Final Project

02807 Computational Tools for Data Science

Important Dates

- 20:00 October 30, 2018: Project plan submission deadline on peergrade.io.
- 20:00 November 28: Preliminary report submission deadline on peergrade.io.
- 13.00-17.00 December 4, 2018: Project demonstration.
- 20:00 December 7, 2018: Final project submission deadline on DTU Inside.

Project

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.

The project should be based on data from Wikipedia. You can see details on how to obtain data from Wikipedia on `https://en.wikipedia.org/wiki/Wikipedia:Database_download`.

Your project must use at least one of the topics we have discussed during the course or that is in the textbook of 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.
- 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.

If you are in doubt if your problem, chosen tools, etc. is a good fit for this project, please feel free to discuss them with the course staff.

Should you already have a perfect idea for a project that does not entirely follow the rules described in this document, please come and talk with the course responsible (without explicit approval from the course responsible, the rules described in this document must be followed).

Examples

Below are a few high level ideas of what you could do in this project:

- Plagiarism: Construct a tool to check whether a document contains phrases from Wikipedia articles.
- Categorize: See how well you can categorize articles (without using the original categories of course).
- Language guesser: Make a tool that is able to guess what language a phrase is.

Practical information

Groups You have to complete this project in groups of 2-3 students. All other group sizes are **not** allowed (without explicit approval from the course responsible).

Project plan Before the project weeks starts, you must complete a project plan. The project plan should contain a list of the group members, a description of what you have decided to do in your project and a high level timeline indicating how you plan on finish this project in time. The project plan should be approximately one page long.

Report There are no exact requirements to the length of the report. It should be long enough to present the work you have done, the decisions you have made, etc. But not longer than that (ie. no repeating, convoluted explanations, etc).

You must hand-in both a report and your program code as a group on DTU Inside before the submission deadline. You should provide sufficient instructions for us to be able to use the program code in some meaningful way.

Project demonstration On the last Tuesday of the semester (December 4), all groups have to present their projects. Here, you will also have the opportunity to see what the other groups have made. This will be a quite informal event, where you are not required to have prepared slides or a big presentation. You should simply demo and explain what your project is about in 5-10 minutes, and be ready to answer questions from the teachers and other students. In other words, this is your chance to show the world what a cool project you have managed to make.

Peer feedback In order for you to get feedback on your project plan and report, we will be using peer feedback. You should use this opportunity to get feedback on your ideas and on your report in order to improve both. Therefore, you must submit your project plan and preliminary report to peergrade (see deadlines in top). Right after the submission deadline, **you have 5 days to give peer feedback to two other groups – this is mandatory**.

Grading This project will mainly be graded on the final report you hand in. Your peer feedback to the other groups (project plan and preliminary report) will also count in the evaluation (other's feedback on your project will not count in your evaluation). Some key criterias in the evaluation are:

- The problem you solve must be sufficiently technically challenging (ie. solving a very easy problem well will not give you a high grade).
- The report must be concise and well-written. Remember to argue for the choices and consideration you have had during the project.

Help During the project weeks our teachers and teaching assistants will be available and ready to help you with your project as normally on Tuesdays. There will generally not be any lectures during this period.