Getting Started with Text Analysis Project

Project Deliverables

You will be required to provide the following deliverables.

- A python notebook with your solution.

Instructions

Background Information

The management of a certain Marketing Firm would like to track the sentiments of their customers. This would help in shortening the amount of time that it takes to act on feedback.

Problem Statement

Your task for this project will be to create a model that can predict whether the sentiment of a tweet is positive or negative. The desired accuracy of your model is 70%.

Below are the text processing steps that you will be required to perform in this project:

- **Text Cleaning/ Text Processing**
  - Removing all URLs/links
  - Replacing @ and # Characters
  - Feature Construction (No. of Punctuation Characters)
  - Removing Punctuation Characters
  - Feature Construction (Lowercase, Uppercase and Proper case words)
  - Conversion to Lowercase
  - Splitting Concatenated words
  - Spelling Correction
  - Feature Construction (Counting the no. of stop words/tweet)
  - Removing Stop words
  - Lemmatization
- **Text Feature Engineering Techniques**
  - Length of text
  - Word Count

- ○ Word density (Average no. of Words / Tweet)
- ○ Noun Count
- ○ Verb Count
- ○ Adjective Count
- ○ Adverb Count
- ○ Pronoun Count
- ○ Polarity
- ○ Subjectivity
- ○ Word Level N-Gram TF-IDF tweet_word_tfidf
- ○ Character Level N-Gram TF-IDF tweet_character_tfidf

You can use the following Guiding Template [Link].

Dataset

- ● Datasets for this project can be found here: [https://bit.ly/31kqByD].
- ● You can load the dataset from the URL.

Dataset Source: Twitter