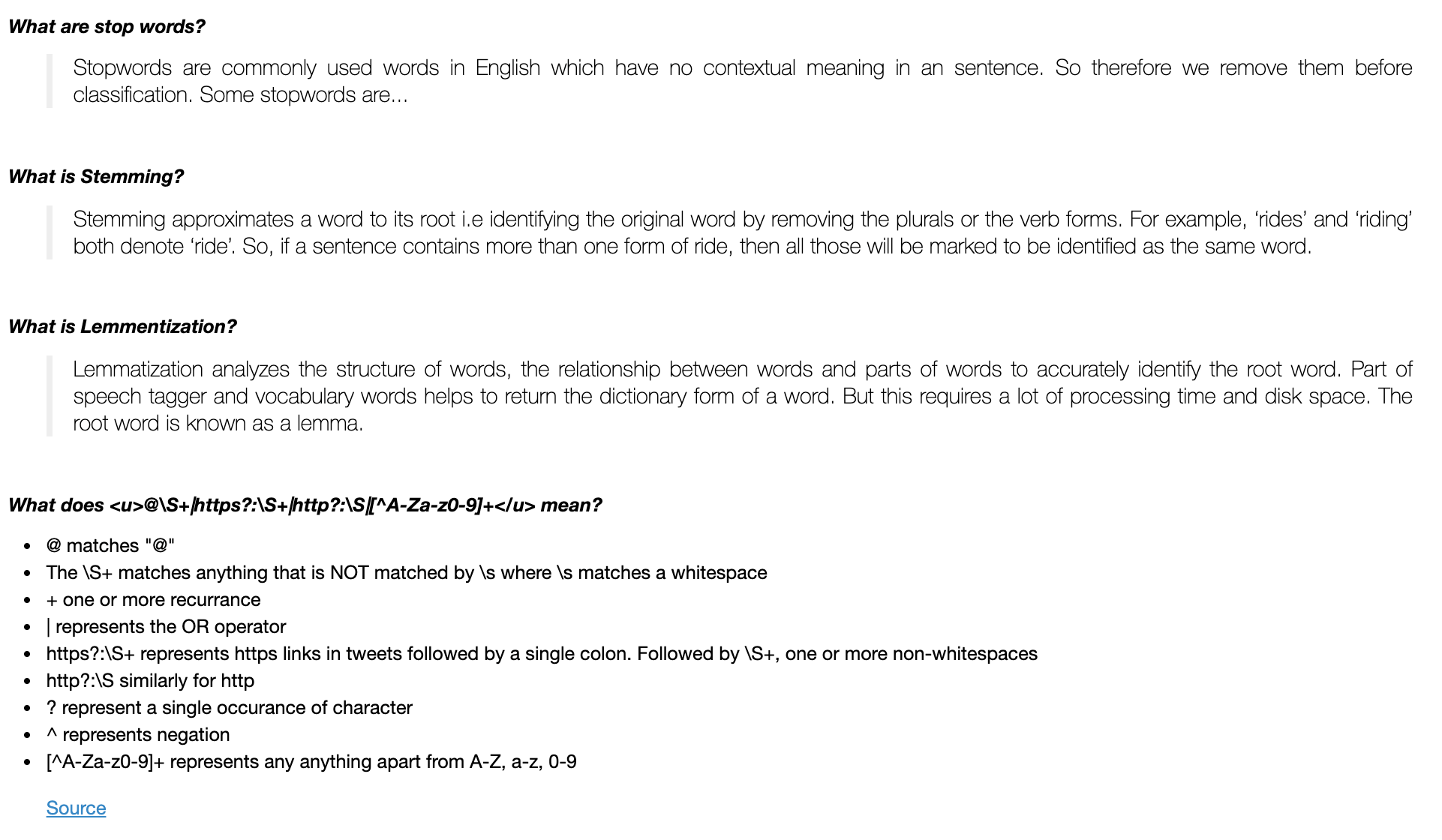
Capstone Project 3: Milestone Report II

As per Milestone Report I, we have performed a initial analysis on the data and explored it as well

Now, we will classify the tweets into positive or negative using a neural network model.

