LINMA2472 – Algorithms in Data Science

Final Homework

The goal of the last Homework is to pick a dataset of your choice and analyse it with non-trivial algorithms, leading to interesting conclusions and/or observations about the dataset. The algorithms can be taken from this year's course, but not necessarily limited to them.

This dataset can be original (i.e. created by yourself as part of the homework) or taken from elsewhere (e.g. from a Kaggle challenge). For ideas, you can refer to Moodle branch “Sample datasets for the final projects”, where you will find links to interesting datasets.

You are encouraged to be creative, both in the choice of the dataset and the analysis performed on it. `Fun' and `realistic' topics are equally welcome. Once you have a project idea, you must first submit it on Moodle (branch “Dataset submission”). We shall then approve it or suggest changes. Once it is approved you can start working on it.

Please describe your dataset briefly (what it's about, size, where did/will you find it) and the kind of analysis you wish to perform (type of algorithms, type of conclusions, why it's interesting). It doesn’t have to be long, just enough for us to provide feedback/advice on feasibility and suitability for the project. It's not a 'contract', it doesn't commit you to do exactly what you announced.

The final product will be a written report of about 10 pages, and a short 5-mins oral presentation. The oral presentation will take place during the week starting on the 13th of December, on either Monday or Wednesday (more details will follow closer to the date). You will present an overview of the data you use, the algorithms you utilize, and the results you have obtained or plan to obtain. You hence need to be precise and clear in the way you present in order to stay under 5 minutes. The report itself will contain more details about your results and algorithms used. The deadline for the report will be on Sunday the 19th of December at 23h59.

Pay particular attention to your visualisation techniques, keeping in mind what you were taught on Wednesday the 1st of December.

Best of luck!