<div align="center">

Final Project

02807 Computational Tools for Data Science

</div>

Deadline for handing in

23:59 on December 2, 2022
To be submitted on DTU Learn

Project Details

In this project, you should demonstrate you are able to use relevant computational tools for data science for a problem of your choice. This project is rather open, so it is important you make a realistic plan for the project such that you will manage to finish in time but at the same time do something that is technically challenging.

Your project must use at least one, preferably two, of the topics we have discussed during the course. Beyond this, it should include at least one new thing that was not directly taught in the lectures, but that you find and understand on your own. In this project, you should:

- Find/make a problem that is relevant within data science and that is technically challenging.

- Motivation/explanation of your problem should be part of the report.

- Analyze the problem and choose relevant tools to solve the problem.

- Implement/use the relevant tools to actually solve the problem.

- Argue clearly for the choices made when designing and developing the solution.

- Write a well-written and concise report about your problem and solution.

Note that it is okay to use existing Python packages for your project, but in order for the project to be considered technically challenging enough for you to receive a good grade, you will likely need to implement some things yourself.

Example high-level project ideas

- Recommendation systems
- Detecting fraud or plagiarism
- Sentiment analysis

Groups

The project work must be completed as a team task. The group/team size is 4 students. Only under rare circumstances will we allow a group size of 3 or 5, which has to be approved by the course responsibles. You must sign up for groups on DTU Learn.

Project report

The report should be 3 pages long. You are allowed to go beyond 3 pages if this is due to including pictures, illustrations, tables etc. However, do not artificially stretch the report by putting information that should be written out into pictures, illustrations, tables etc. In the report you should briefly explain/motivate your problem, present the work you have done, the decisions you have made, report your results, etc. Avoid repetition, convoluted explanations, etc. As an appendix to the report you should submit a Jupyter Notebook that includes your code. The notebook should be well documented. Also, sufficient instruction should be provided for how to use the code. Furthermore, you must include an outline explaining why and how each person did roughly an equal amount of work for the project. This outline can also be part of the appendix. The report together with the appendix must be uploaded on DTU Learn at latest on the 2nd of December.

Gathering data

The project should make use of some source of data. For example you could use Pandas to extract some data from a website. You can also take a look at https://www.kaggle.com, which contains many different types of data sets.

Grading

The project will be graded on the final report and the Jupyter Notebook you hand in. Some key criterias in the evaluation are:

- The problem you solve must be sufficiently technically challenging (i.e. solving a very easy problem well will not give you a high grade.).

- Ideally the problem should also be relevant/interesting to a bigger community. (However, a clearly outlined merely technical problem could also be used.)

- What did you learn/implement additionally on your own? Which technical problems before/during implementation have you had to solve?

- The report must be concise and well-written. Remember to argue for the choices and consideration you have had during the project.