TWITTER AIRLINE SENTIMENT ANALYSIS

In class, we studied the naïve Bayes algorithms and its application to text classification. We also had a lab on this topic. In this assignment, you will apply the NB classifier to the Twitter US Airline Sentiment dataset, which is available at:
https://www.kaggle.com/crowdflower/twitter-airline-sentiment/version/2

You will work with the file Tweets.csv and primarily be concerned with three columns:
- airline_sentiment
- airline
- text

You have to write a Python script to perform the following tasks:

1. Your program should have one argument that reads in the location of the Tweets.csv file from your computer.
If you are not familiar with this, you can read more about it here:
https://www.pythonforbeginners.com/system/python-sys-argv

2. Read the above 3 columns into a dataframe

3. Perform the following text pre-processing steps:
- convert text to lowercase

- transform the text data using CountVectorizer and TfidfTransformer, just like we did in the lab.

- convert the airline_sentiment from categorical to numerical values. You can use the LabelEncoder class in sklearn to do this:
http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html

4. Now split the data into two parts: training and testing. 10% of the data should go to the test part. You can use the train_test_split method in scikit learn to accomplish this.

5. Build a Multinomial Naïve Bayes (MNB) model using the training dataset. You have to choose the best set of parameters.

6. Apply your model on the test part and output the accuracy.

7. Repeat this process 5 times with different parameter choices and output the parameters and accuracy in a tabular format.

8. The following is not related to naïve Bayes, but you can use the above data to answer the following question:
Using the numeric value of airline_sentiment, output the average sentiment of each airline and report which airline has the highest positive sentiment.

What to Submit:

- Your code written in Python. Note that your script should read in the location of the input file. Do not hard code the path.
- Summary of your results
- README file indicating how to compile and run your code.