Project Report

Team Member's Details
Group Name: Scalable Minds
Members: Name: Igwebuike Eze and Lucy Nowascki
Email: samson12193@gmail.com and quantlucy@gmail.com
Country: UK
College/Company: Queen Mary
Specialization: NLP

Problem Description

The term hate speech is understood as any type of verbal, written or behavioural communication that attacks or uses derogatory or discriminatory language against a person or group based on what they are, in other words, based on their religion, ethnicity, nationality, race, colour, ancestry, sex or another identity factor. In this problem, We will take you through a hate speech detection model with Machine Learning and Python.

Hate Speech Detection is generally a task of sentiment classification. So for training, a model that can classify hate speech from a certain piece of tweet can be achieved by training it on a data that is generally used to classify sentiments. So for the task of hate speech detection model, We will use the Twitter tweets to identify tweets containing Hate speech

Data Understanding

We were presented with two datasets namely:

a) Train - train_E6oV3lV.csv

b) Test - test_tweets_anuFYb8

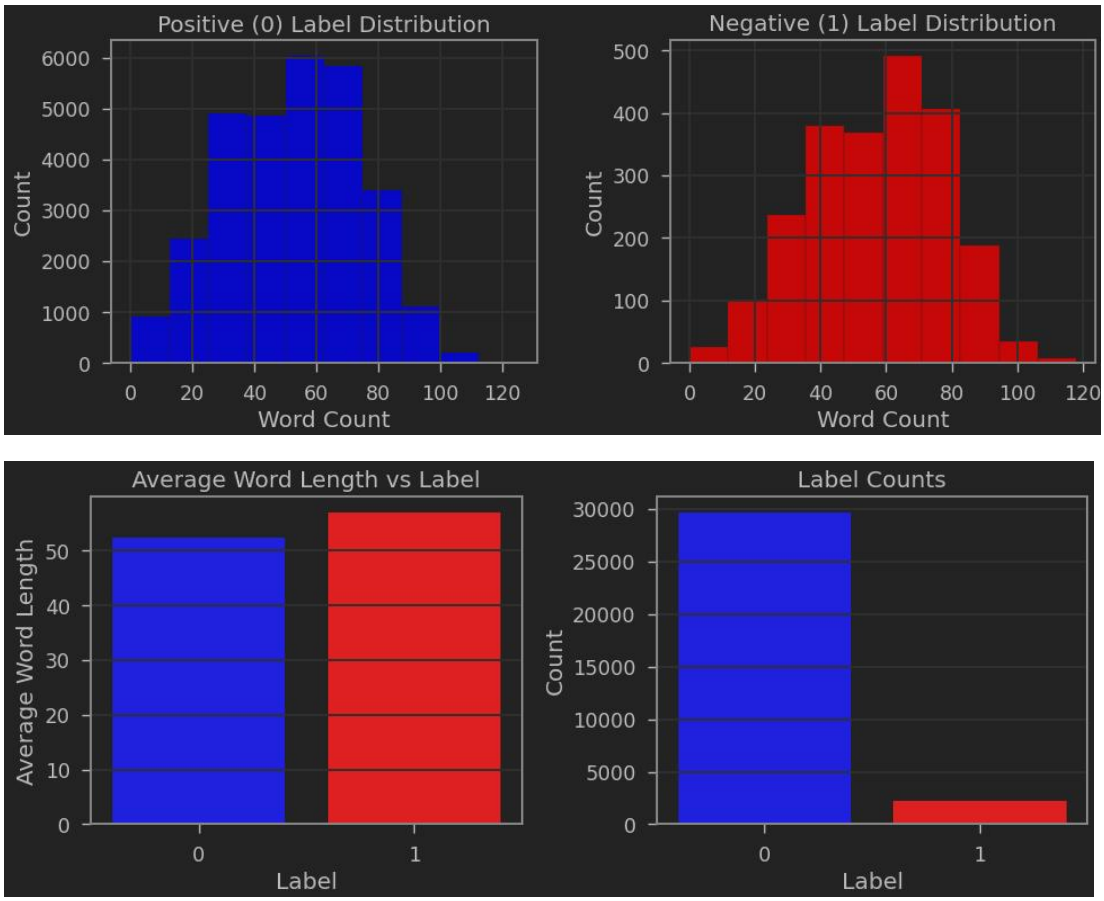
A summary of the two data sets show the following:

