IN6227 Data Mining – Assignment 2

In this assignment, you will get practical experience working with association rule mining tools. The gist of the assignment is to use one of the available tools for association rule mining that implements Apriori or any other algorithm and compare it to the brute-force approach.

Assignment

Familiarize yourself with any association rule mining tool that you prefer. You can find a list of some of the tools in the links below (but you are also free to choose any other tool you find):

- http://www.butleranalytics.com/6-free-association-rules-mining-tools/
- https://analyticsindiamag.com/6-top-open-source-tools-for-association-rule-mining/

Alternatively, we can search "association rule mining" in **GitHub** to find the relevant open-source code to implement different algorithms (e.g. Apriori, FP-growth, brute-force, etc).

Prepare **several (4–8) datasets of different sizes** (you may find datasets for association rules from Kaggle or any other websites). They can vary in size in either the number of transactions or the number of unique items. Try to isolate the variable, i.e., if you are varying sizes on the number of unique items, make sure the number of transactions doesn't change; if you are varying sizes on the number of transactions, try to maintain the same number of unique items and average width of a transaction.

Run the AR mining program on different datasets and measure the time it takes to complete. You can measure the time for frequent itemset generation and for high-confidence rules generation separately or combined.

Estimate the time that would be required to do the same on a dataset of that size for a brute-force approach. Plot the measured time (using AR mining tools/algorithms) and the estimated time (using brute-force approaches) on the same plot. You may refer to "IN6227 Data Mining Assignment 2 - Reference" for estimating the time of brute-force approaches.

Reporting

Your submission for this assignment is a single PDF file with a report on the assignment. Your report should be exactly **one page**. Somewhere at the top of the page, there should be: your matric number, full name, and a line "IN6227-2023-Assignment-2". The only requirement for report formatting is that it is readable, otherwise you are free to arrange information in any way you prefer.

Submission

Submission should be done in NTULearn. Access the assignment submission page through the left navigation bar by selecting "Assignments". Submit a **single PDF file**. Submissions are accepted up to **Friday, 24th March 2023, 23:59:59**.