.2022
Due:	06.02.2022, 23:59 Uhr
Points:	100 Points

Organizational Information

The subject of the case study is the analysis of a dataset presented at the Business Processing Intelligence Challenge 2013 (BPIC 2013). The challenge provides an event log from Volvo IT Belgium. The log contains events from an incident and problem management system called VINST. Provided are the original CSV file as well as three event logs that were extracted from this CSV file with the incident or problem number as case ID. Next to the CSV file and the three logs, there is also a document detailing the data in the CSV file and providing background to the process. Furthermore, this document contains several questions of the problem owner. Finally, the manual of the VINST system is included. The challenge with data, descriptions and interesting submissions can be found in Moodle as well as on the website of the challenge¹.

¹ <https://www.win.tue.nl/bpi/doku.php?id=2013:challenge>

Task Description

Volvo IT Belgium commissions you to analyze a data set using process mining techniques. The process owners hope to gain practical insights and recommendations for action from their analyses. Create a target group-oriented written elaboration and present your results to the responsible persons. Keep in mind that the addressees do not necessarily have a data science background. The focus is on producing interpretable, understandable results that are tailored to the target audiences in the form of a report. The target group are both business units that are involved in the process as well as the product owner of VINST.

Your task is to first familiarize yourself with the problem of the data set using the information provided. Then, analyze the provided data using the methods covered in the lecture. You may use any software tools available to you for the analysis. For every analysis, describe your approach step-by-step, the results and discuss the results. Derive concrete answers and recommendations for action (e. g. as suggestions for improvement or descriptions of the causes of problems) for the process participants and the process owner and formulate the results in the form of a report. Address your recommendations to the process owner (organizational) and the organizational units (functional). Furthermore, address the methodology and techniques used.

The following points and questions are an unbinding guide, which may help you gain insights and recommendations. However, you do not have to answer these questions step by step:

- (1) Your task is to first familiarize yourself with the problem of the data set using the information provided.
 - a. What is the log about?
 - b. What activities, people, resources, and other fields are included in the data? Specifically address the data collection methodology of creating the log. Interpret and explain the *Status* and *Sub-Status* data fields.
 - c. How many cases and events are there? Over what period of time does it record?
- (2) Describe the data set using descriptive methods: How many different process variants can you identify, how do they differ, for example, in terms of processing times, what is the statistical distribution of process parameters, etc.? Compare the behavior of users of different organizational units and service levels, focusing on differences in terms of the functional areas used. Are there any other interesting facts?
- (3) For the Process Owner, a more detailed understanding of conspicuous process behaviors is of great interest. For this purpose, identify various process behaviors, such as:
 - a. **Push to Front (incidents only)** Is there evidence that cases are pushed to the 2nd and 3rd line too often or too soon?
 - b. **Ping Pong Behavior** How often do cases ping pong between teams and which teams are more or less involved in ping-ponging?
 - c. **Wait User abuse** Is the "wait user" sub-status abused to hide problems with the total resolution time?

- d. **Process Differences per Organization** Where do the two IT organizations differ and why?

Besides the input of the process owner, can you identify more bottlenecks and suspicious process behavior in terms of process performance and throughput times?

- (4) Read and understand the proceedings (submissions by other authors) of BPIC13. Are the above issues addressed there? What other findings were made? Are there discrepancies between your findings and the findings in the proceedings?

Assessment

As the process takes place between different organizational units, the interpretability of the report is strongly emphasized. We aim to provide feedback that can be easily interpreted by end-users in the industry. Submissions will be evaluated according to the following criteria:

1. Understandability: the results are easy to understand and interpretable by the end-users in the domain.
2. Precision: the insights and suggested improvements are informative for the defined stakeholders.
3. Originality: the report proposes new techniques or uses existing techniques in an original and interesting way for this specific scenario.

Notes

1. A methodology for process mining projects is *PM*². You are welcome to use this as a guide for working on the case study.
2. We encourage you to develop your own scripts to solve it. Therefore, you can also work on and submit the case study in the form of a Jupyter notebook. If you choose to do so, convert the notebook with the results into a PDF and hand it in additionally.

² http://www.processmining.org/_media/blogs/pub2015/pm2_processminingprojectmethodology.pdf

Submission and Presentation

Summarize the results of your analyses in the form of a written report. The required length 7000 words +/- 10%. Figures and tables are explicitly encouraged but do not count towards this length. Use the format: page spacing left, right, and top 4 cm, bottom margin 3.5 cm, font size 12 points, font Times New Roman or Times, 1.5 times line spacing. The following points should be addressed in the paper, but there is no explicit structure:

- Introduction and problem description
- Descriptive analysis of the data set
- Description of the software tools used for the analysis
- Answering the questions
- Recommendations for the process owner

The results will be presented after the groups have submitted their papers. The focus should be on the answers to the questions posed and the identified potential for improvement. The time allotted for the presentation is 30 minutes, divided into a 20-minute presentation of the results and a ten-minute question section. Please hand in your presentation slides in advance. You have to do this until immediately before your presentation.

Important Notes

To answer the questions, you are allowed to use the results for the Business Processing Intelligence Challenge 2013 (BPIC 2013) as a guide. The transparent presentation of your solution path is crucial for the assessment. Document all intermediate results (e. g. process model, scripts, evaluations) of your analyses and attach the corresponding representations to your submission. Of course, cite whenever you use ideas and information from other sources.