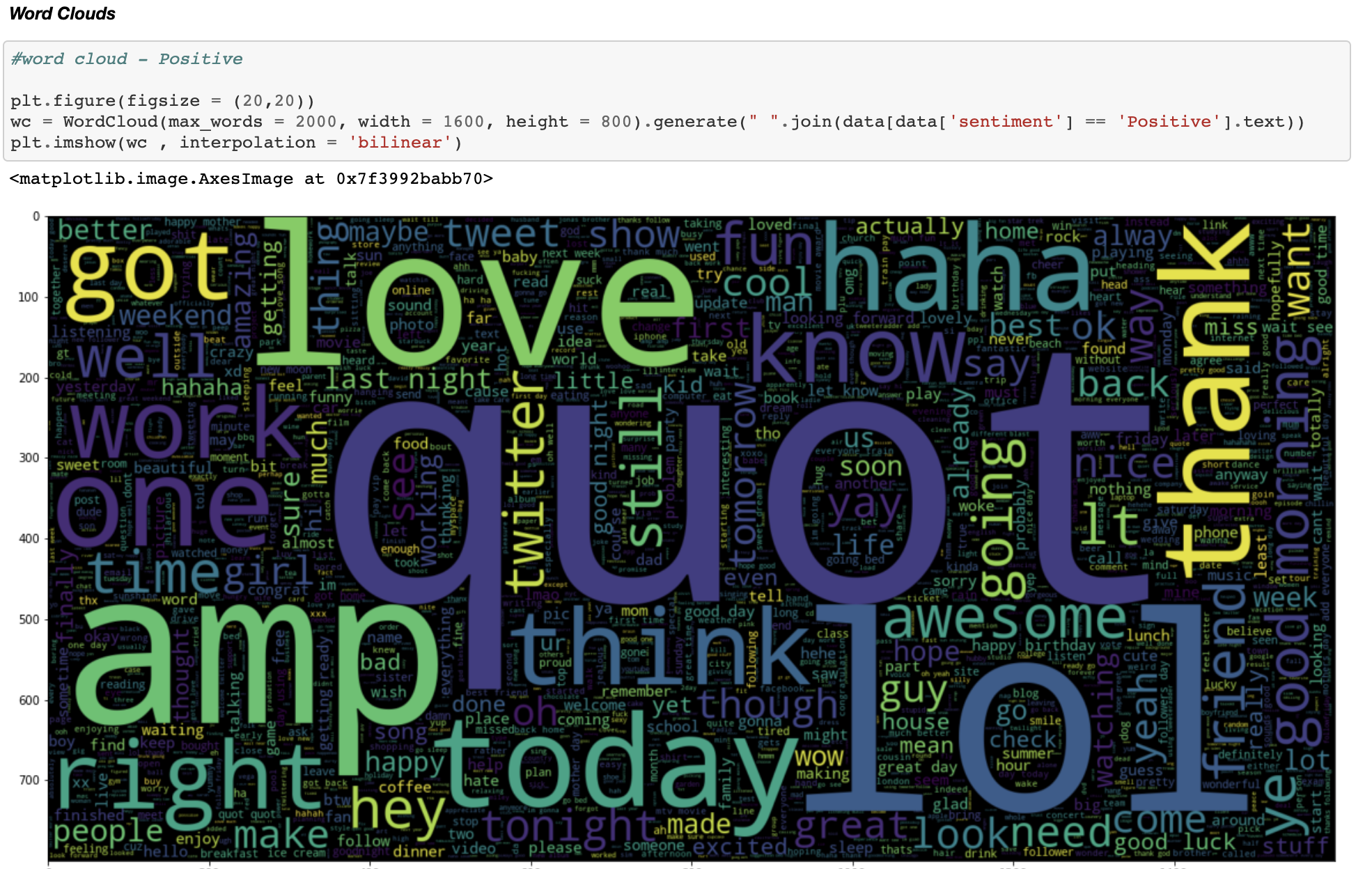
Before that we have to preprocess the data

