Week 11 Deliverables

Group Details

- Name: William Harvey
- Email: will.harvey@att.net
- Country: United States of America
- College: University of California, Berkeley
- Specialization: NLP Analyst

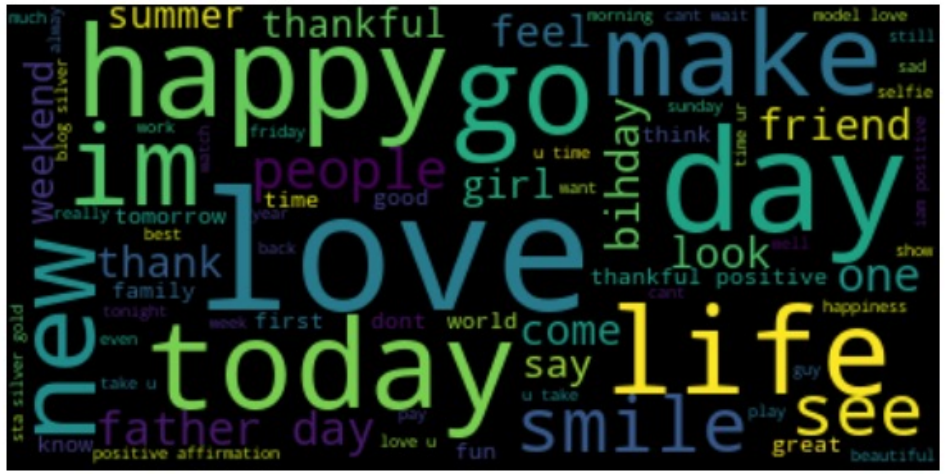
Problem Description

- Classifying hate speech in tweets on Twitter

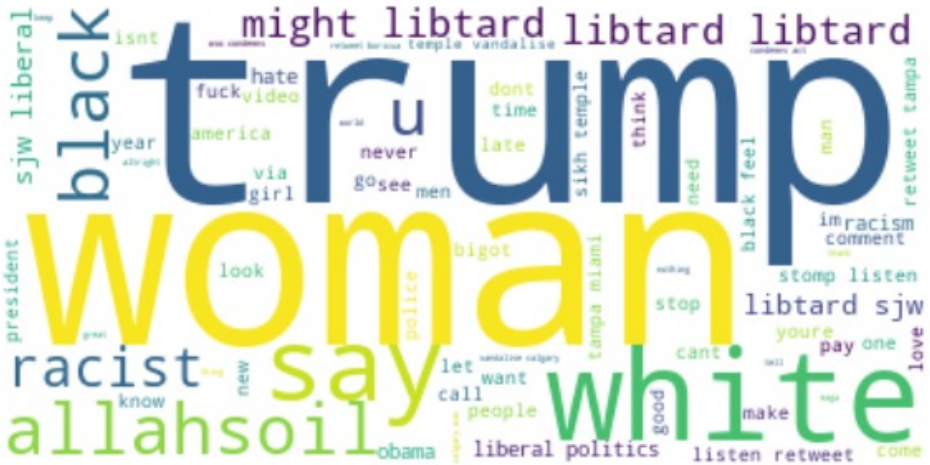
EDA Presentation for Twitter

Word Cloud Comparisons

Most Popular Words in Non-Hate Comments



Most Popular Words in Hate Comments



Impact on Modeling

- As we could see, the words “trump” and “woman” appeared very heavily in hate speech tweets, while words like “love” or ”happy” appeared the most in non-hate speech tweets.
- Our model will look to those words, among the other bigger and bolder words in order to classify our test data as either hate or non-hate

Technical Model Recommendation

- A classic Logistical Regression model will be explored, as the only input features will be the text, so overfitting isn’t an issue
- An XGB Classifier is also being discussed, as this type of classifier is a very powerful way to include gradient descent in our results

Repo Link

https://github.com/wsharvey/DG-Harvey/blob/main/DG_Proj%20(1).ipynb