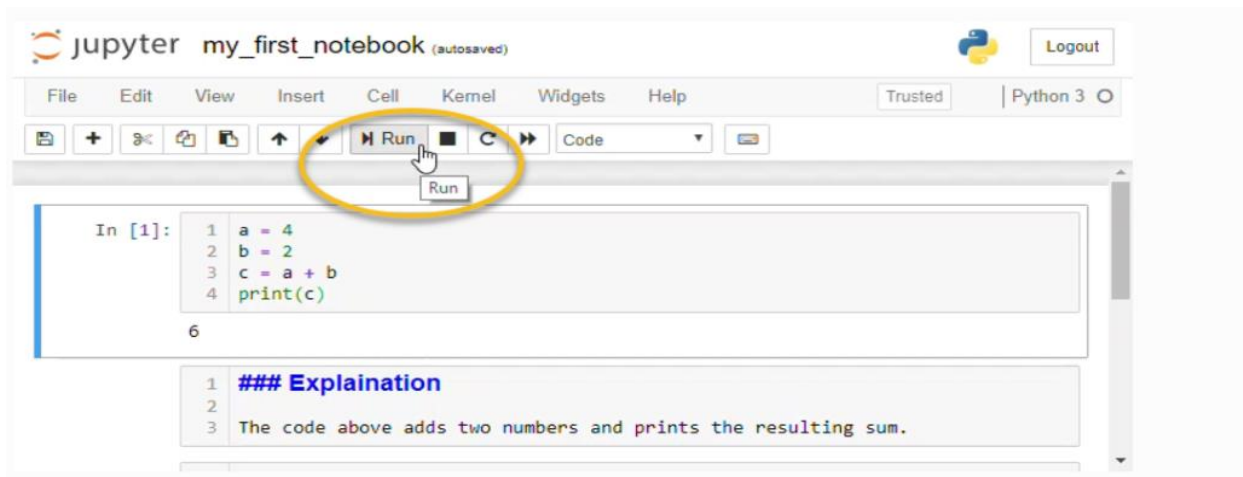


How to run the module?

Throughout this module you will encounter both text and code cells. Please run each cell in this Notebook by clicking "Run" button in the Toolbar or by pushing Shift+Enter keys



How to Download Jupyter Notebook of Each Module and Work on Your Local Computer (e.g., Your Laptop or Desktop Computer)

The environment on which a Jupyter Notebook of each Module running on datamine.unc.edu is very different from a Jupyter Notebook running on your local computer. You may fail to run the Module's Jupyter Notebook on your local computer if you had simply downloaded it from datamine.unc.edu. The following steps will allow you to download the Jupyter Notebook of each Module on your local computer and work on it as you want without worrying about losing the information you have done.

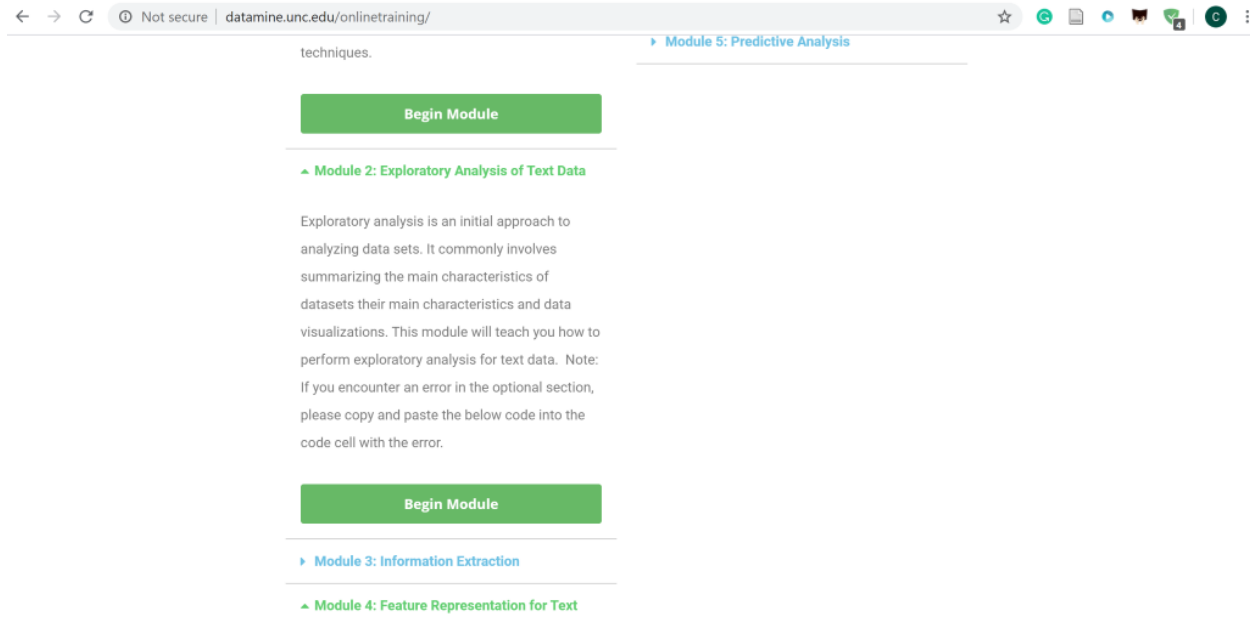
Note: You have to install Python 3.7 and Jupyter Notebook on your local computer. Follow instructions at <https://jupyter.readthedocs.io/en/latest/install.html> to install them.

