Lab 4

Develop naïve Bayes and logistic regression classifiers to predict whether a statement is positive or negative using the training data you have been provided. Use an open source and publicly available library to build the classifiers. There should be two versions for each classifier – one that includes text normalization and one that does not.

You are also to write a report about the lab. The report should contain:

1. An explanation for your choice of libraries
2. How you evaluated your classifiers and your results. You should have results for each version of each classifier.
3. A discussion of your results. Why do you think you got the results that you did? Compare the results.

Submission Details

1. You should have a folder called *lab4* in your Github repository. Inside the directory, there should be a program called *lab4.py*. There should also be the Jupyter version of the program inside the directory.
2. Your program should be runnable from the command line using the following command: python lab4.py <classifier-type> <version> <testfile>. classifier-type can either by nb for naïve Bayes or lr for logistic regression. version can either by n for normalized or u for unnormalized. E.g. "python lab4.py nb u test.txt" should run the unnormalized version of your naïve bayes classifier on a file called test.txt.
3. Your program should output a file called results-<classifier-type>-<version>.txt which should print the output for each test sentence on a new line. E.g. results-nb-n should print the results of the normalized version of your naïve Bayes classifier.
4. Your code should be well commented. There should be a comment explaining what each line of code is doing except for lines that are very obvious such as import statements. Essentially, a programmer who knows nothing about machine learning should be able to look at your code and comments and know what each line of code is doing.

Assigned: 29th October 2018

Due: 11th November 2018 before 2:45pm.