* Stemming
* Removing unwanted characters through regex
* Stop Words removal

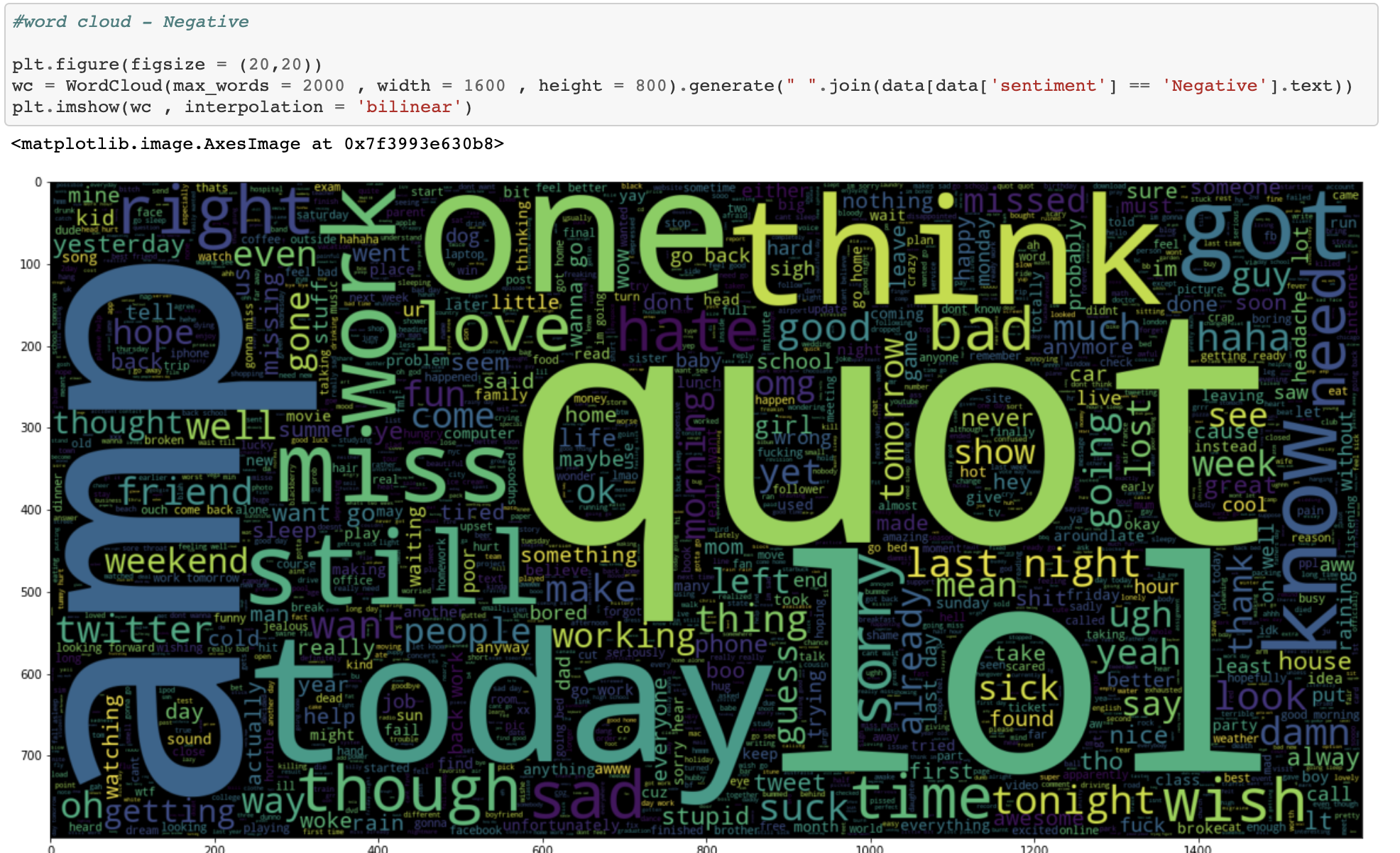


Once the above steps are completed, we can visualize the positive and negative tweets using word cloud.

