Introduction to Algorithmic Data Analysis (JADe)
Autumn 2022
Computing Assignment #1: Frequent Itemset Mining
Initial submission on 20.09.22 at 23:59, then revisions weekly on Tuesdays

UNIVERSITY OF EASTERN FINLAND

Please carefully read and follow the general instructions regarding computing assignments. Failing to meet the requirements might lead to penalties. https://moodle.uef.fi/mod/page/view.php?id=2775059

If you suspect that something is wrong with some task instructions, please contact the lecturer.

If you face persistent issues while working on a task, do ask for help, e.g. during a course meeting or by contacting the lecturer via email.

Datasets

abalone from https://archive.ics.uci.edu/ml/datasets/Abalone

bats from https://www.european-mammals.org/

groceries from https://www.kaggle.com/irfanasrullah/groceries

house from https://archive.ics.uci.edu/ml/datasets/congressional+voting+records

Tools

`fim_resources.py` some potentially useful code snippets for loading and saving datasets from and to different formats.

! Imports of external libraries other than those that appear in the `fim_resources.py` file are not allowed.

Task 1. Implement methods to load and prepare the data (you may use snippets from `fim_resources.py`), and to compute the support of a given itemset. Specifically, you should implement a function which, given a data set as a list of transactions and an itemset, computes the support of the itemset in the data set.

Task 2. Implement a simple level-wise search to extract frequent itemsets given a dataset and minimum support threshold, following the enumeration process presented in the lecture.

Task 3. Experiment with your algorithm on different datasets and with different parameter settings. Compare your results to the results obtained in the Python notebook #1.

Task 4. Implement a procedure for generating association rules given a dataset, a collection of frequent itemsets and minimum confidence threshold.