**NLP Sentimental Analysis Assignment Solution**

Reference: Analysing Messy Data Sentiment with Python and nltk:

<https://www.twilio.com/blog/2017/09/sentiment-analysis-python-messy-data-nltk.html>

Install the **nltk** library ( <http://www.nltk.org/)> as shown in the Environment Setup section

Download the tweets data files (neg\_tweets.txt, pos\_tweets.txt)

Use the **Jupyter** notebook if you like ( Twilio-nltk-scikit-learn.ipynb **)**

Follow the examples in sections:

Preparing the Data

Building a Classifier

Classification

Accuracy

**Select 10 random tweets from Twitter:**

**Classify the tweets as 'pos' or 'neg’**

Solution:

ex1 = "Everyday is a new opportunity to make someone smile."

print(classifier.classify(format\_sentence(ex1)))

ex2 = "The future depends on what we do in the present."

print(classifier.classify(format\_sentence(ex2)))

ex3 = "As we look forward, I want our first steps to reflect what matters most to you. Share your thoughts with me at http://Obama.org ."

print(classifier.classify(format\_sentence(ex3)))

ex4 = "I'm still asking you to believe - not in my ability to bring about change, but in yours. I believe in change because I believe in you."

print(classifier.classify(format\_sentence(ex4)))

ex5 = "I won't stop; I'll be right there with you as a citizen, inspired by your voices of truth and justice, good humor, and love."

print(classifier.classify(format\_sentence(ex5)))

ex6 = "Cindy McCain says that neither President Trump nor Melania Trump reached out to her after John McCain's funeral."

print(classifier.classify(format\_sentence(ex6)))

ex7 = "Beautiful, positive, things happen in your life when you distance yourself from the negative things."

print(classifier.classify(format\_sentence(ex7)))

ex8 = "That anger you have. Is it making you and those around you happy?"

print(classifier.classify(format\_sentence(ex8)))

ex9 = " that’s not the argument. Your a hater"

print(classifier.classify(format\_sentence(ex9)))

ex10 = "It’s never too late to go back and make a wrong that you did, right again."

print(classifier.classify(format\_sentence(ex10)))

pos

pos

pos

neg

pos

pos

pos

neg

neg

neg

**Discuss the accuracy of the classification (i.e., is the classification "correct"?)**

Solution:

The model worked pretty well in classifying the tweets from twitter. However, when we mix tweets with different sentiments it sometimes miss-classifies the tweets as positive or negative tweets. One such example which I had is ex4. It wrongly classified the positive motivational tweet to a negative tweet only because of the word ‘not’. Naive Bayes doesn’t see for relationship between words, because of which it was not able to catch the word ‘not’ and, ‘but’ which says that there is something positive after that. It read it as two negative words and classified it as such.

**How can the accuracy be improved?**

Solution:

The current dataset which we have consists of grammatical errors and the data is a messy data with lots of typos, abbreviations etc. More processing needs to be done on the data so as to improve the accuracy before processing the data through the model. This will not only help in having all tweets in a generalized way but will also remove the messy data and will improve the accuracy of the model.