

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 you are not necessarily limited to them.

This dataset can be original (i.e. created by yourself as part of the homework) or taken from somewhere else (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 (activity "Dataset and outline of final project"). We shall then approve it or suggest changes. Once it is approved you can start working on it.

For this first submission, 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, expected conclusion, 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, and a short 5-mins oral presentation followed by some questions. The report should be no longer than 15 pages, however the report shouldn't be much longer than any of the previous assignment (an excellent report could fit in only 7 pages, so you don't have to go up to 15 pages). The oral presentation will take place during the week starting on the 19th of December, on either Monday or Wednesday (more details will follow closer to the date). You will present an overview of the data and algorithms you use(d), 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 the 5 minutes limit. The report itself will contain all the details about your results and algorithms used. The deadline for the report will be on Wednesday the 23rd of December at 23h59.

Feel free to include a section of your reports about the subjects from the external speakers (ethics, sustainability, visualisation, etc...)