Process Mining Team Assignment

Utrecht University
Department of Information and Computing Sciences
Faculty of Sciences

INFOMPROM

Team Assignment
Version 1.0

Academic Year 2024-2025
Period 4

Lecturer and course coordinator:
dr. ir. Xixi Lu

Instructor:
Suhwan Lee

Assignment supervisors:
Xixi Lu
Suhwan Lee

1. Introduction

In this team assignment, together with your team members, you will select a track and complete the assignment accordingly. There are three different tracks. Track 1 focuses on conducting process mining research of your interest. Track 2 focuses on designing and implementing your process discovery algorithm. Track 3 focuses on analyzing real-world event data using Process Mining techniques and tools covered during the lectures and tutorial sessions.

Track 1: Process Analytics

The objectives of this assignment track are multifold. Firstly, you should be able to apply or follow a process mining project methodology of your interest, which in general includes multiple different stages. In the initial face, you should be able to explore the event data, discover process maps/models, and come up with business questions of your interest with motivations. Next, to answer the event data, you may perform data/log preprocessing techniques and apply process mining techniques (such as process discovery, conformance checking, compliance checking, process enhancement, process prediction, or process optimizations).
Finally, you should be able to create a data science report that documents the questions, the objectives, the methods, the analysis conducted, and the insights obtained into the selected process.

Throughout this assignment, you should develop a practical approach for performing process mining and related data analytics activities on real-world event data and get acquainted with the use of contemporary tool environments for process mining.

For process mining techniques and analysis, participants are encouraged to use any tools, techniques, or methods at their disposal.

For this assignment, we will use the Business Process Intelligence Challenge (BPIC) logs, which provide participants with real-life event logs and questions to analyze these data using whatever techniques are available.

In particular, the participants can choose three different real-world event data sets:

- The BPIC 2014 log [Difficulty 1.5/3]
  The log is made available via Blackboard. All information about the challenge and the type of contributions for it can be found at https://www.win.tue.nl/bpi/2014/challenge.html. For this assignment, please use the preprocessed event log provided via the Blackboard.

- Sepsis log [Difficulty 2/3]
  Alternatively, you may also choose to work on the sepsis event log (see https://data.4tu.nl/articles/Sepsis_Cases_-_Event_Log/12707639)

- *the BPIC 2019 log [Difficult 3/3]*
  https://icpmconference.org/2019/icpm-2019/contests-challenges/bpi-challenge-2019/

Track 2: Process Mining Research

For the research track, you will be conducting a research project on one of the process mining topics of your selection, under the condition that the topic or research problem has been consulted with and agreed upon by your assignment supervisor.

The research focus can be but is not limited to

(1) selecting a recent process mining paper that proposes a new technique or a new approach and replicating the experiment in this paper,
(2) conducting a systematic literature study on a recent, novel process mining topic,
(3) applying a newly proposed technique or approach to a new set of event data and analyzing its performance and usability

As examples, you may find recent research papers on process mining published in the International Conference on Business Process Management (https://link.springer.com/conference/bpm) or in the International Conference on Process Mining (https://dblp.uni-trier.de/db/conf/icpm/index.html).

Track 3: Process Discovery Algorithm

As a team, you will implement a process discovery algorithm and evaluate the quality of the process model discovered. The process discovery algorithm that you implement can either be designed by yourself or an adaptation of an existing one (such as improving Inductive Miner, Transition System miner, ILP miner, or any other miner that returns a formal, semantically clear process model). In addition to the final report, you should provide an executable implementation, in Python.

The following recent studies may provide some examples of the recent process discovery algorithms.

[1] Augusto, A., Conforti, R., Dumas, M. and La Rosa, M., 2017, November. Split miner: Discovering accurate and simple business process models from event logs. In *2017 IEEE International Conference on Data Mining (ICDM)* (pp. 1-10). IEEE.

[2] D. Brons, R. Scheepens, and D. Fahland, “Striking a new Balance in Accuracy and Simplicity with the Probabilistic Inductive Miner,” in *2021 3rd International Conference on Process Mining (ICPM)*, Oct. 2021, pp. 32–39. doi: 10.1109/ICPM53251.2021.9576864.