An alternative solution will be using Anaconda which is available here:
<https://www.anaconda.com/products/individual>

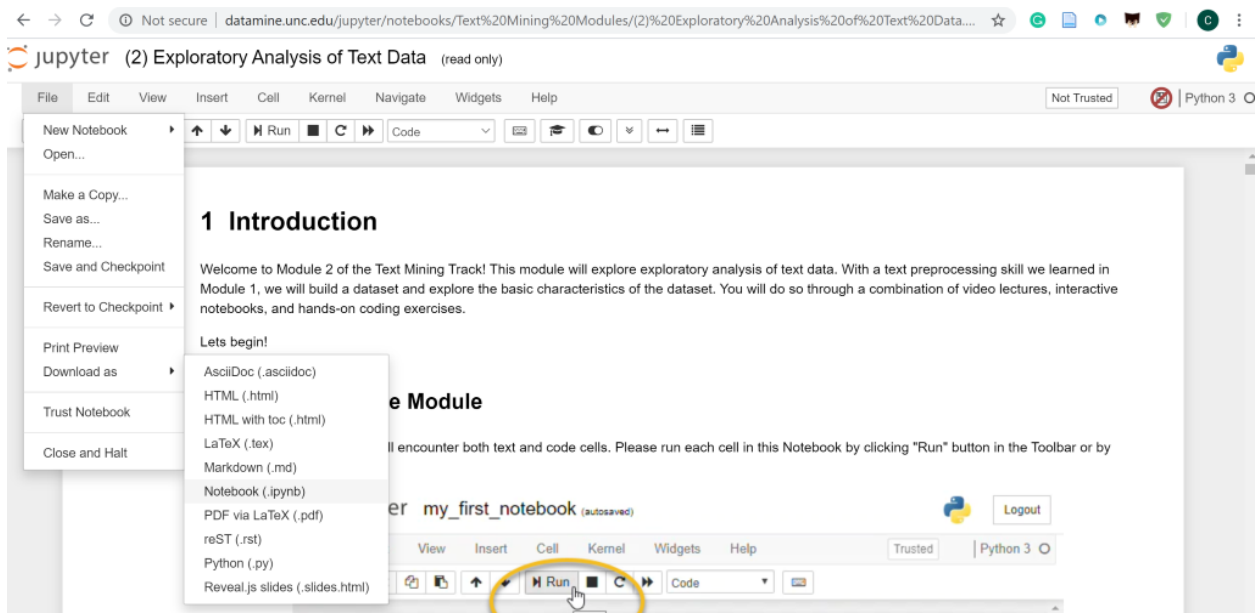
After downloading and installing, from the Anaconda navigator install Jupyter Notebook and you should be good with both python 3.7 and Jupyter notebook runnable on your computer. For setting up your environment, refer to document SettingEnv.docx.

