NLP Hate Speech Detection

Ankita Manna
Week 11
LISUM 38
Data Glacier Internship

Table of contents

Overview

The goal is to accurately detect tweets containing hateful content while minimizing incorrect predictions. Hate speech refers to any form of verbal, written, or behavioral communication that attacks or uses derogatory or discriminatory language targeting individuals or groups based on characteristics such as race, religion, gender, or ethnicity. This project focuses on developing a reliable detection system to identify and mitigate such harmful content.

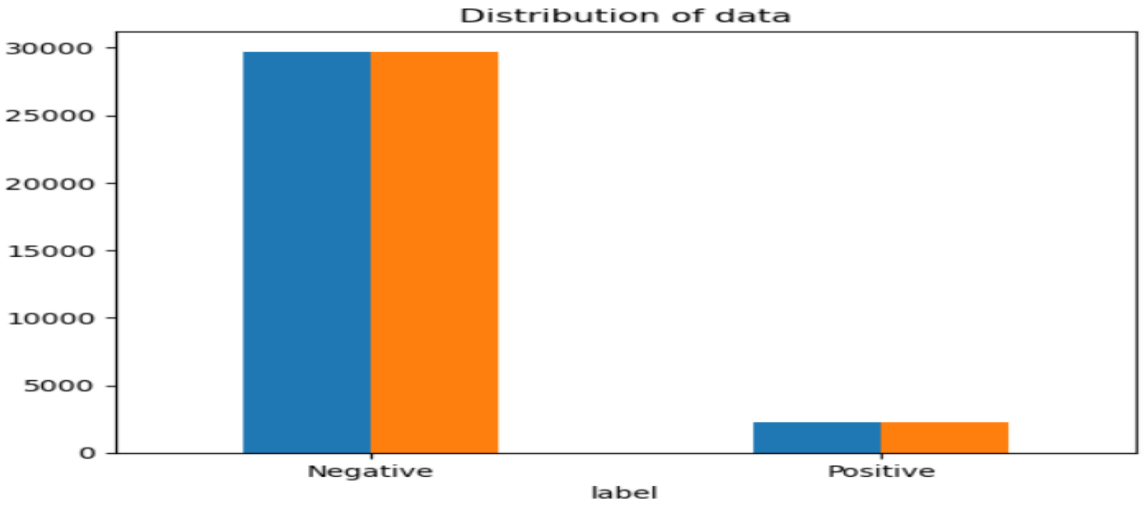
Steps

The general approach of the project is divided into important major section . They are outlined as follows:

1.Downloading the data
2.Preparing and Preprocessing the data
3.Data Visualization
4.Transform the data
5.Model development
6.Training and Validation
7.Model Inference
8.Deployment

EDA(1 of 1)

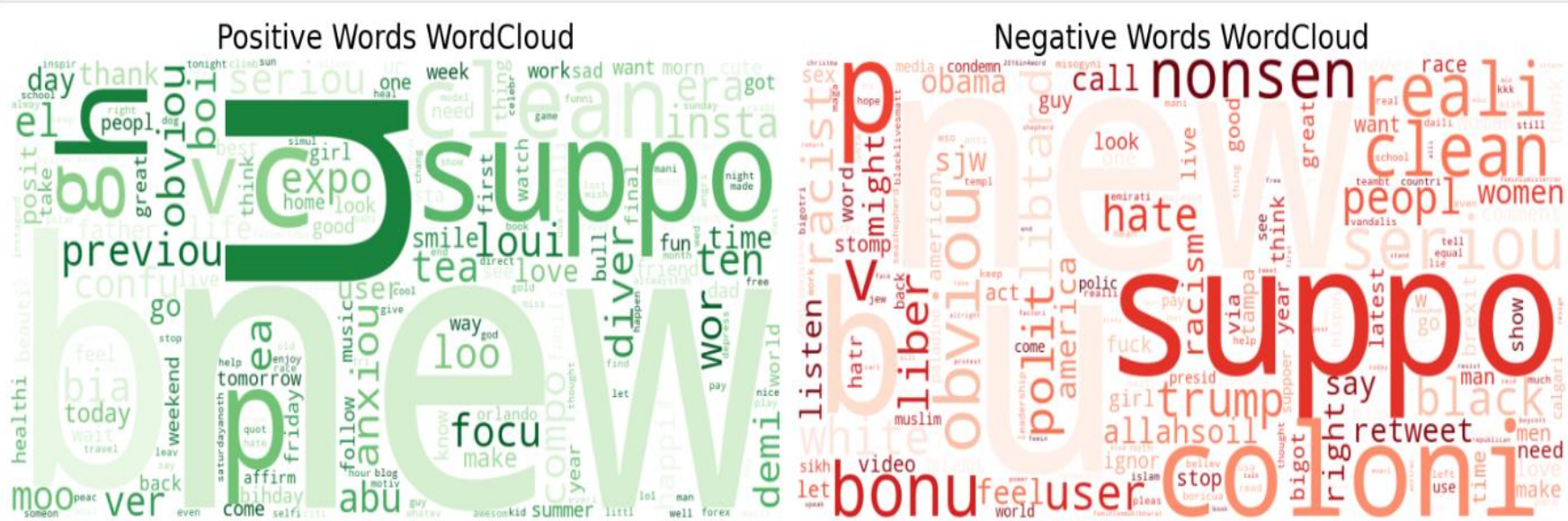
Count of positive and Negative Words:



INTERPRETATION: Number of negative tweet is much higher than the number of positive tweets.

EDA(2 OF 2)

Positive and Negative word cloud:



Summary & Recommendation:

• The EDA has been done on the textual data to see which words used the most on each class and found that the word ‘User’ has been used much on both classes. This word is nothing, but the common word used by Twitter. For positive words, ‘love’ comes second, whereas for negative words, ‘trump’ comes next.’

• Also, we found many local slang words in the corpus that are not correctly spelled in English. Attempts will be made to replace those words by building a dictionary with key as the slang words and their correct word as value and replacing them across the corpus. Building such a dictionary in would be useful in the long term in developing the model.

• Class Imbalance has been found on the data, so for designing the model, the under-sampling technique adopted as oversampling may overfit the model. There would be some loss in the model, but it can be compensated in a long run by acquiring more data with the other class in future.

Thanks!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**