Data Mining och Data Science Laboratory instructions

View and interact with Jupyter notebooks and learn about data mining techniques.

Prior Knowledge
Data Mining and Data Science lectures.

Software Requirements
The notebooks are available in the course GitHub repository:
https://github.com/UppsalaIM/2IS063

Binder is a hosted service that takes care of all the installation for you, and runs the notebooks on the cloud. For running on the cloud using Binder you will need only:

- A Web browser: Google Chrome or Mozilla Firefox.

For running on your computer you will need to install:

- Python 3.7 or newer
- Jupyter
- Extra Python packages: `pandas`, `textmining3`, `wordcloud`, `matplotlib`, `mlxtend`, `graphviz`, `scikit-learn`, `okpy`
- Graphviz binary installation.

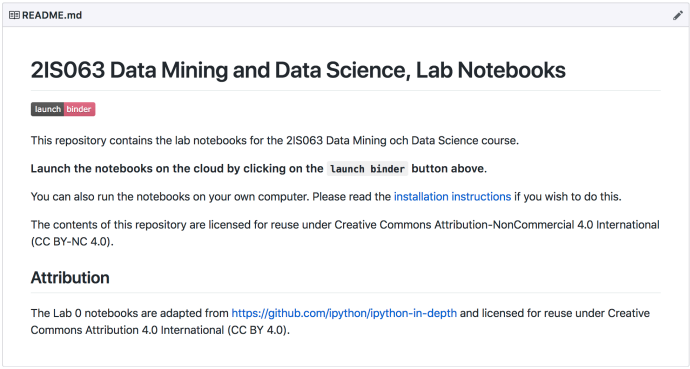
Full installation instructions can be found in repository documentation (visit the URL given above).

Starting the Labs
The laboratory sessions for this course are run in Jupyter.

Jupyter is an environment that is web based, and allows you to do interactive programming inside a Web browser. Jupyter allows you to view and create computational notebooks, which are like Web pages that contain cells that are static content and cells that are Python code that you can run and view the output in the browser.

1. Visit the GitHub source code repository for the Labs at
   https://github.com/UppsalaIM/2IS063

   Scroll down and you should see some content that looks like this:


README.md

2IS063 Data Mining and Data Science, Lab Notebooks

launch binder

This repository contains the lab notebooks for the 2IS063 Data Mining och Data Science course.

Launch the notebooks on the cloud by clicking on the `launch binder` button above.

You can also run the notebooks on your own computer. Please read the installation instructions if you wish to do this.

The contents of this repository are licensed for reuse under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

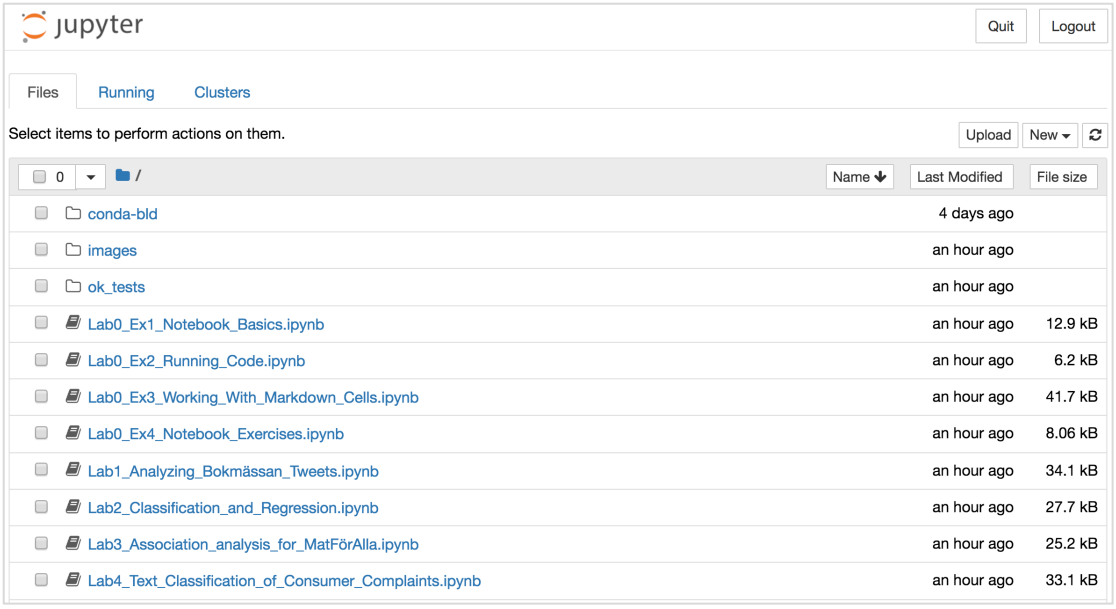
Attribution

The Lab 0 notebooks are adapted from https://github.com/ipython/ipython-in-depth and licensed for reuse under Creative Commons Attribution 4.0 International (CC BY 4.0).

There are two ways to launch the notebooks repository. The simplest way is to use the **`launch binder`** button. If you installed Jupyter on to your own computer, launch Jupyter from the command-line with the command

```
jupyter notebook
```

or if you are using Anaconda launch Jupyter from the Anaconda Navigator app. Once launched, you should see the repository contents in the Jupyter dashboard like this:


jupyter

Quit Logout

Files Running Clusters

Select items to perform actions on them.

Upload New

0 /	Name	Last Modified	File size
conda-bld		4 days ago	
images		an hour ago	
ok_tests		an hour ago	
Lab0_Ex1_Notebook_Basics.ipynb		an hour ago	12.9 kB
Lab0_Ex2_Running_Code.ipynb		an hour ago	6.2 kB
Lab0_Ex3_Working_With_Markdown_Cells.ipynb		an hour ago	41.7 kB
Lab0_Ex4_Notebook_Exercises.ipynb		an hour ago	8.06 kB
Lab1_Analyzing_Bokmässan_Tweets.ipynb		an hour ago	34.1 kB
Lab2_Classification_and_Regression.ipynb		an hour ago	27.7 kB
Lab3_Association_analysis_for_MatFörAlla.ipynb		an hour ago	25.2 kB
Lab4_Text_Classification_of_Consumer_Complaints.ipynb		an hour ago	33.1 kB

To launch a specific notebook, click on any of the **`.ipynb`** files in the dashboard.

2. If you have not already done so, make sure to familiarize yourself with the Jupyter environment by doing the Lab 0 exercises 1-4

   i.e. Open up, read and complete `Lab0_Ex1_Notebook_Basics.ipynb`, `Lab0_Ex2...`, `Lab0_Ex3...` etc.

3. Open the Lab notebook corresponding to the session and complete the questions in the notebook.

   e.g. During Lab 1, open `Lab1_Analyzing_Bökmassan_Tweets.ipynb`, etc.

4. When you are finished, get the attention of a teaching assistant or the lecturer to have a brief discussion about the questions you answered in the notebook.

If you are stuck, do not hesitate to ask for help from one of the teaching assistants (Evelina and Martin) or the lecturer (David)!

Good luck!