Group name: VioletGo
Name: Xinyu Wang
Email: xw3080@nyu.edu
Country: United States
College: New York University
Specialization: NLP

Problem description: To construct a classifier utilizing Natural Language Processing (NLP) techniques. Its primary function is to scrutinize and identify offensive or provocative comments on Twitter, specifically those that contain hate speech or abusive language, with the ultimate aim of fostering healthier online interactions.

Repo link: https://github.com/VioletGo319/sentimental_analysis/tree/Week-10-Deliverables

Final Recommendation:

1. The text analysis revealed a high frequency of terms such as 'user' and 'amp', which appear to be unrelated to the categorization of normal and hate speech. We recommend that these terms be processed further and included in the list of stopwords to improve the efficiency of the classification.

2. An analysis of the dataset reveals that approximately 93% of tweets are categorized as normal speech. The word cloud generated from all tweets suggests a predominance of positive sentiments, with a profusion of terms related to love, smile, and other positive emotions. However, a small but significant 7% of tweets are labeled as hate speech, with a noticeable emphasis on terms linked to racial and political issues, such as 'white', 'black', and 'Trump'. It seems to indicate a politically charged environment, possibly reflecting certain issues of political correctness.

3. In terms of the length of tweets, the majority of them range from 2 to 20 words. This insight could inform future text processing and modeling strategies to optimize the processing speed and efficiency.