Section 1

1. Begin the Module you want to download at datamine.unc.edu.

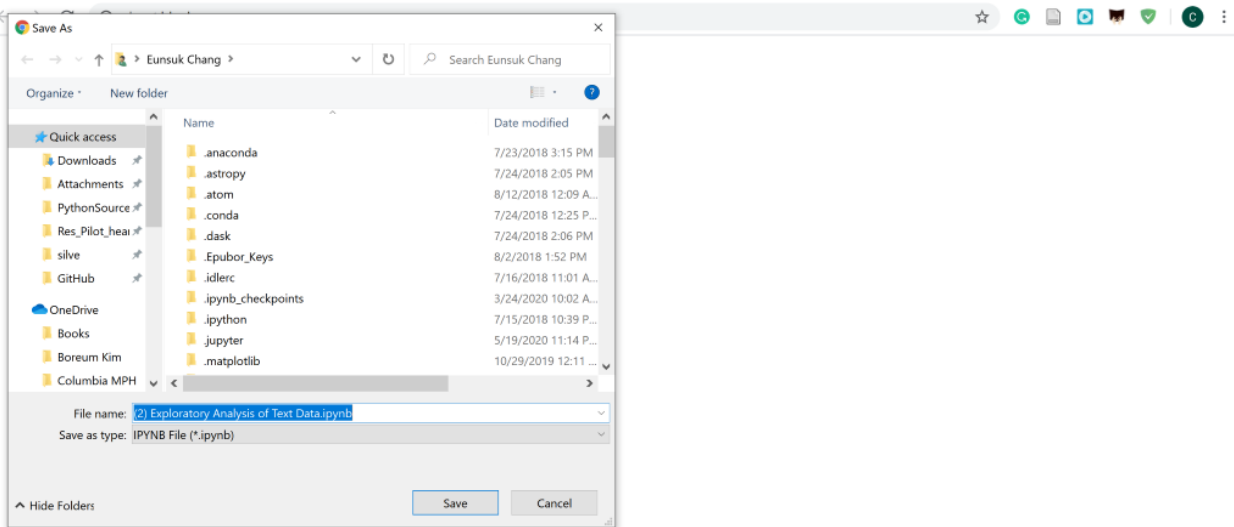


2. Click on File > Download as > Notebook (.ipynb)



3. Download the Notebook in the same directory where Jupyter Notebook was installed.

If II



4. If you do not know where your Jupyter Notebook was installed, open your Jupyter Notebook and try the following code.

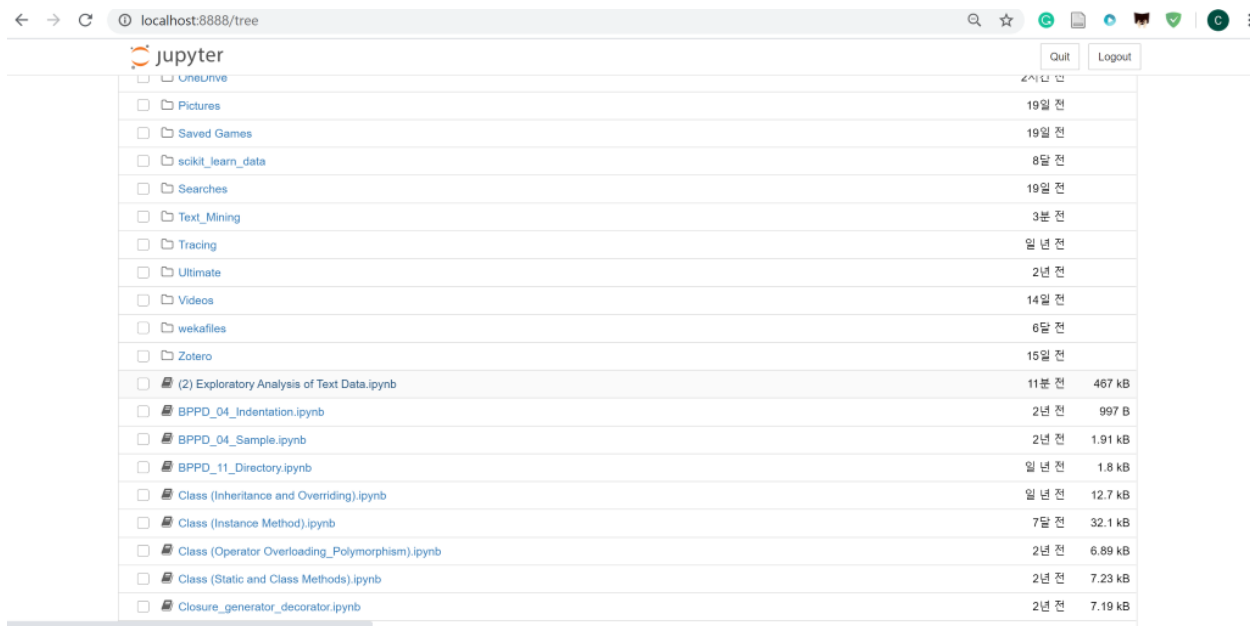
```
In [1]: 1 import os
        2 os.getcwd()

Out[1]: 'C:\\Users\\W\\silve\\Text_Mining'
```

