DAT405 Introduction to Data Science and AI

2021-2022, Reading Period 1

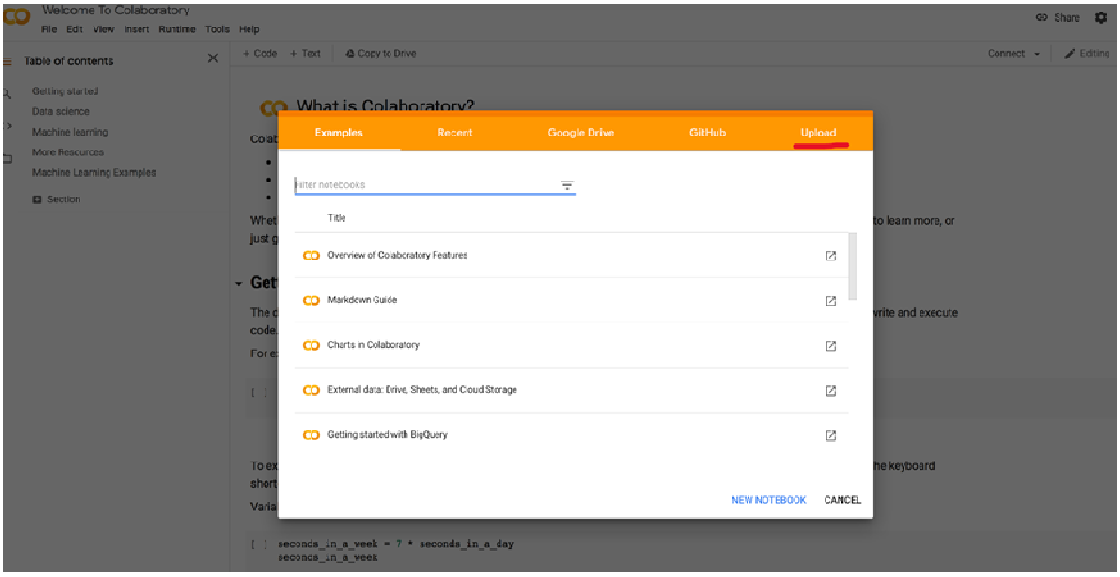
Assignment 4: Spam classification using Naïve Bayes

Practical details

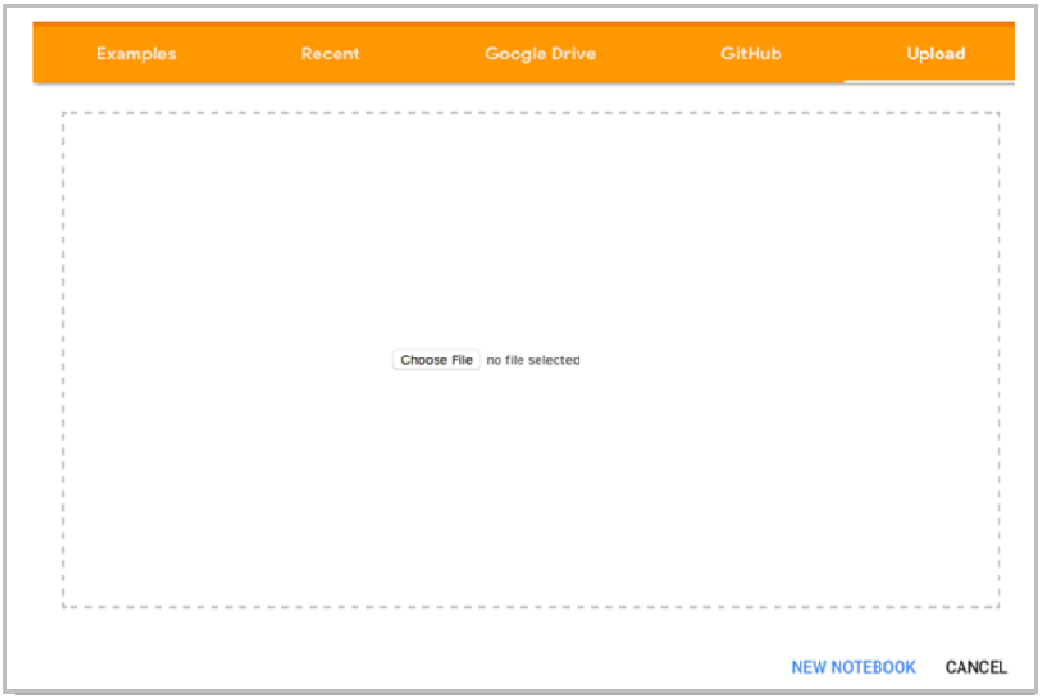
In this assignment, you will work with the Naïve Bayes algorithm using the *scikit-learn* library. The assignment takes place in a notebook environment – all questions and general information about the assignment is provided in the notebook itself, in the **Assignment4.ipynb file**. You are free to continue working in Jupyter on your own computer and submitting **pre-executed notebooks to Canvas**. However, we strongly recommend using **Google Colab** for this and the coming assignments as it allows you to collaborate in groups more easily while applying social distancing.

