*Homework 4: Predictive Process Monitoring*

Li Merila, 03/05/2023

1. (2 points) As part of the log preprocessing, we need to calculate the remaining time for each case in the log. This can be done by finding the difference in seconds between the end timestamp of one event and the end timestamp of the last event in the case. To implement this, a new column called "remtime" should be created in the log, containing this new feature as an event attribute.

In the task, I added the remtime variable indicating the difference between maximum end time and minimum start time in a case in seconds.

2. (2 points) One way to improve the model's precision is by extracting new time contextual features from the timestamps. These features can provide meaningful information to the model about possible seasonal influences on the process behavior. For example, from a start timestamp like "2023-04-18 13:00:00", we can extract the month of the year (e.g., 04), the day of the week (e.g., 2 for Tuesday, where Monday is 0 and Sunday is 6), and the relative time in seconds since midnight (e.g., 46800 for 1:00 PM, as midnight is 0). To implement this, six new columns must be created in the log, containing these three contextual features for the start and complete timestamps. NB! Consider using the weekday() method from Python.

In the task, I added six variables: start time month, end time month, start time weekday, end time weekday, start time from midnight in seconds and end time from midnight in seconds.

3. (3 points) Given the “remtime” column in Task 1 and the contextual features columns in Task 2, train an XGBoost Regressor using single bucketing and aggregation encoding.

In the task,

4. (3 points) Perform Task 3 again, but this time exclude the contextual features. Then, compare the accuracy of the resulting models with the previous step, and explain any differences in the results. Consider whether the use of contextual features impacted the accuracy of the models, and explain why such an effect may or may not have occurred.