

# Week 10 Deliverables

Orkun

9 July 2024

## Team Member's Details

- **Group Name:** Orkun
- **Name:** Orkun Kınay
- **Email:** orkunkinay@sabanciuniv.edu
- **Country:** Turkey
- **College/Company:** Sabancı University
- **Specialization:** NLP

## Problem Description

The goal of this project is to develop a hate speech detection model using Twitter data. The dataset consists of tweets labeled as hate speech or non-hate speech, which will be used to train and evaluate machine learning models. The primary objective is to accurately classify tweets and mitigate the spread of hate speech on social media platforms.

## GitHub Repo Link

<https://github.com/orkunkinay/Hate-Speech-Detection/tree/main/EDA>

## Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was performed to understand the distribution, patterns, and relationships within the data. The following steps were taken during the EDA process:

### Data Visualization

Various plots and visualizations were created to analyze the data, including:

- **Distribution of Tweet Lengths:** The distribution of tweet lengths was plotted to understand the range and frequency of tweet lengths. Most tweets are between 60 and 110 characters long.
- **Class Distribution:** The class distribution of hate speech and non-hate speech tweets was plotted. The dataset is imbalanced with a higher number of non-hate speech tweets.
- **Word Clouds:** Word clouds were generated for both hate speech and non-hate speech tweets to visualize the most common words.

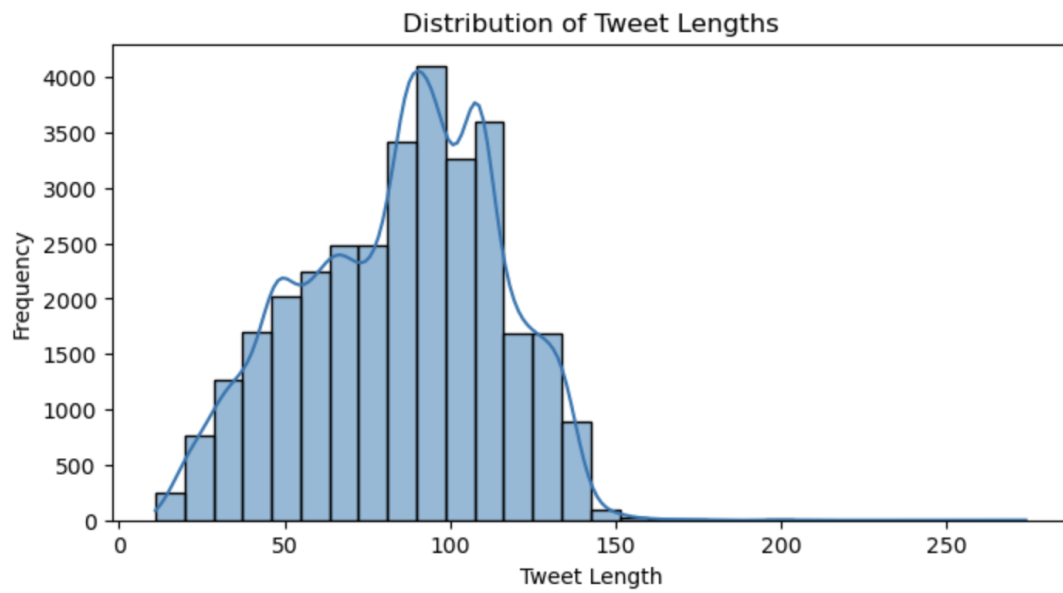


Figure 1: Distribution of Tweet Lengths

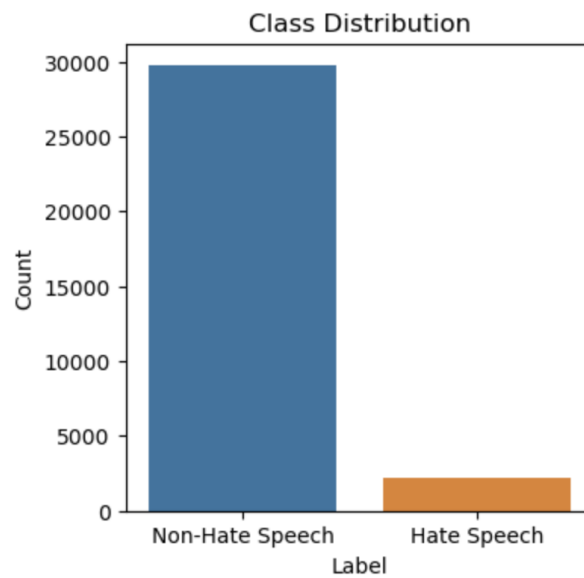


Figure 2: Class Distribution

## Statistical Analysis

Statistical analysis was conducted to identify significant differences between hate speech and non-hate speech tweets. This included:

- **Tweet Length Statistics:** Statistical analysis of tweet lengths showed the following results:

count	31962.000000
mean	84.739628
std	29.455749
min	11.000000
25%	63.000000
50%	88.000000
75%	108.000000
max	274.000000

[illegible]

- **Common Words:** Frequency analysis of common words in non-hate speech and hate speech tweets:

- \* day (2797)
- \* love (2745)
- \* happy (1679)
- \* u (1578)
- \* amp (1325)
- \* life (1221)
- \* time (1205)
- \* im (1112)
- \* today (1069)
- \* get (949)
- \* like (948)
- \* positive (932)
- \* thankful (925)

- \* father (919)
- \* new (917)
- \* bihday (856)
- \* good (820)
- \* smile (812)
- \* make (804)
- \* people (790)
- **Hate Speech Tweets:** Common words include:
  - \* amp (283)
  - \* trump (216)
  - \* white (153)
  - \* libtard (149)
  - \* black (146)
  - \* like (140)
  - \* woman (120)
  - \* racist (109)
  - \* politics (97)
  - \* people (95)
  - \* liberal (92)
  - \* allahsoil (92)
  - \* u (89)
  - \* might (77)
  - \* sjw (74)
  - \* new (71)
  - \* hate (69)
  - \* obama (68)
  - \* retweet (67)
  - \* dont (67)

## Final Recommendation

Based on the EDA, the following recommendations were made:

- Focus on tweets with higher lengths, as they are more likely to contain hate speech.
- Pay attention to frequently occurring words and phrases in hate speech tweets for better feature extraction.
- Implement advanced text preprocessing techniques to enhance model performance.