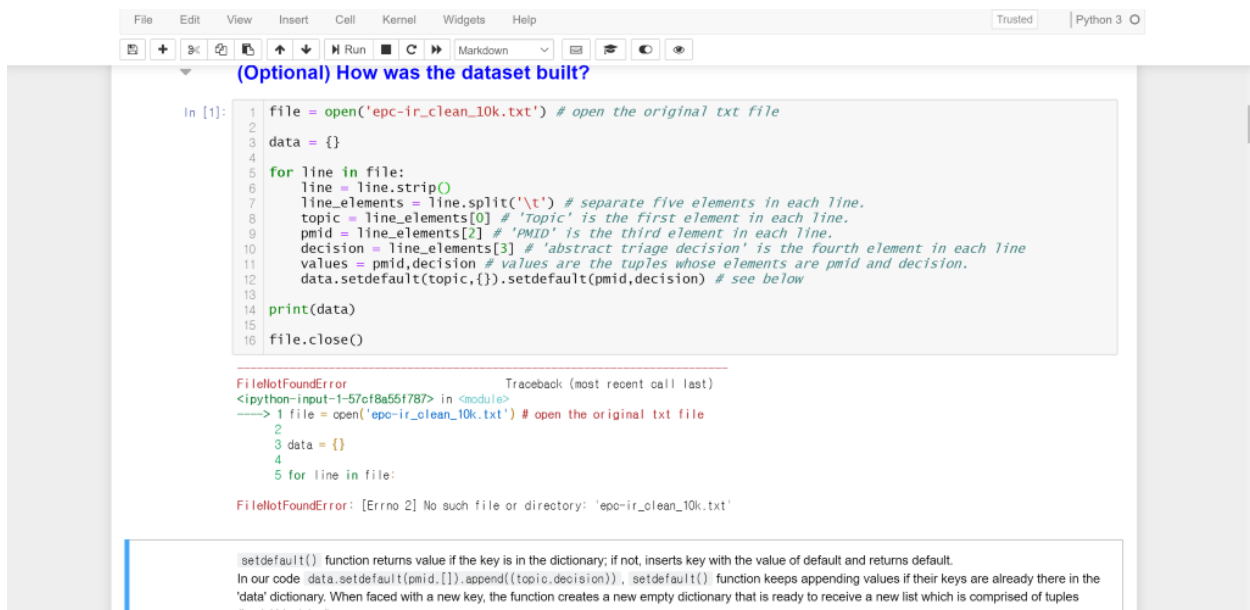
5. Now you just downloaded the Jupyter Notebook of the Module on your local computer.

Name	Date modified
Ultimate	8/20/2018 12:30 P...
Videos	5/20/2020 2:16 PM
wekafiles	11/19/2019 10:53 ...
Zotero	5/20/2020 12:19 A...
(2) Exploratory Analysis of Text Data.ipynb	6/3/2020 2:09 PM
.condarc	7/23/2018 3:17 PM
.gitconfig	5/18/2020 5:58 PM
BPPD_04_Indentation.ipynb	7/15/2018 11:06 P...
BPPD_04_Sample.ipynb	7/17/2018 2:43 PM
BPPD_11_Directory.ipynb	5/25/2019 6:15 PM
Class (Inheritance and Overriding).ipynb	6/30/2019 10:40 A...

6. To run the downloaded Notebook, go to your Jupyter Notebook and locate the Notebook you have just downloaded.



7. Click and run the Notebook. You will see error messages that look like this.



The error message went off because you don't have the data file 'epc-ir_clean_10.txt' on your local computer. You have to download the data file from datamine.unc.edu/jupyter.

8. Go to datamine.unc.edu/jupyter and click on the data file ('epc-ir_clean_10.txt') you want to download.

← → ↻ Not secure | datamine.unc.edu/jupyter/tree/Text%20Mining%20Modules ☆ [extension icons] [profile icon] ⋮

jupyter Quit

Files Running Clusters Nbextensions

Duplicate Rename Move Download View Edit [red icon] Upload New ↺

1 / Text Mining Modules Name Last Modified File size

..	몇 초 전	
<input type="checkbox"/> Data	20시간 전	
<input type="checkbox"/> Pictures	20시간 전	
<input type="checkbox"/> (1) Text Preprocessing.ipynb	Running 20시간 전	174 kB
<input type="checkbox"/> (2) Exploratory Analysis of Text Data.ipynb	Running 20시간 전	467 kB
<input type="checkbox"/> (3) Information Extraction.ipynb	Running 20시간 전	45.9 kB
<input type="checkbox"/> (5) Feature Representation of Text.ipynb	Running 20시간 전	208 kB
<input type="checkbox"/> contractions.py	20시간 전	3.03 kB
<input type="checkbox"/> epc-ir_clean_10k.csv	20시간 전	14.9 MB
<input checked="" type="checkbox"/> epc-ir_clean_10k.txt	20시간 전	314 kB