	Train	**Test**
RangeIndex	0 - 31962	0 - 17197
Nos columns	3 (id, label and tweet)	2 (id and tweet)
Data Type	id - int64 label - int64 tweet - object	id - int64 tweet - object
Nos Unique Values	0	0
Label Value Count	0 - 29720 1 - 2242	No Label
Nos Missing Values	0	0
Memory Usage	749.2 KB	268.8 KB

A further in depth review of the train dataset showed the following:

Statistic	Positive (0)	(0)Negative
Mean	70.665007	77.4005351
Median	73.000000	81.000000
Standard Deviation	28.269149	25.8280043
Variance	799.144778	667.0857804

Skewness	-0.039524	-0.3854375
Kurtosis	-0.691766	-0.561919



We categorized the tweets into positive (label 0) and negative (label 1). The results are visualized in a 2x2 grid of subplots, providing insights into the distribution of word counts and the characteristics of the tweets based on their labels. The statistical parameters provide further depth to this analysis.

1) Positive Tweets Word Count Distribution The histogram in the top-left subplot illustrates the distribution of word counts in positive tweets (label 0). The distribution shows that most positive tweets contain between 20 to 60 words, with noticeable peaks around 30 and 50 words. This suggests that positive tweets tend to be concise but informative, falling within a moderate range of word lengths.
2) Mean: 71.61 Median: 73.00 Standard Deviation: 28.39 Skewness: -0.06 Kurtosis: -0.70 The average word count is approximately 71.61 words, and the median is 73 words. The standard deviation indicates a moderate spread around the mean. The negative skewness (-0.06) suggests a slight left-skew, meaning there are slightly more tweets shorter than the mean. The negative kurtosis (-0.70) indicates a flatter distribution with fewer outliers than a normal distribution.
3) Negative Tweets Word Count Distribution The histogram in the top-right subplot depicts the word count distribution for negative tweets (label 1). The distribution

is more varied compared to positive tweets, with a significant number of tweets around the 10-20 word range and several spikes indicating higher word counts up to around 60-70 words.

4) Mean: 78.28 Median: 82.00 Standard Deviation: 26.03 Skewness: -0.39 Kurtosis: -0.57 The average word count for negative tweets is higher at around 78.28 words, and the median is 82 words. The standard deviation is slightly lower than that of positive tweets, indicating slightly less variability. The more pronounced negative skewness (-0.39) suggests a greater number of shorter tweets in this category. The kurtosis (-0.57) also indicates a flatter distribution, similar to positive tweets but with slightly more outliers.

5) Average Word Length vs Label The bar plot in the bottom-left subplot compares the average word length for positive and negative tweets. The results show a distinct difference:

6) Positive Tweets (label 0): Average word length is approximately 71.61 words. Negative Tweets (label 1): Average word length is around 78.28 words. This indicates that negative tweets are generally longer than positive tweets. Longer tweets may contain more detailed expressions of negative sentiment or elaborate on specific issues.

7) Label Counts The bar plot in the bottom-right subplot displays the count of positive and negative tweets in the dataset. There is a significant imbalance, with positive tweets (label 0) being much more prevalent than negative tweets (label 1). The count of positive tweets is around 30,000, while the count of negative tweets is significantly lower. This imbalance is crucial for understanding the dataset and indicates the need for careful consideration during model training to avoid bias and ensure balanced representation.

In conclusion, positive tweets are generally moderate in length and more prevalent. Negative tweets exhibit greater variability in word count and slightly longer average word lengths. Both distributions exhibit slight left skewness and a flatter-than-normal distribution, with negative tweets having slightly fewer outliers. These insights are essential for developing effective models to detect hate speech and understand the sentiment distribution in social media content. Understanding the distribution and characteristics of the tweets helps in better handling of the data for tasks such as sentiment analysis, hate speech detection, and overall text analytics.