Final Project Report

Group Details

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

Problem Details

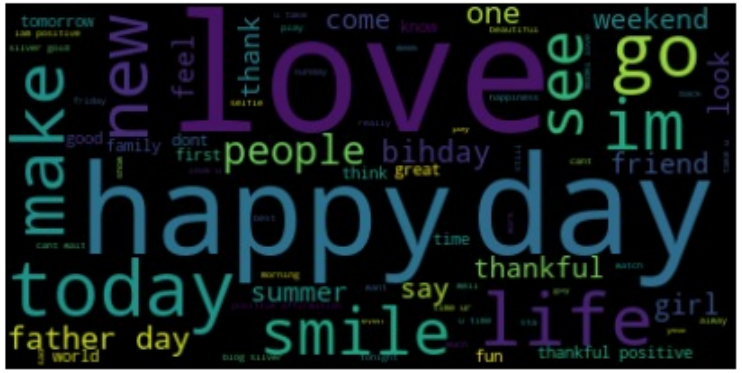
- I am doing the advance NLP project on detecting hate speech on Twitter. Hate speech is a very important problem, as it attacks a person for their identity by using defamatory language. It is commonly used in today’s world because Twitter makes it so you can connect with millions of people from across the globe, however hate speech should be monitored and shut down. I will be using sentiment Twitter data to train a classifier to identify hate speech on the platform.

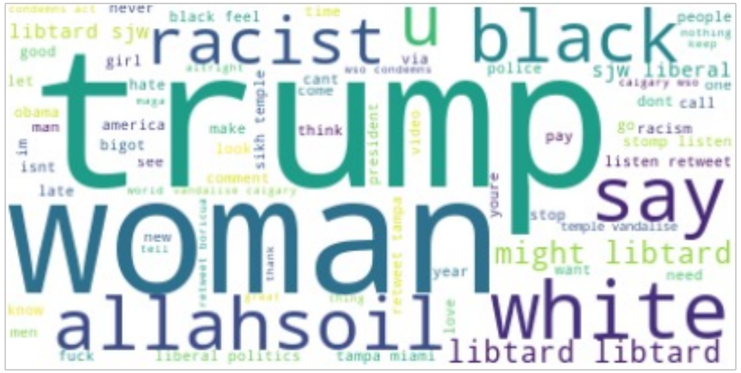
Raw Data Understanding

- No NaN values existed, so there was no need to clean the data based on this
- However, data imbalance existed, as there were many more tweets that were NOT labeled as hate speech, so we used a resample to make the data balance even