9. Click on 'Download.'

← → ↻ Not secure | datamine.unc.edu/jupyter/tree/Text%20Mining%20Modules ☆ [extension icons] [profile icon] ⋮

jupyter Quit

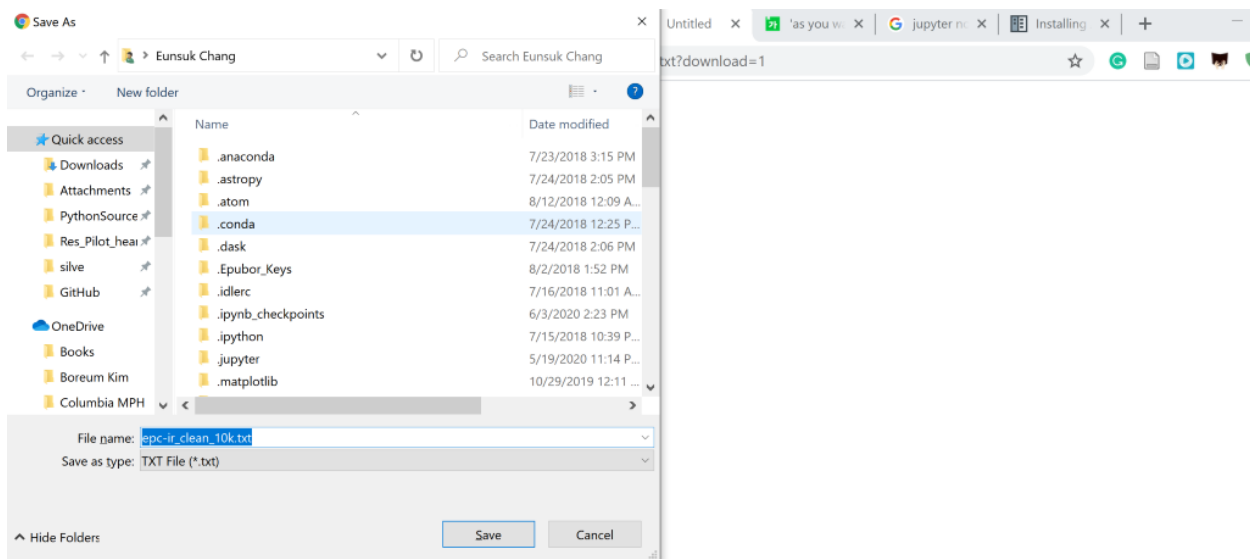
Files Running Clusters Nbextensions

Duplicate Rename Move Download View Edit [red icon] Upload New ↺

1 / Text Mining Modules Name Last Modified File size

..	몇 초 전	
<input type="checkbox"/> Data	20시간 전	
<input type="checkbox"/> Pictures	20시간 전	
<input type="checkbox"/> (1) Text Preprocessing.ipynb	Running 20시간 전	174 kB
<input type="checkbox"/> (2) Exploratory Analysis of Text Data.ipynb	Running 20시간 전	467 kB
<input type="checkbox"/> (3) Information Extraction.ipynb	Running 20시간 전	45.9 kB
<input type="checkbox"/> (5) Feature Representation of Text.ipynb	Running 20시간 전	208 kB
<input type="checkbox"/> contractions.py	20시간 전	3.03 kB
<input type="checkbox"/> epc-ir_clean_10k.csv	20시간 전	14.9 MB
<input checked="" type="checkbox"/> epc-ir_clean_10k.txt	20시간 전	314 kB