To use **Google Colab**, you will need to have a Google account – as for Gmail, Google Docs, etc.

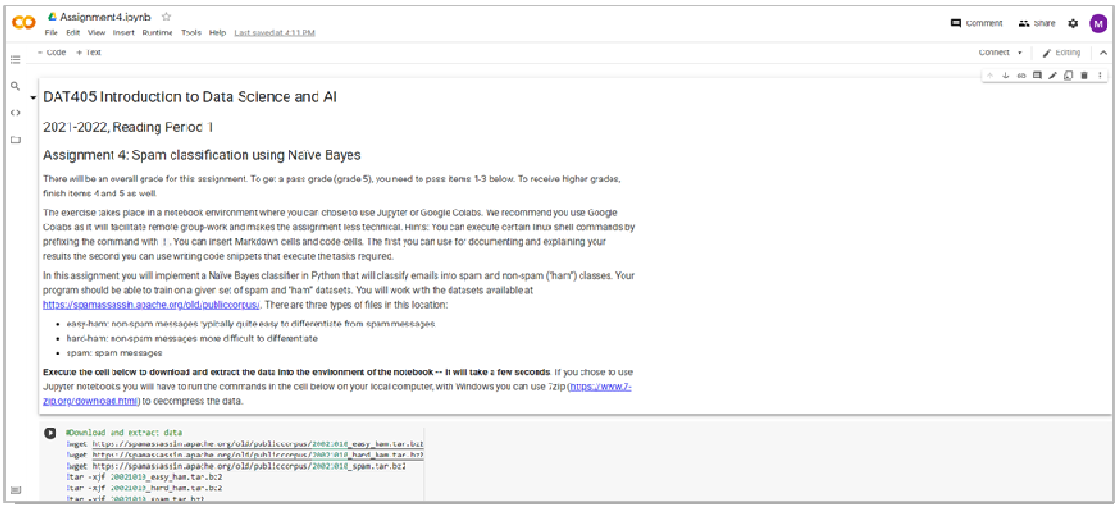
- Make sure you have downloaded the Assignment4.ipynb file from Canvas.
- Then go to http://colab.research.google.com/ in your browser. You will be prompted with a screen which looks something like this



- Press Upload in the top-right corner (underlined in the screenshot above) and the following screen should appear.



- Select Assignment4.ipynb on your local computer and upload it. Now the assignment should be open in your Google Colab environment, and look something like this:



You use this exactly as you use a Jupyter notebook, and all the libraries you need for the assignment are already pre-installed. Cool, huh?

Now you can share the notebook with your team-partner via the share button in the upper right corner of the notebook.

An option for submitting your assignment is to share a link to your completed notebook in Canvas.

Submitting the assignment for grading

In general, when submitting your reports, please make sure all results and discussions are clearly visible and readable. **That means, notebooks should be executed, and all code output should be visible and readable**. In Google Colab, first go to the Runtime menu and select Factory Reset-Runtime and then go to the Runtime again and select Run all.

You have the following options to submit the assignment:

- Submit a link to a completed and fully executed Google Colab notebook (please make sure it is executable and editable for anybody with the link).
- Submit a completed and fully executed Jupyter notebook (.ipynb-file) from Colab or in Jupyter.

Deadline: Sept 27, 2021 at 23:59.

Grading

Grading will be based on a qualitative assessment of each assignment. It is important to:

- Present clear arguments.
- Present the results in a clear and pedagogical way, for instance using tables or plots.
- Show understanding of the topics, by reasoning around your method and results.
- Give correct solutions.
- Make sure that the code is well structured and commented.