052412 VU Business Intelligence II (2025W) --- Group Project

This document describes the goal and the setup of the group project that will be conducted as part of the Business Intelligence II course, which accounts for 35% of your final course grade. Your project will focus on replication of process mining research, as described in detail below.

Group composition: The assignment will be conducted in groups of about 3-5 students. You are free to form your own groups. You can use the Moodle forum to find a group or additional members or you can also register as an individual or incomplete group, asking to be assigned with others. To keep things manageable, a maximum of 8 groups will be established.

Key dates: The timeline for the group assignment is as follows:

- **Thursday December 11:** registration & paper selection deadline (23:59 CET)
- **Thursday January 22:** Group project presentations

Project Tasks

Your group will select a recent paper on process mining, try to replicate the reported evaluation results and, aim to gain further insights into the results.

Paper selection: As a basis for your replication study, you are free to select any paper that meets the following requirements:

1. It is a research paper accepted for the main research track (no Forum, workshop, or demo track papers) for one of the following venues in 2023-2025:
    - The *International Conference on Business Process Management (BPM)* (Foundations or Engineering sub-track)
    - The *International Conference on Advanced Information Systems Engineering (CAISE)*
    - The *International Conference on Process Mining (ICPM)*

   Tip: use DBLP to get the lists of publications, e.g., for the research track of BPM 2025 https://dblp.org/db/conf/bpm/bpm2025.html
2. The paper proposes a process mining approach/technique/algorithm, i.e., any method that takes an event log (or event stream) as input.
3. The paper should have an available implementation and at least part of the data is publicly available.

If in doubt about the suitability of a paper, just ask. **Once you have a paper in mind, list it on the wiki on Moodle ("Paper selection")!** Every group should have a different paper; first group to enter a paper on the wiki can use it.

Note for deep learning/LLM papers: Unfortunately, we cannot provide compute resources or funding to use OpenAI etc. Therefore, you need to select a paper for which you have the means available to run their experiments.

Project goal: The goal of the project is twofold:

1. **Replication:** First, you will try to replicate the evaluation results reported in the paper. This involves getting the provided implementation to work, obtaining the data, and analyzing and comparing the results to what has been reported in the paper.
   Important: It is sufficient to replicate the results on one machine, you do not have to be able to replicate on all machines of all group members.

2. **Additional investigation:** Depending on the replication success, do the following:
   a. **If replication failed:** If you were not able to replicate the reported results, dive deeper into the differences between your findings and what was reported. Try to understand what caused this, when the results differ, and if there are steps that can be taken to come closer to the reported results.
   b. **If replication was successful:** Expand upon the evaluation that has been reported in the paper. For this, you will further explore the performance of the approach, for instance by doing one or more of the following:
      i. Applying the approach on additional data and checking the results for that.
      ii. Computing additional evaluation measures that can provide additional insights.
      iii. Analyzing the results in more detail, e.g., what factors influence the technique's performance across different logs, etc.
      iv. Compare the approach against additional baselines.

Contacting authors: If you have questions about the implementation or are struggling to get it to run, you can reach out to the paper's authors via e-mail (generally the first author is the best to contact). However, be sure to be polite in your requests and understand that researchers are not professional developers. Furthermore, be patient in expecting answers.

Deliverables

Before December 11: Make sure you have set up your group, have selected a paper, and that the basic requirements are met in terms of having data to use for the replication and that you are able to run the code.

Presentation (January 22): You will present your work in a 15-minute presentation, followed by a brief Q&A. Some pointers:

- **Brevity: Time is limited!** Keep the number of slides small and the slides themselves clear and concise. Use appropriate visualizations, focus on concrete bullet-point statements, rather than full paragraphs.
- **Presenters:** It's best to have at most two presenters who present on behalf of the entire group.
- **Attendance:** Attendance is mandatory for the entire group for the entire session, everyone in the group should be able to answer questions regarding the project.
- **Repository link:** For us to review the resources you used to replicate the results (code, raw data, etc.), provide a link to a single repository in your presentation (see below)

Your presentation should cover the following aspects:

1. **Paper overview:** Summarize the paper that you selected. Describe the context of the work, the problem it addresses, how it tackles this problem, and its novelty. You must be able to explain these aspects in your own words, without getting lost in technical terms! Your fellow students should be able to understand you. It is highly preferable to have illustrations here.
2. **Reported evaluation:** Describe how the authors evaluated their work.
3. **Your replication experience:** Describe your experiences while trying to replicate the evaluation results. Including describing what was provided by the authors in terms of implementation? Which tools, code, data were needed?
4. **Show your success:** Show that you managed to get the approach running and what output you obtained.
5. **Replication results:** Present the insights of your replication study.
6. **Additional insights:** Present The insights obtained during your additional investigation.
7. **Conclusion:** Conclude with a brief reflection on what you did and what could have still be done.
8. **A slide that explains the distribution of work in the group** (no need to present this)

Repository: Your presentation should contain a link to a repository (preferably GitHub/Gitlab) that should contain the following:

- **Step-by-step instructions** on how to obtain the results that you report.
- **Necessary code** for the replication and any download links for code/libraries/data.
- **Raw results** (only if relevant, e.g., more detailed results than shown in the slides, not yet aggregated results, …).
- **System information** on which machine(s) you ran the replication experiment on.

There are no further deliverables beyond your presentation, the corresponding slides, and the repository. Make sure to submit a PDF of your slides after the presentation session via Moodle.

Grading

You can receive a maximum of 35 points for your project. Your project will be assessed based on the clarity and contents of all aspects outlined above. Particular attention will be paid to:

- Your ability to describe and your understanding of the paper's research motivation, goals, and novelty.
- The experiments you conducted and the clarity and validity of the insights you drew from them.
- Your ability to answer questions about your project.

The actual success of your replications, i.e., **whether you were able to obtain the results reported in the paper, does not affect your grade**. This aspect is out of your control; it is your task to investigate how well the paper is reproducible.

Questions & Group Issues

For questions, the easiest way is to ask me in-person around the Thursday sessions. Otherwise, please reach out to me via e-mail.

In case of group-related issues, reach out to as soon as possible! I cannot help you if I only hear about issues close to or after the deadline.