10. Download the data file on the same folder your Jupyter Notebook was installed.



11. Go back to your Jupyter Notebook and run the cell again. You will find it is working now.

```
In [2]: 1 file = open('epc-ir_clean_10k.txt') # open the original txt file
        2
        3 data = {}
        4
        5 for line in file:
        6     line = line.strip()
        7     line_elements = line.split('\t') # separate five elements in each line.
        8     topic = line_elements[0] # 'Topic' is the first element in each line.
        9     pmid = line_elements[2] # 'PMID' is the third element in each line.
        10    decision = line_elements[3] # 'abstract triage decision' is the fourth element in each line
        11    values = pmid, decision # values are the tuples whose elements are pmid and decision.
        12    data.setdefault(topic, {}).setdefault(pmid, decision) # see below
        13
        14 print(data)
        15
        16 file.close()

{'ACEinhibitors': {'10024335': 'E', '10027665': 'E', '10027935': 'E', '10028936': 'E', '10029645': 'I', '10029788': 'E', '10030325': 'E', '10047639': 'E', '10049657': 'E', '10051289': 'E', '10053176': 'E', '10063787': 'E', '10066961': 'E', '10067800': 'E', '10069777': 'E', '10072218': 'E', '10073748': 'E', '10073783': 'E', '10073846': 'I', '10075143': 'E', '10075388': 'E', '10077374': 'E', '10078088': 'E', '10080414': 'E', '10080455': 'E', '10080457': 'I', '10082500': 'E', '10083055': 'E', '10086852': 'E', '10088066': 'I', '10089936': 'E', '10090111': 'E', '10090348': 'E', '10090351': 'E', '10091486': 'E', '10091816': 'E', '10091821': 'E', '10093076': 'E', '10093769': 'E', '10095798': 'E', '10097934': 'E', '10099033': 'E', '10099034': 'E', '10099084': 'E', '10099075': 'E', '10099910': 'I', '10100063': 'E', '10100064': 'E', '10100068': 'E', '10100083': 'E', '10100105': 'E', '10130617': 'E', '10150320': 'E', '10150325': 'E', '10155301': 'E', '10164060': 'E', '10169388': 'E', '10170450': 'E', '10172139': 'E', '10178959': 'E', '10190395': 'E', '10190404': 'E', '10190422': 'E', '10190653': 'E', '10192229': 'E', '10192233': 'E', '10192373': 'E', '10193712': 'E', '10193713': 'E', '10195086': 'E', '10204820': 'E', '10209075': 'E', '10211343': 'E', '10216796': 'E', '10218721': 'E', '10218737': 'E', '10219067': 'E', '10220232': 'E', '10220238': 'E', '10220239': 'E', '10220541': 'E', '10221346': 'E', '10226660': 'I', '10228295': 'E', '10229744': 'E', '10232497': 'E', '10232685': 'E', '10235686': 'E', '10321443': 'E', '10321444': 'E', '10323377': 'E', '10326170': 'E', '10326532': 'E', '10327129': 'E', '10329066': 'E', '10333052': 'E', '10333343': 'E', '10334415': 'E', '10335324': 'E', '10335771': 'E', '10335912': 'E', '10336919': 'E', '10338458': 'E', '10338459': 'E', '10340259': 'E', '10344043': 'E', '10344062': 'E', '10347337': 'E', '10348094': 'E', '10350103': 'E', '10356671': 'E', '10361863': 'E', '10362194': 'E', '10362195': 'E', '10362383': 'I', '10367600': 'E', '10368114': 'E', '10369535': 'E', '10369612': 'E', '10371365': 'E', '10373219': 'E', '10374374': 'I', '10376178': 'E', '10376615': 'E', '10376852': 'E', '10377929': 'E', '10379633': 'E'}}
```

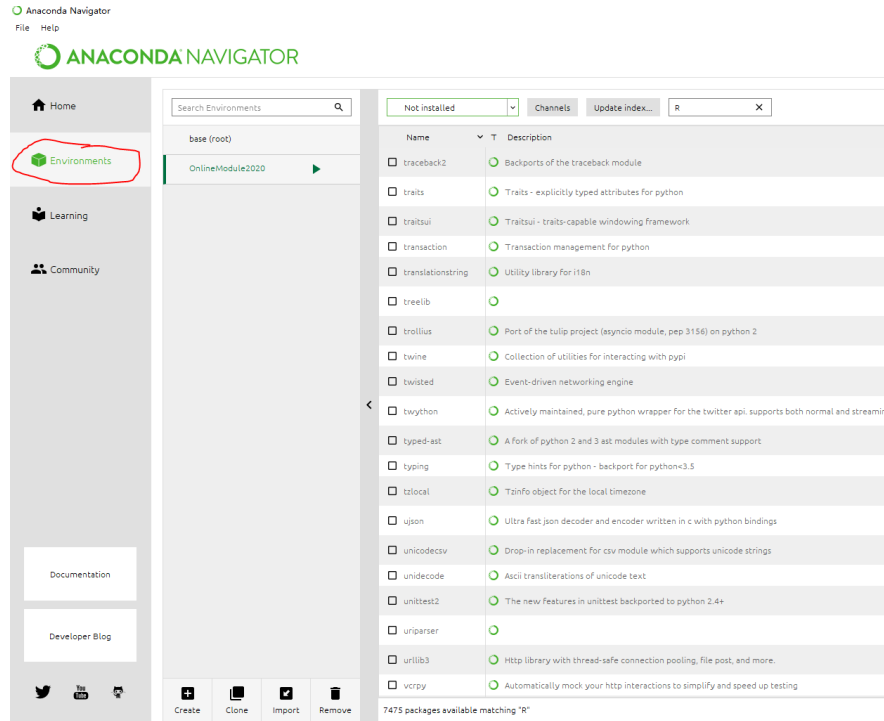
Those steps can be repeated to download other Module Notebooks and data files. But remember,

You should download Module Notebooks and data files on the directory on which your Jupyter Notebook was installed.