Positive Word Cloud



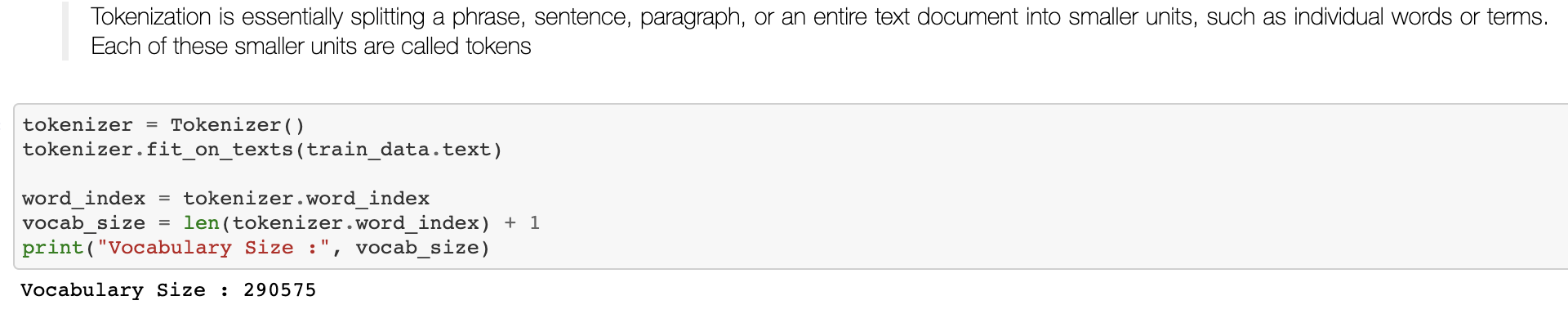
Negative Word Cloud



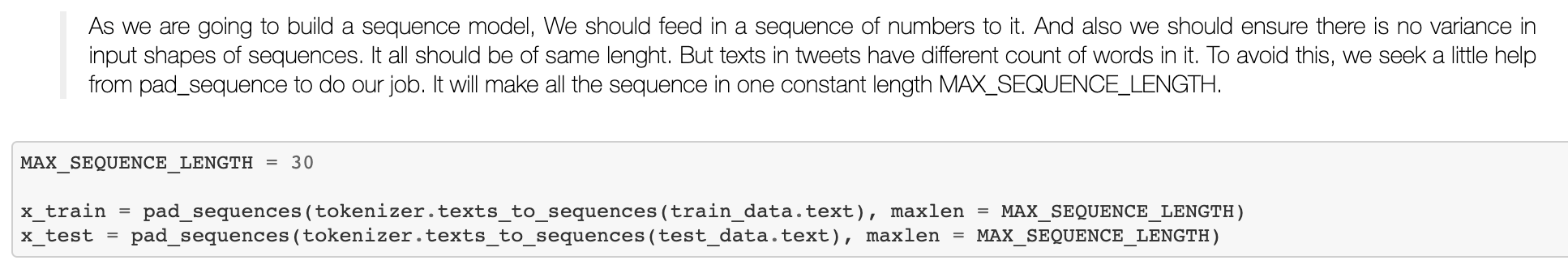
Test-Train Split:

