

Assignment Predictive Process Monitoring

Group 9

Contact Person: Remco Dijkman, r.m.dijkman@tue.nl

Important Dates:

- 1st office hour: 28th October, 9:30-10:00h online via Teams, send your questions no later than 28th October, 08:00h
- 2nd office hour: 18th November, 9:30-10:00h online via Teams, send your questions and presentation draft no later than 18th November, 08:00h
- Hand-in presentation and implementation: 24th November, 23:59h
- Presentation: 28th November, 13:45-14:30h (Group A)

Instructions for Deliverables:

- Implementation:
 - Programming language is Python 3.x, follow common best practices (check, e.g., <https://peps.python.org/pep-0008/>) and use ideally object-oriented structure
 - README.md with installation instructions and instructions how to run your code <https://tilburgsciencehub.com/topics/collaborate-share/share-your-work/content-creation/readme-best-practices/>
 - requirements.txt containing all necessary dependencies https://pip.pypa.io/en/stable/user_guide/
 - Deductions possible for too messy code (variable and function names not self-explanatory, structure unclear, etc.), wrong or incomplete information in README.md and/or requirements.txt
- Presentation: follow the template provided in Canvas

Topic Description:

Predictive Process Monitoring is about predicting what is likely to happen next in a business process: what the likely next activity is, when that activity will be ready, or what the outcome of the process will be.

There have been various specific improvements to predictive process monitoring over the past few years. In this assignment, you will reproduce the one that is proposed by Senderovich, di Francescomarino, and Maggi (2019), which uses inter-case features to improve predictive process monitoring results. While traditional predictive process monitoring predicts results for a single case, using only features from that case, algorithms that use inter-case features also use features from other cases or from the

process as a whole (such as how many cases are in the system at a particular moment).

Your main goal is to implement the improved algorithm and evaluate whether the algorithm with inter-case features performs better than the algorithm without inter-case features.

The paper presents multiple variants of the proposed algorithm, and tests different settings on multiple datasets, with different features. You do not have to reproduce all variants, settings, datasets, and sets of features. Keep your main goal in mind, which is: evaluating the effectiveness of using inter-case features for predictive process monitoring. In particular, we recommend using only one of the datasets: the one that is publicly available. Also, the paper presents many formal definitions. You do not have to understand and implement all of these definitions in detail. Instead, focus on understanding and reproducing the essence of the paper.

Note that in your presentation towards your fellow students, you have to focus on presenting the algorithm in such a way that they can understand it. Your precise method for evaluation then becomes less important.

Dataset(s) for Evaluation:

D. Levy. Production Analysis with Process Mining Technology. Eindhoven University of Technology, 2014, doi:10.4121/uuid:68726926-5ac5-4fab-b873-ee76ea412399

References:

Arik Senderovich, Chiara Di Francescomarino, Fabrizio Maria Maggi. From knowledge-driven to data-driven inter-case feature encoding in predictive process monitoring. Information Systems. Volume 84, 2019, Pages 255-264, doi:10.1016/j.is.2019.01.007.