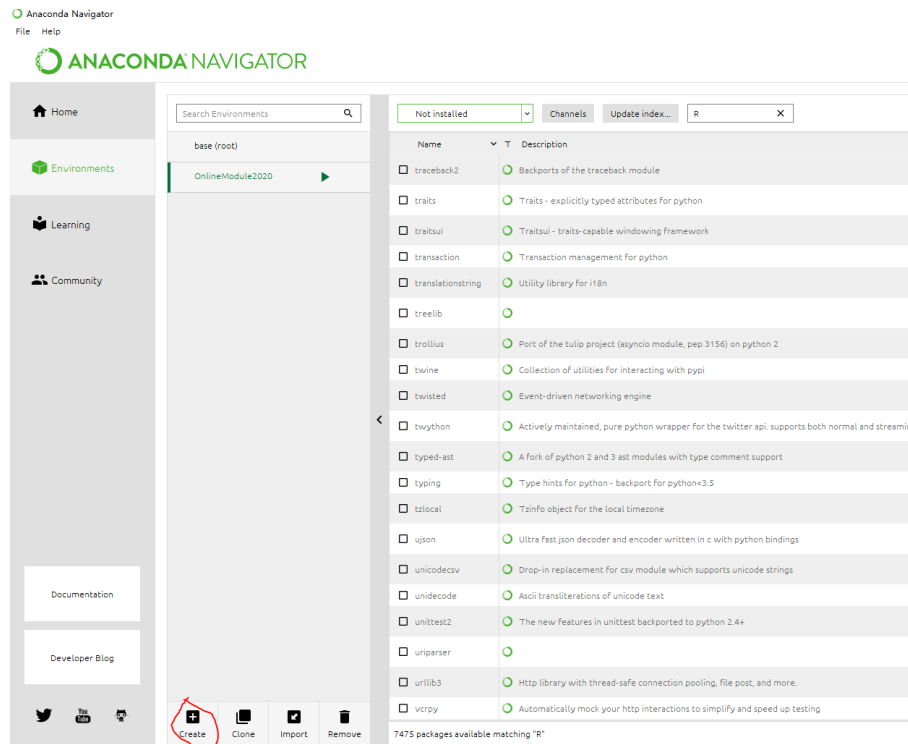
Section2

Following these steps for setting up your environment that works for the online training module.

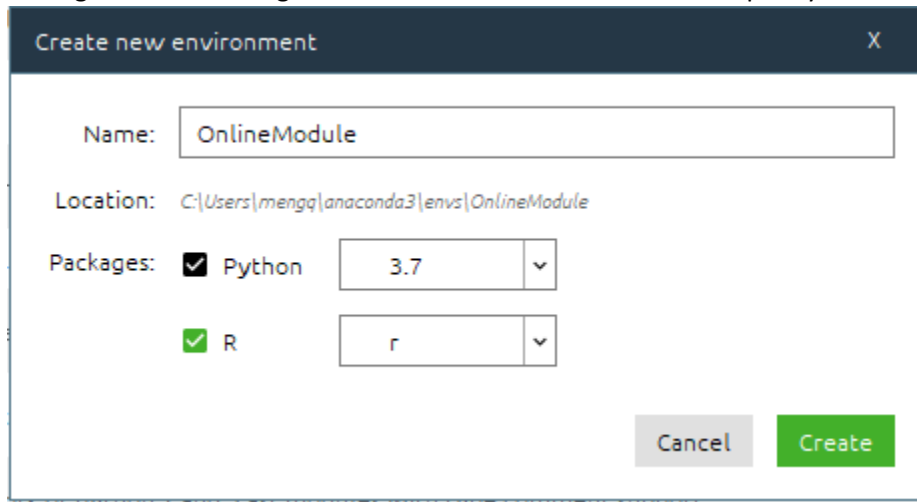
1. From Anaconda Navigator, navigate to “Environment” panel with the button on your left-hand side.



