

# FLIP 01 PROJECT REPORT

Jing Miao  
Qingdao University of Technology, China

## Introduction

With all of the tweets circulating every second it is hard to tell whether the sentiment behind a specific tweet will impact a company, or a person's, brand for being viral (positive), or devastate profit because it strikes a negative tone. Capturing sentiment in language is important in these times where decisions and reactions are created and updated in seconds. The purpose of this task is to detect hate speech in tweets. For the sake of simplicity, if there is racist or sexist sentiment on Twitter, we will say that it contains hate speech. Therefore, the task is to classify racist or sexist tweets from other tweets. If there is a system that can detect this type of text, it will definitely make the Internet and social media a better, non-malicious communication space.

## Data Analysis

- View the data of the training data set and the test data set, in which the training data set contains text ID, text, extracted text (the original text reflects the emotion part), emotion classification; The test set contains text ID, text, sentiment classification.
- There are 27481 pieces of data in the training set, among which one text and extracted text data are empty. The emotion classification is divided into three parts: positive, negative and neutral. By drawing Funnel-Chart to count the proportion of texts under different sentiment classification, it can be concluded that, the neutral sentiment accounted for the majority, followed by the positive and the least negative.
- So further analysis is needed to check the difference in Number of Words and the Jaccard Scores similarity across Different Sentiments between Text and selected text. We can do that by visualizing the data, It can be seen that the number of tweets with a Jaccard similarity of 1 between text extraction and text is more neutral sentiment. In conclusion maybe we can use neutral text as it is for selected text in test data submission. We can see from the Jaccard Score Plot that there is peak for negative and positive plot around score of 1. That means there is a cluster of tweets where there is a high similarity between text and selected texts, if we can find those clusters then we can predict text for selected texts for those tweets irrespective of segment.

## Data Cleaning

First, make text lowercase, remove text in square brackets, remove links, remove punctuation and remove words containing numbers. Then to remove the stopwords. We can see that the top 20 most common words and the text in the selected text are almost the same.

## Data Feature Analysis

- Find unique word (a word to represent some kind of emotional), according to the classification of emotional all words in the text can be divided into three collections, there must be some words also belong to the positive, negative and neutral or both of the three, there must be some words also belongs to a collection of emotion, then these terms can be used as the only word on behalf of a certain emotion.
- By Looking at the Unique Words of each sentiment, we now have much more clarity about the data, these unique words are very strong determiners of Sentiment of tweets.

## WordClouds

- By drawing the word cloud, the words in the original text can be viewed more clearly, and the words of different emotional categories can be viewed more clearly and intuitively.

## Model

- Two training models were used for testing, namely NER and BERT.
- Through the training of NER model, the test text obtained contains some non-text data such as web addresses, which leads to unsatisfactory test results. Compared with the NER model, the BERT model has a higher testing accuracy, and the extracted text can reflect the emotion category of the original text.

## Conclusion

Better use of real data and analysis, can find the hidden information in life, natural language processing, information analysis can grasp the future trend and create beneficial value.