KSD DM Project

Form groups of 2-4 members

You should work out:
- Topic: A general problem area
- Data: What data you will use (make sure you can access it easily)
- Methods: Methods that'll help you address the problem

Then, enter the details in the document linked to on LearnIT.

Groups of 1 will be automatically added to incomplete groups on **November 23**.

The grade will be the same for all members, so get your management, diplomatic, and coaching skills ready.

Due date: **Monday, 10 January 2022, 14:00**

Assignment content:

The report should be 6-8 pages (double column format / standard margins and font), excluding source code but including figures, and should be well written. The report should answer the following questions:

- What knowledge are you trying to extract from what data?
- What methods will you be using?
- Why did you choose this data set and this data?
- How do the algorithms work? (Only a brief overview necessary.)
- How did you extract, store and manage the data?
- Which parameters were used with the algorithms? Did you do any tuning? Why/why not?
- How good are the results, and how valid? What do they really show? (Remember to compare with a baseline, e.g. most common class)
- If applicable (in most cases it is), what is the societal impact and what are the privacy aspects of your work?
- Why did it work / did it not work?

You should also submit the code and data via LearnIT; a link to a CoLab is recommended.

The assignment must be completed in Python. You may use any libraries you like, and are encouraged to use scikit-learn's tools.

The following technical tasks should be undertaken:

    Preprocessing (except if obviously unnecessary)
    Descriptive statistics
    Appropriate visualization
    Appropriate validation of results

You should use algorithms from at least two of the following groups (except if you have a very good reason to use something else, needs pre-approval from me):

    Clustering
    Supervised learning (classification/prediction)

Report templates in both Word and LaTex formats. Templates originate from IEEE (https://www.ieee.org/publications_standards/publications/authors/author_templates.html). You can use the ITU Overleaf access to write your report online.

Please note that the last part of template (with author biography and picture) is *not* required.

https://learnit.itu.dk/pluginfile.php/183789/mod_page/content/3/IEEEtran_LatexTemp.zip

https://learnit.itu.dk/pluginfile.php/183789/mod_page/content/3/IEEEtran_WordTemp.docx