Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Business Process Intelligence Assignment 2

This assignment is to be executed individually. The grade for this assignment counts for 20% of the final grade. The deadline for the assignment is July 15, 2018, 23:59 and it is a hard deadline.

There is only this possibility to submit the assignment. Only the final test can be retaken, assignments can only be redone in the next academic year.

The log you obtain is taken from a financial institute. This log contains some 262.200 events in 13.087 cases. Apart from some anonymization, the log contains all data as it came from the financial institute. The process represented in the event log is an application process for a personal loan or overdraft within a global financing organization.

The data owner is interested in all valuable information and especially *estimators for the total cycle time*, *which resources generate the highest activation rate of applications, how does the process model look like, which decisions have great influence on the process flow and where are they taken*.

As you have seen, the log consists of three types of activities:

1. Event names that start with "App_" refer to status changes of *applications*. People make applications for a loan online.
2. Event names that start with "P_" refer to status changes of *proposals*, i.e. if the financial institute sees an opportunity to sell a loan to a customer, a proposal is sent to the customer.
3. Events that start with "W_" refer to activities performed by employees in the process, for example the call-center employees.

During this entire process, there is a constant contact with the applicants through the call center. The events starting with "W_" for *workflow* refer to these contacts.

In this assignment, you are expected to discover the actual structure of the process. The process owner only has a partial picture about how instances of the process are carried out. Roughly, an application submits an application. Then, some automatic checks are performed which may result in the application being declined automatically (for example if the amount is incorrect, or if a so-called BKR check fails). When this happens, the event log records the execution of an instance of activity "App_Rejection".

For this assignment, we are interested to discover how instances of this process are actually being executed. However, the process clearly shows much variability and the activities that are actually executed depend on the final outcome. Therefore, it is not possible to discover a single process model.

Therefore, you are asked to discover different process models:

1. The model of the application's lifecycle
2. The model of the proposal's lifecycle

3. Models that combine these two models into one showing the lifecycle of the application and the proposal together. Due to the high variability, you should discover one different model for each possible outcome, namely whether the application is finally rejected, cancelled or approved.
4. Analyze the C-net of the proposal process and explain it.
5. Based on your analysis, create a Petri net of the proposal process by your own. For example, do not consider infrequent paths and also outlier behaviors.
6. Analyze the performance of Application and work process. What are the bottlenecks? What are your recommendations to the company to increase the performance of the process?

Providing process models is not enough because the process owner is diffident with any result which he/she is simply provided with. Therefore, he/she wants evidence of the goodness of these models. For this purpose, you should perform alignment-based conformance checking. Using the result of the conformance checker, you should motivate why the chosen models are better that any possible alternative (e.g. because of mediating fitness, precision, generalization and simplicity). For each model, also provide its textual description.

In addition to that, the process owner is interested in obtaining answers to the following questions:

1. Are some decisions in any of the models driven by the application's amount?
2. Are there clear indications that the same employees are always involved in cases that are declined, approved or cancelled?
3. The process owner would like to see a comparison of the throughput times between the non-approved applications, i.e. those cancelled or rejected by the applicant, and the approved applications. In particular, he/she is interested in the time between the moment in which an application is submitted and when it is approved, compared with the time between the application is submitted and it is rejected or cancelled.

For this assignment, you have to use ProM. You can use any plug-in and visualizer. It is not allowed to use Disco for this assignment. Any result obtained through Disco will not be graded.

Please note that the event log is extracted from the information system while the system is running. Therefore, the event log contains several incomplete cases.

RWTH AACHEN UNIVERSITY

Report requirements

The assignment should be submitted as a single PDF file named ***StudentNumber*_BPI_assignment_2.pdf** (StudentNumber is your student number) to the OASE folder "Submit_Assignment2" before the deadline. For instance, if your student number is *123456*, the file should be named *123456_BPI_assignment_2.pdf*.

The PDF should report on:

1. The student name and id number.
2. The analysis performed, including the details about how ProM plug-ins have been applied, etc.
3. The findings, accompanied with ProM screenshots of different process models, conformance information, time information, etc. Make sure that you draw sound conclusions.

The report would be be well-structured, start with an introduction and end with a conclusion (keep them concise). The report should not exceed 20 pages of length. Please consider that any part exceeding the 20-page limit will not be graded.

Rules of the assignment

1. For this assignment, you have to use ProM. You can use any plug-in and visualizer. However, it is not allowed to use Disco for this assignment. Any result obtained through Disco will not be considered.
2. Omitting student name and/or id number in the report will incur in a penalization of 1 point.
3. The screenshots are necessary to convincingly support your findings by evidence. Any finding that is not supported by a ProM screenshot will not be considered.
4. A ProM screenshot that is not readable is equivalent to having no screenshot. Hence, the findings supported by that screenshot will not be considered.
5. The text of the assignment has been double checked by multiple persons. If anything is not fully clear, please use the instructions to avoid any doubt. If you make some assumption, please state it in the text. Any assumption is accepted as long as it is considered as reasonable.
6. The structure and quality of the report will also be assessed and graded. Ensure that the report is of sufficient quality.
7. The report should not exceed 20 pages of length. Any part exceeding the 20-page limit will not be graded.