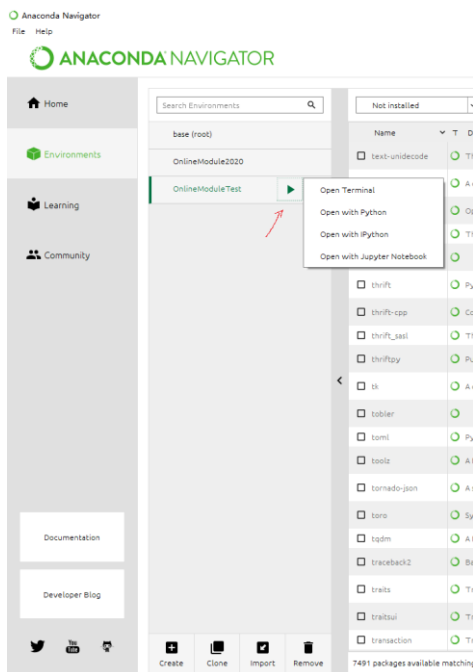
2. Hit “Create” to build a new environment.



3. Give your environment a name, and make sure you select both python and R for both Text mining and Data mining modules. Then click “Create”. This step may take a few mins.



4. To install all required packages, click on the triangle on your environment button, and hit “Open Terminal”



5. In the Terminal window you just opened, type in “pip install <Package Name>” to install the following packages: (e.g. for numpy==1.17.2, do **pip install numpy==1.17.2**)

ipywidgets

jupyter_contrib_nbextensions

numpy==1.17.2

pandas==0.25.1

scipy==1.3.1

scikit-learn==0.21.3

scikit-image==0.14.3

matplotlib==3.1.1

seaborn==0.9.0

statsmodels==0.10.1

pydotplus==2.0.2

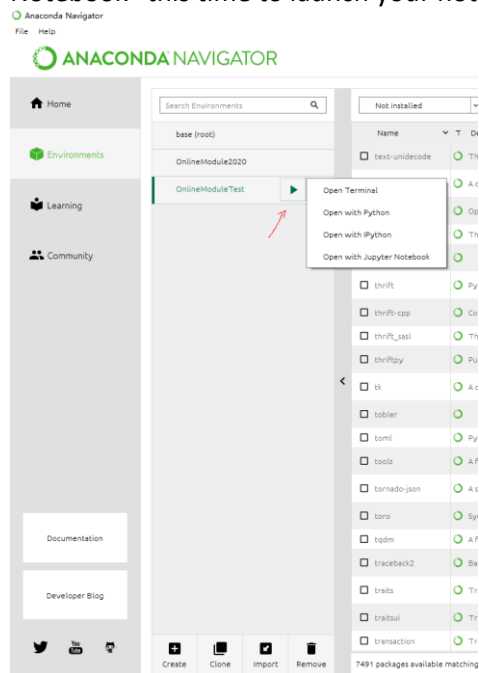
nltk==3.4.5

biopython==1.76

IPython==7.4.0

spacy

6. Now you should be good to go. Hit the little triangle again but click “Open with Jupyter Notebook” this time to launch your notebook.



7. Note that the initial path that your notebook launched from should be under your user folder. i.e. the path shown in your Anaconda terminal when you open it. As shown in the picture, my path is C:\Users\mengq.

```
(OnlineModuleTest) C:\Users\mengq>
```

You could work in this directory if you are good with it. Just create a new folder under this path and download our notebooks to the folder.

8. Otherwise, you can change the startup path through the following steps:

8.1 type in **jupyter notebook --generate-config**

```
(OnlineModuleTest) C:\Users\mengq>jupyter notebook --generate-config  
Writing default config to: C:\Users\mengq\.jupyter\jupyter_notebook_config.py
```

- 8.2 Then type in **open <your path>\.jupyter\jupyter_notebook_config.py** to open the file.
(e.g. in my case it will be **open C:\Users\mengq\.jupyter\jupyter_notebook_config.py**)
You can also navigate to the directory and open the file manually. If your system asks you to choose an application, a simple notebook or any type of text editor should be good.

8.3 In the file, find line **#c.NotebookApp.notebook_dir**

```
## The directory to use for notebooks and kernels.  
#c.NotebookApp.notebook_dir = ''
```

- 8.4 Delete the “#” and add the path you would like to work with in the quote.

```
## The directory to use for notebooks and kernels.  
c.NotebookApp.notebook_dir = 'D:/'
```

- 8.5 Save the file then reopen your Jupyter Notebook, you should be in the new directory now.