[3] D. Sommers, V. Menkovski, and D. Fahland, “Process Discovery Using Graph Neural Networks,” in *2021 3rd International Conference on Process Mining (ICPM)*, Oct. 2021, pp. 40–47. doi: 10.1109/ICPM53251.2021.9576849.

For all tracks:

For the team assignment, you are required **to work in teams of four people**. Your final assignment submission should contain a pdf report of at most 12 pages, including figures and references, using the LNCS format (https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines) specified by Springer (available for both LaTeX and MS Word). A latex template is provided via the Blackboard.

By the end of the course, you will submit your final report for the assignment on Blackboard, and it will be graded. The document (and possible appendices) should adhere to the rules and standards of the LNCS format.

Note that your assignment submission may differ from the actual BPIC submissions. For example, it might be shorter and more condensed than the challenge submission. It may also focus on more advanced process techniques and analyses. However, it should adhere to the following focus, which is stated as "*the originality of the results, the validity of the claims, and the depth of the analysis of specific issues identified. We expect participants can focus on a specific aspect of interest and analyze this aspect in great detail*".

2. Assignment Deliverables

Deliverable 0: Form teams and create a team (April 30th at 17:00)

Download the zip provided on Blackboard that contains the latex template for creating a team picture. Take a team picture and fill in the TODOs.

Submit your draft paper in the Blackboard before the deadline.

Deliverable 1: Draft of the assignment topic (May 23rd at 17:00)

For this deliverable, you will write a document **of max. 8 pages length.**

For the process analytics, you report two or three business questions you have identified. For each question, you sketch process analysis and the steps you plan to perform to answer this question.

For the business questions, you may get inspired by the list of questions on the challenge home page (see the links provided in the introduction) to choose which topic/question to address in the challenge. Besides these, as the challenge description states, "we encourage the participants to come up with new and interesting insights". If you can think of other interesting questions that can be answered by analyzing the data, please don't hesitate to suggest them in your draft paper and discuss them with the lecturer/instructor.

You are also encouraged to apply any kind of process analysis approach to the data, not only process mining in a classical sense like control flow discovery. If, e.g., you can base your analysis on a good data visualization derived from the provided event data, go ahead. In case of doubt whether a topic is appropriate to work on, consult the lecturer/instructor.

Specific requirements for the Process Analytics track

- In the draft paper, propose three business questions from process stakeholders' perspective. For each of the questions:
    - Describe the research problem or business question briefly -
    - Motivate the question by discussing possible benefits that would be achieved by answering this question.

    - Argue a set of factors (for example, KPIs, patterns, and variables; maximal three) that are important, considering the business question asked.
    - Sketch how you would compute and analyze these factors using the concepts, analysis, or techniques you have learned during the course. (It is sufficient to make plausible that a technique or combination of techniques would work, no details are necessary)
    - Draft how the results would be presented (e.g., using mockup figures)

Specific requirements for the Process Research track:

- In the draft paper:
    - discuss the research problem,
    - discuss the related work,
    - describe the research method, and
    - provide some early results if possible.

Specific requirements for the Process Discovery Contest track

- In the draft paper:
    - Discuss the limitation of the related existing process discovery algorithms.
    - Describe the idea of the new or improved discovery algorithm
- Provide a working prototype implementation of your discovery algorithm

General requirements for deliverable 1

- The page limit for the draft papers is 8 including references. The appendices do not account for this total.
- The report should follow the LNCS/LNBIP format (https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines) specified by Springer (available for both LaTeX and MS Word).
- The document format should be PDF.

Submit your draft paper in the Blackboard before the deadline.

Please note that failing to finalize this assignment on time will result in the deduction of one point per day from your final assignment grade, starting on the day after the deadline. Furthermore, late submissions may not be included in the peer review process.

You must receive at least a sufficient for the draft paper to pass to this course.

Please try to hand in your papers as complete as possible so you can get useful feedback from your fellow students. In case you are in doubt, please get in touch with your assignment supervisor.

Deliverable 2: Peer review (May 30th at 17:00)

For this deliverable, **each course participant** will review one draft report from another team that is assigned to you. The reviews are intended to help these teams to improve their assignment concerning the final assignment report they will create.

When you have obtained a proper report to review, you need to provide textual feedback (i.e., remarks and suggestions for improvement) based on the criteria listed here below. Read them carefully to understand the specifics of where to pay attention to when reviewing.