EDA

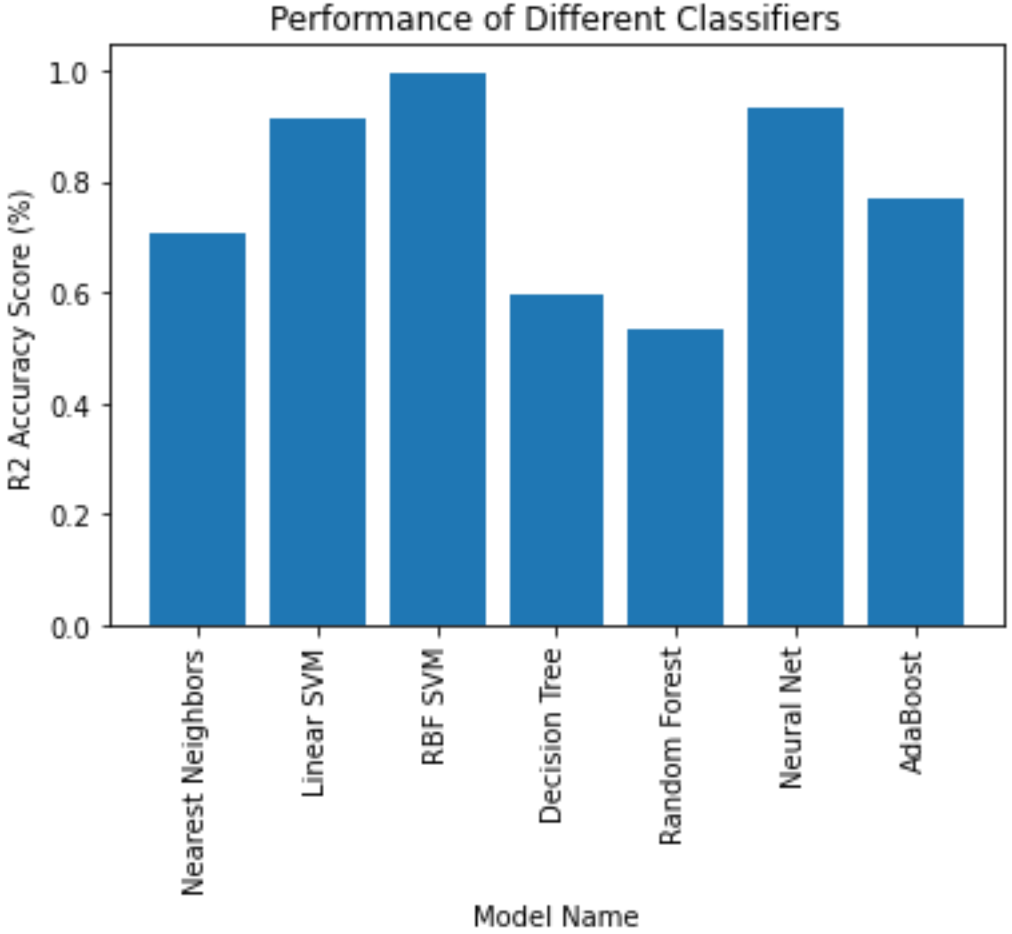
- I generated the below WordClouds to see which words were most common in hate speech versus no hate speech


Most Popular Words in Non-Hate Comments


Most Popular Words in Hate Comments

Model Selection

- I tested 7 different types of classifiers, measuring their F1 and accuracy scores. The result is visualized below

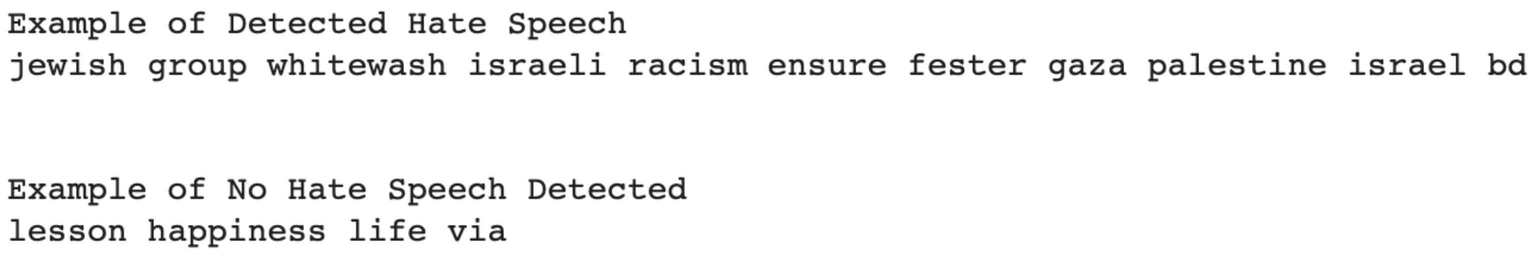


Model Performance

- The RBF SVM model performed at a 99% accuracy score, but now we looked at precision and recall to make sure that this model was the best to use
- Both precision and recall scores were above 99%, leading to the RBF SVM model to be the recommended model

Model Performance (part 2)

- We passed in the second dataframe to our created model, and the results were consistent with our earlier predictions. Below is an example of detected hate speech versus no hate speech detected



```
Example of Detected Hate Speech
jewish group whitewash israeli racism ensure fester gaza palestine israel bd


Example of No Hate Speech Detected
lesson happiness life via
```

Repo URL

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