INSTRUCTIONS

**Libraries/Modules**

The libraries used were:

* nltk
* datetime
* random
* numpy
* pandas
* csv
* json
* collections
* tensorflow
* sklearn
* keras
* pickle
* pyprind
* matplotlib
* wordcloud
* itertools (used within a function from sklearn that plots confusion matrices)

All are publicly available.

**Dataset**

We used the Yelp academic dataset. Smaller samples of the dataset can be found in train\_Yelp.csv and test\_Yelp.csv.

**Instructions for Files**

First, Move all submitted files into your Python directory.

After installing the libraries above, download the Yelp Academic Dataset used to build our model. Note that the Building\_Model\_PYONNIERS.ipynb file is the only one that requires the full dataset. The other Notebooks do not. We attempted both KNN and Neural Networks. Once all the files have been moved into your Python directory, and all the libraries installed, you may open the Model\_Demo\_PYONNIERS Jupyter Notebook and demo it by typing in the area between the “”s that define the string inside the list “your\_input”

It should give you a rating between 1 – 5 that estimates the sentiment of your statement.

Downloading the Full Dataset

Download the Yelp academic dataset from:

<https://www.yelp.com/dataset>

Clicking the “Download Dataset” button at the center of the page will take you to the following link:

<https://www.yelp.com/dataset/download>

Note the button looks like this:

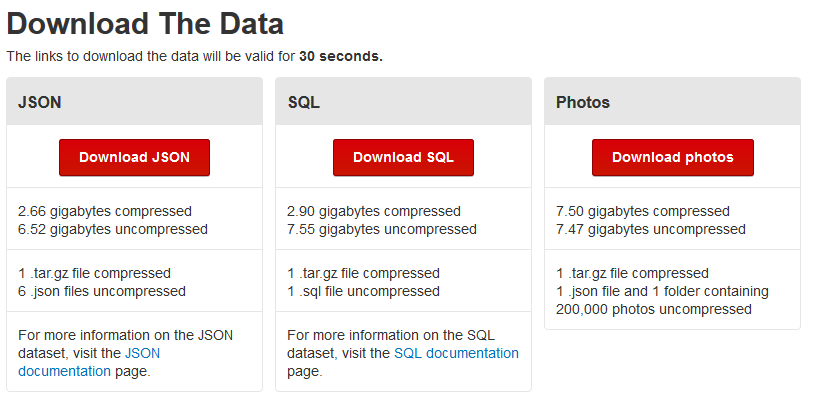


Then once on the webpage, you need to fill out your name and email address, along with signing your initials (need to confirm that you have read the Dataset License.

Click the “Download” button. It should look like this:

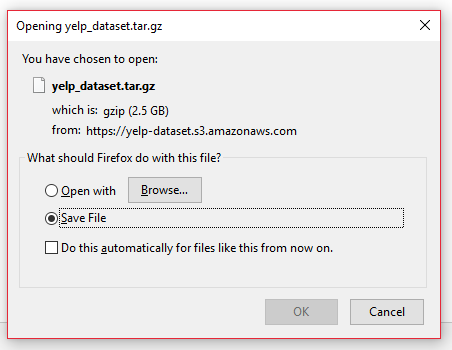


This will take you to a new webpage that looks like this:



Click the “Download JSON” button.

A prompt should pop up that looks like this:



Click on the prompt and click “OK.”

Note that the file is 2.5 GB so you need to ensure you have enough space in your Downloads folder.

The name of the file should be:

yelp\_dataset.tar.gz

Once it has finished downloading, use an unzipping program, such as 7-Zip (<https://www.7-zip.org/>) to open the file.

Opening yelp\_dataset.tar.gz will show a file named:

yelp\_dataset.tar

Opening that will show a folder called:

dataset

Opening that will show a number of files, one of them being:

review.json

Copy and Paste review.json into your Python directory.

Rename the review.json file to:

reviewMarch2018.json

**File Descriptions**

The Building\_Model\_PYONNIERS file shows the code used to build our final model. Note that it requires the entire dataset that can be acquired by following the instructions above.

The Model\_Demo\_PYONNIERS file is what we use to demo our model. It requires the PYONNIERS\_model.h5 file and PYONNIERS\_tokenizer\_5m.pickle file to be in your Python directory.

The Positive and Negative Words xlsx file contains the Positive words used for KNN in the first sheet and the Negative words used for KNN in the second sheet.

The PYONNIERS\_KNN\_Attempt file shows our attempt at KNN. It requires the test\_Yelp.csv and train\_Yelp.csv files to be in your Python directory.

The PYONNIERS\_model.h5 file contains our final model. The Building\_Model\_PYONNIERS file creates a new model and rewrites this file.

The PYONNIERS\_tokenier\_5m.pickle file contains our Tokenizer for the Neural Network.

The test\_Yelp\_csv file and train\_Yelp.csv files contain our train and test sets for KNN.