We use different packages in nlp

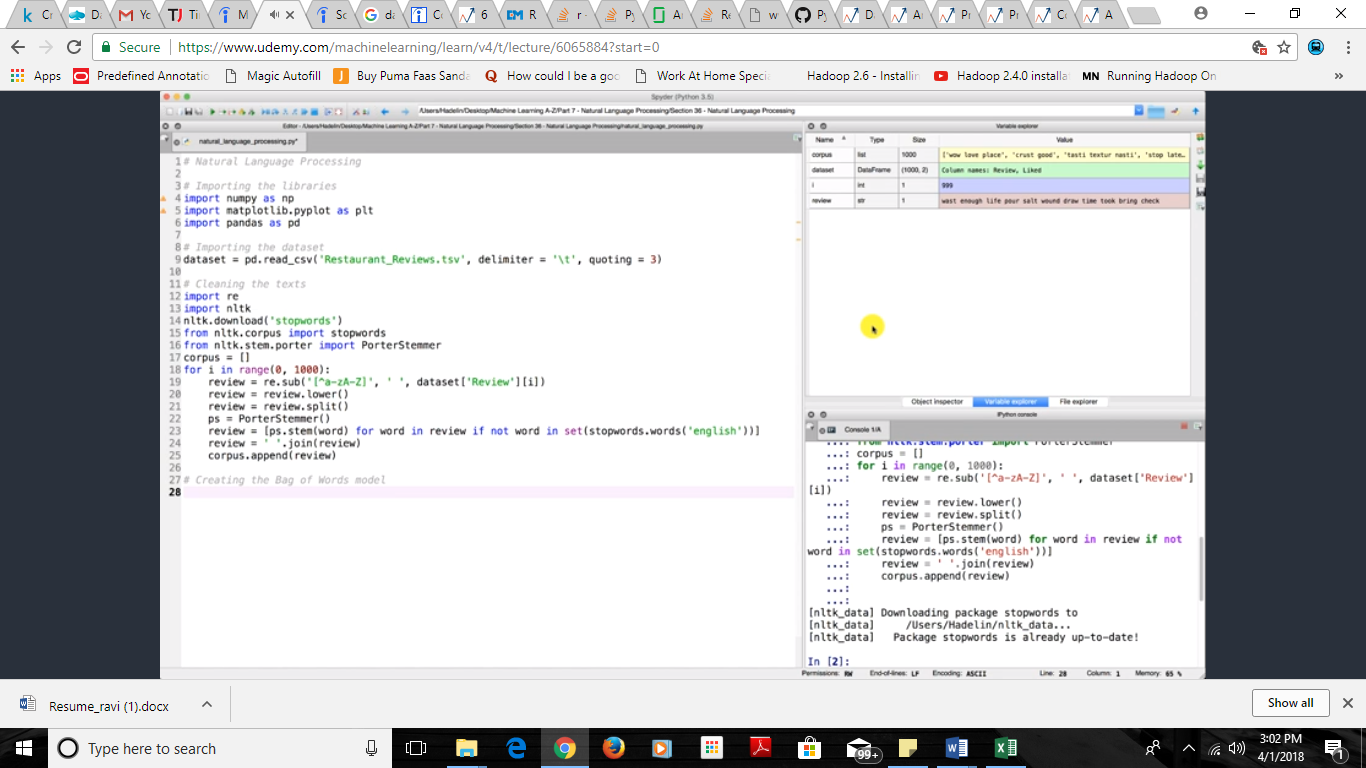
They are

Re

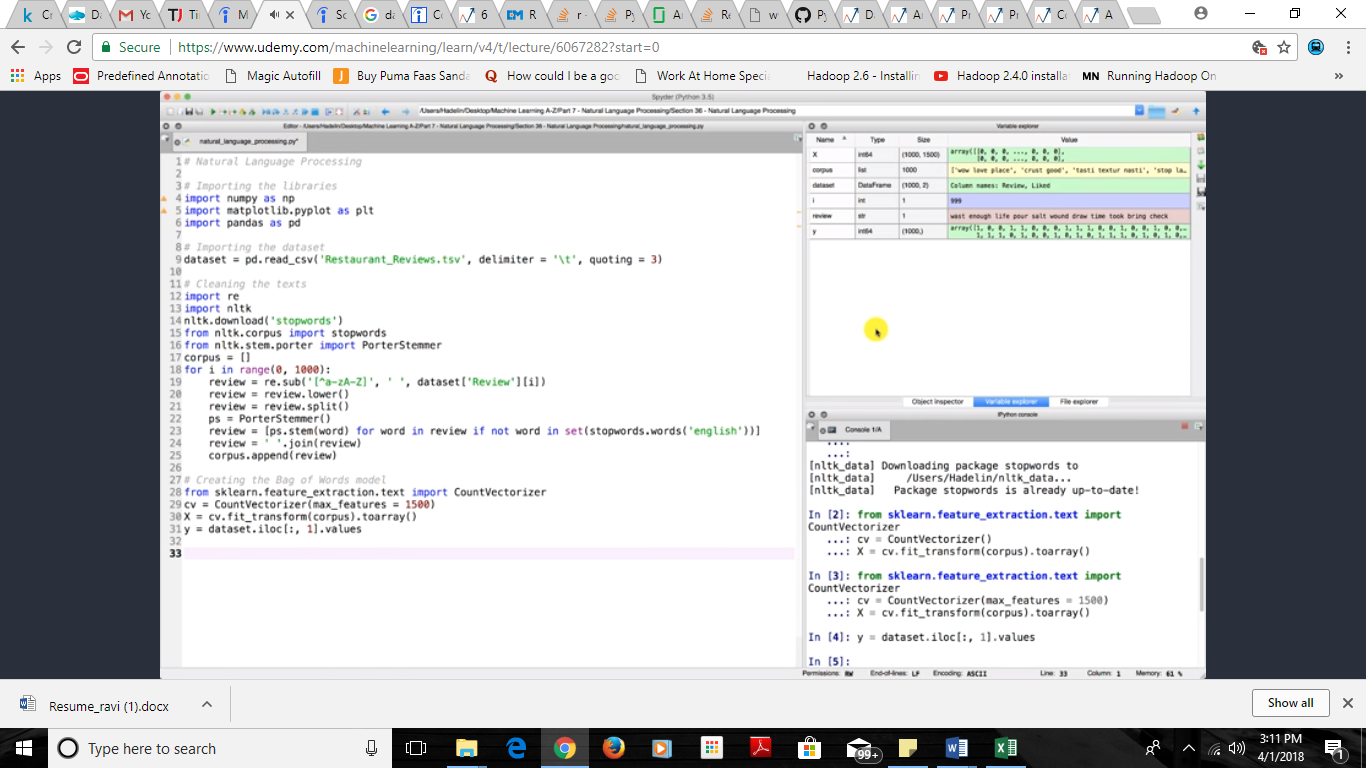
Nltk

Re is regular expression package where we remove digits,split string, lower case conversion,take only alphabets like these can be done using this

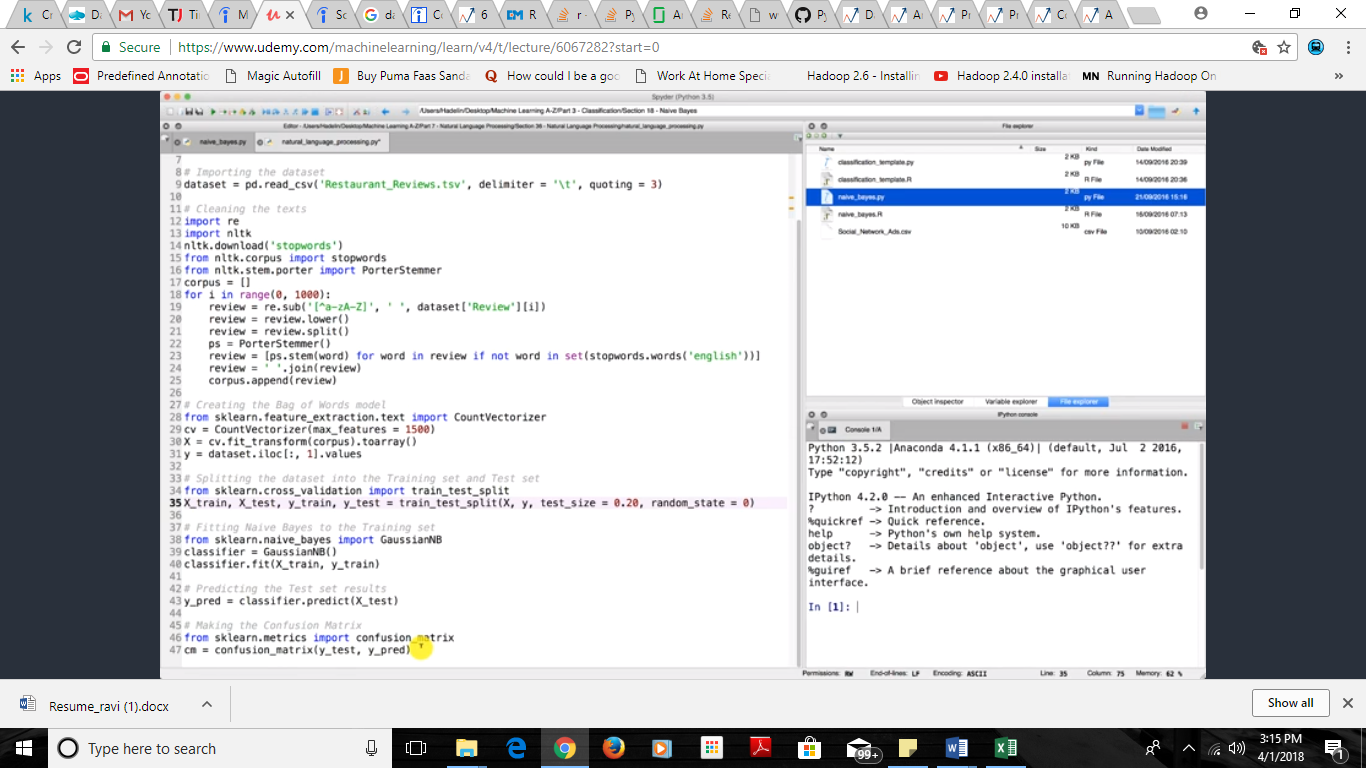
Nltk is for using stopwords so that our text don’t contain stop words and stemming is also done using it(love,loved,loving) all these are convertd to love by Stemming(porterstemming)



Converting data to bag of words (like number of times word appeared in that comment) we can select the number of features by giving number which takes he max number of features by their importance that is the count of the feature appeared in comments



Afteecleaning the total data now apply the ML algo like u do normally



In R:

We use tm package to work on it

For stop words we need snowball package to work