Requirements for deliverable 2 – Peer review

- The review should contain the following:
  - summarizing the analysis problems described in the draft paper
  - discussing 2-3 positive aspects of the draft paper
  - discussing 2-3 weak aspects of the draft paper **and providing constructive feedback and concrete suggestions for each weak aspect**.

- For the positive or weak aspects of the draft paper, you may consider using the following subjects to discuss:
  - whether a particular analysis problem or question is appropriately described.
  - whether the potential benefits of a solution to the analysis problem are presented properly.
  - how clearly it is explained with which technique or setup an analysis solution is to be achieved.
  - if the way of presenting an analysis solution is appropriately sketched (e.g., by an example visualization).
  - compliments or concerns you have when considering the feasibility, novelty, or other aspects of a proposed analysis.
  - Any other things that your classmates did particularly well or could improve (e.g., structure, introduction, data description, references, etc.).

- You should submit your review via the FeedbackFruits assessment on the “Submissions” page on Blackboard.

Each course participant must submit your personal review individually in the Blackboard before the deadline. Failing to submit your review in time will result in an insufficient grade for this deliverable, which may cause you to fail the course.

As soon as the deadline for providing feedback has passed, you will be able to access the feedback you received for your paper via Blackboard.

Deliverable 3: Final presentation (June 6th or 11th at 13:00)

Each team will present about their assignment results. The presentation will take 10-15 minutes, after which there will be a 5-minutes Q&A session.

The presentation should contain the following elements:

- Introduction, including the motivation of the selected theme and the main article (practical problems, theoretical problems, existing gap(s))
    - Background information or research context
- Business questions / Research questions / Process discovery
- Related work (only for Track 2 and 3)
- The process analysis performed / the research methods / or your process discovery algorithm
- Preliminary, interesting results, discussing the key findings and implications
- Conclusion

The team submits a pdf-version of the presentation to Blackboard before the presentation. The rubrics of the presentation will be made available via Blackboard.

Deliverable 4: Final assignment report (June 18th at 17:00)

General requirements for deliverable 1

- The style of the report paper must follow the LNCS Springer template.
- The document format should be PDF.
- The page limit is 16. The list of appendices does not account for this total, for example, you can submit a paper of 16 pages including references excluding appendices.

Specific requirements for the Process Analytics track

- Have performed the two or three process analyses you sketched in the draft paper.
- Have discussed the results and derived concrete recommendations based on the insights obtained.

Specific requirements for the Process Research track

- Have conducted the research proposed in the draft paper.
- Have obtained results, discussed the results, and derived a conclusion.

Specific requirements for the Process Discovery Contest track

- Have implemented the process discovery algorithm
- Have evaluated the process discovery algorithm using the PDC 2023 logs.

Submit your final report in the Blackboard before the deadline.
Failing to finalize the report on time will result in the deduction of one point per day from your paper grade, starting on the day after the deadline.

To get acquainted with the writing style of BPI Challenge reports, look at the submissions available from other editions of the challenge:
https://icpmconference.org/2019/icpm-2019/contests-challenges/bpi-challenge-2019/#the_student_category (BPI Challenge ‘19)
https://www.win.tue.nl/bpi/doku.php?id=2018:challenge#the_student_category (BPI Challenge ‘18)
https://www.win.tue.nl/bpi/doku.php?id=2017:challenge#all_submissions (BPI Challenge ‘17)

Fraud and Plagiarism

The final submissions will be checked for plagiarism. We expect you to adhere to the Utrecht Code of Conduct and the UU Academic policies and procedures. Failure to adhere to the code of conduct and our policies may result in penalties, up to and including automatic failure in the course and reference to the Examination Committee. You may learn from the BPIC submissions or cite statements from the BPIC14 submissions to support your findings. Copying and pasting text from such submissions or any other sources that are not your intellectual property, without using quotation marks and referring to the source will be considered as plagiarism. Please have a careful look at https://students.uu.nl/en/practical-information/policies-and-procedures/fraud-and-plagiarism.

Disclaimer

The contents of this assignment are subject to adaptation during the running course. Any specific negotiations with the lecturer may supersede the regulations stated in this document.

Changelog

Version	Changes