Sentiment Analysis using Natural Language Processing

Sentiment analysis

Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations. Brands can use this data to measure the success of their products in an objective manner. In this challenge, you are provided with tweet data to predict sentiment on electronic products of netizens.

Problem Statement

Sentiment analysis remains one of the key problems that has seen extensive application of natural language processing. This time around, given the tweets from customers about various tech firms who manufacture and sell mobiles, computers, laptops, etc, the task is to identify if the tweets have a negative sentiment towards such companies or products.

Implementation Approach

We will use the python language for implementing this assignment.

Dataset

- The train set contains 7,920 tweets
- The test set contains 1,953 tweets

Text Cleaning and Preprocessing

You should apply the below text preprocessing items on the training and testing tweets sets:

- URLs removal: We have used Regular Expressions (or RegEx) to remove the URLs.
- Punctuation marks removal: remove any punction marks from the text.
- Numbers removal: replace any digits in the tweets with space.

- Whitespaces removal
- Convert the text to lowercase.

Classification Model Building and Evaluation

You can use any classification model for this problem e.g. neural networks, BERT , ELMO models, …

Submission:

- Each group of three students should register and submit their solution for the below public contest URL, one submission for the group.
    - https://datahack.analyticsvidhya.com/contest/linguipedia-codefest-natural-language-processing-1/
- For the groups who will not submit for the contest will take ZERO .
- Your rank in the contest will be considered in the grading.
- The project will be marked out of 10 grades.
- **Copies** will take -5
- The highest 5 groups with highest ranks will take 5 grades as BOUNS.
- **Deadline is 10 MAY 2023**

References that may help:

- https://www.analyticsvidhya.com/blog/2021/06/nlp-sentiment-analysis/
- https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+NLP101+2018_T1/about?utm_source=practice_problem_Identify_The_Sentiments&utm_medium=Datahack
- https://www.analyticsvidhya.com/blog/2018/11/tutorial-text-classification-ulmfit-fastai-library/
- https://www.analyticsvidhya.com/blog/2018/07/hands-on-sentiment-analysis-dataset-python/