Data mining and network analysis ITI8370

Final Project (Exam)

Submission deadline 48h before the defense.

General requirements:

- No plagiarism in any form. Please cite all the sources you used.
- Prepare your solution in such a way that it may be executed on any computer with R-studio.
- Prepare a short write-up with the analysis of achieved results. Maximum 4 pages 12pt. submitted via Taltech Moodle.
- Submit your code by means of https://gitlab.cs.ttu.ee, provide the lecturer and teaching assistant (dmgolo@taltech.ee) with the developer access for your project.
- R codes and functions naming convention. Distance, feature selection and silhouette functions: {student_initials}_nameofthefunction.R. Main codes {student_initials}_ex_{number}_nameofthecode.R. Please avoid using capital letters.
- During the examination you will have to demonstrate your solution and will be asked few questions. Note it is mandatory to attend chosen examination date.
- If you are unsure about using some third party function contact your teacher.
- Report file naming convention DM_2021_[your name]_[your surname]_FP.pdf
- Please indicate (using bold letters in the beginning of your report) if you are willing to present your work in the class or online.
- The initial plan is to have defense in the hybrid mode.
- Please follow ÕIS to see dedicated examination days.

Problem

Choose real life (or close to real life) problem such that it solution would require to use at least three different data mining techniques. It is important to use real life data either acquired as an experiment or from numerous data banks.

Solution

Demonstrate necessity to apply data mining techniques and justify application of all chose techniques. Perform all the necessary steps of data mining, report the results, explain their accuracy and provide interpretation of the achieved results. Be Creative!

Good luck!