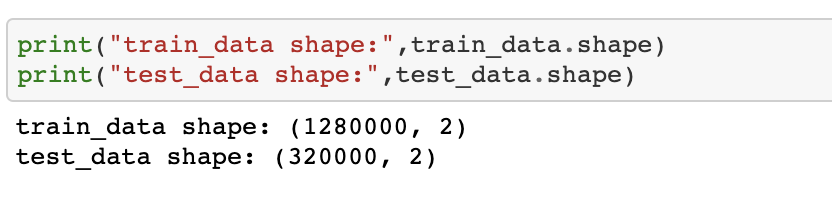


Tokenization: Over here we extract the word index for the corresponding words in the document

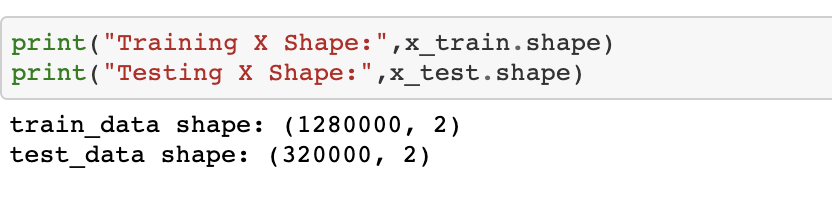


Pad Sequence: This is done in order to build a sequence model with a standard input shape

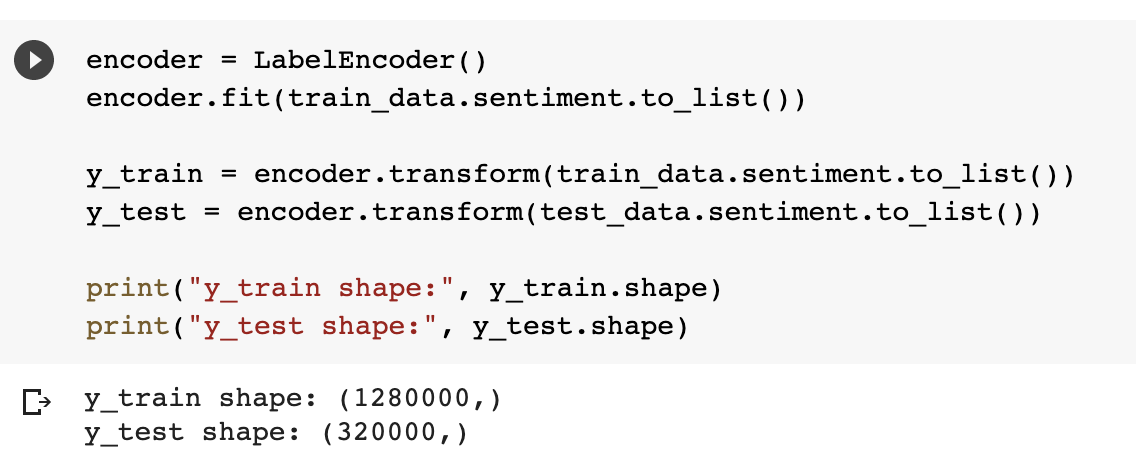


Shape before padding:

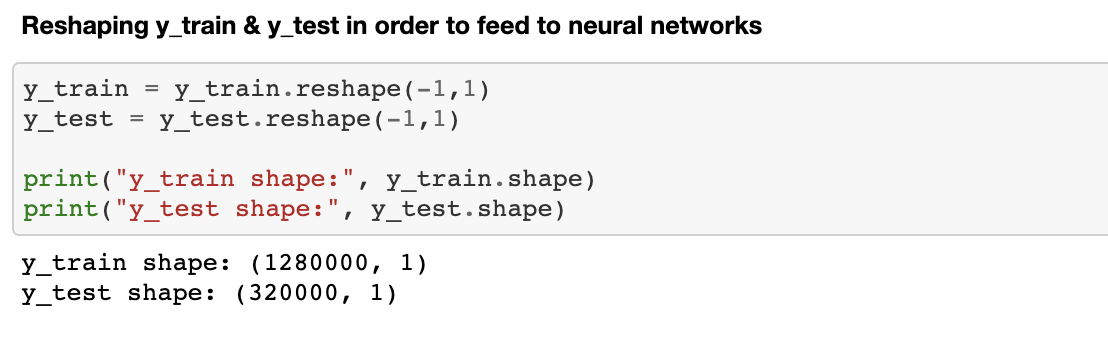
Shape after padding:



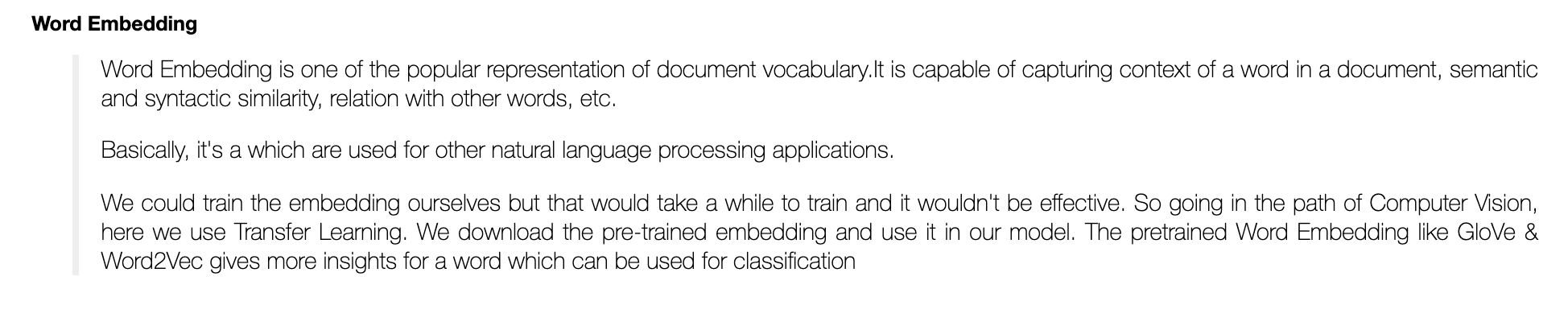
Label Encoder:



