# **Business Process Intelligence – SS 2021**

# Assignment 04 Clustering

First, who are you? Enter your names (as well as Team number (from OLAT) and Matrikelnummer)

* …
* …
* …
* …

This assignment is about clustering. Feel free to use any tool or software of your choice, that supports you in clustering and process mining.

**Assignment**: Given is an event log “CreditRequestEventLog.csv” for requesting credits:

* It has 18 cases with 112 events in total
* **Your task:** Cluster the traces of the event log in order to explain and compare the behavior of traces in different clusters:
  1. Select a trace profile (e.g., activity profile, resource profile, etc.) by your own and compute the measures (number of activities, number of resources, etc. per trace).
  2. Cluster the traces in the event log according to the profile measures and a clustering approach (e.g. k-means). Suggestion: do not experiment with more than k=2 or k=3 clusters, because the event log is quite small and simple
  3. Briefly explain similarities and differences of traces in your clusters. Can you discover and visualize the process models for each cluster?

**Data**:

* The event log is in the file: **CreditRequestEventLog.csv** (see OLAT: Vorlesungsunterlagen / Assignments / 04 Clustering)

For the assignment again use the following answer template:

**Approach we used:**

**What we saw:**

**Our analysis:**

For each of these, fill them in as follows:

***Approach we used:*** *Describe exactly what you did (e.g. which tools did you use, which* *settings in the tools did you change, which plug-ins did you use), in such a way that others can easily reproduce your results. You only need to describe where/if you deviate from the instructions.*

***What we saw:*** *Briefly describe the resulting analysis screen (also includes this as a screenshot), explain, what we see, without including observations or conclusions.*

***Our analysis:*** *What can you conclude from the analysis result? How do you interpret the result? What is the answer to the question?*

Do not write to much, 1-4 sentences for each item are enough. If necessary or helpful, create screenshots and paste them into the document, for example to show configurations or results. Include source code if it supports the understanding of your approach.

# **Clustering**

**Cluster the traces of the event log in order to explain and compare the behavior of traces in different groups**

***Approach we used:*** …

*Document here which tool (for example RapidMiner or Python or anything else) and which configuration (for example the operators in RapidMiner and their parameters or functions in Python) you used. Explain which event log profile you used.*

***What we saw:*** …

*Document your result, i.e., display the clusters and discovered process models.*

***Our analysis:*** …

*What are commonalities and differences of the processes discovered from your clusters?*