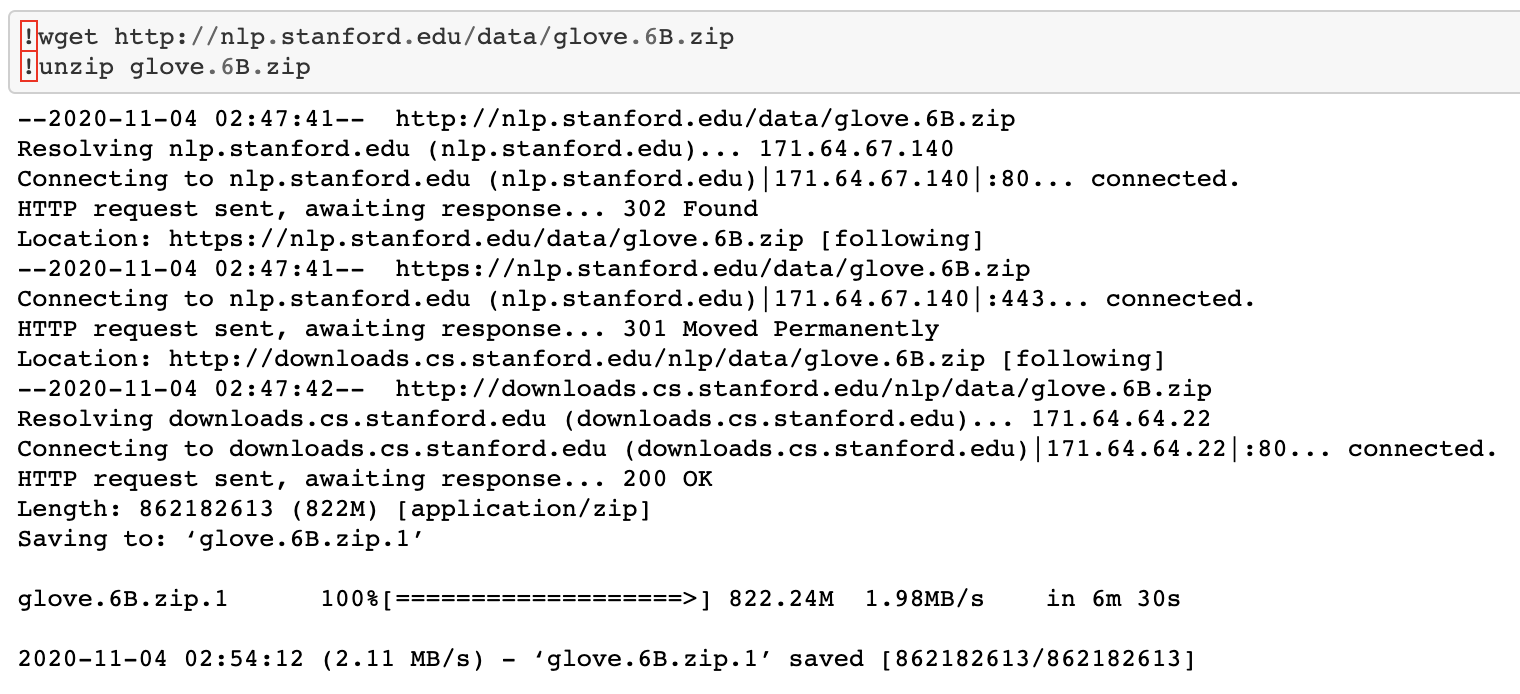
Reshaping:



Word Embedding: There are 2 types – Glove & word2vec. We will be using Glove



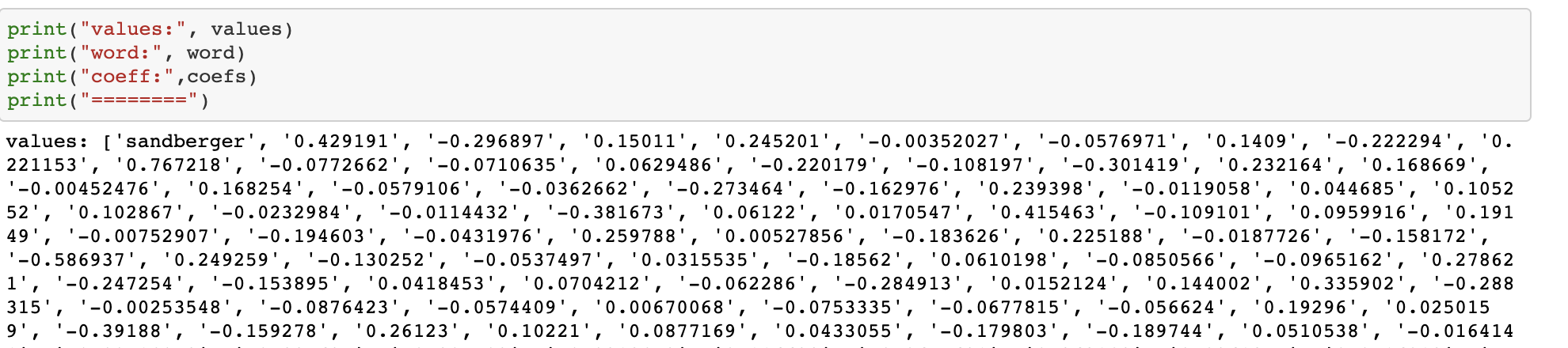
Downloading Glove Model:



Extracting the words and their coefficients in glove model and store it to a variable called embedding\_index:



Printing the values extracted from glove:



Looking at the coefficients of the word “the”:

