

## Airline Tweets

SENTIMENT ANALYSIS TO DISTINGUISH BETWEEN POSITIVE AND NEGATIVE TWEETS

Khanyile Dlamini | Social Media Analytics | Dec. 2019

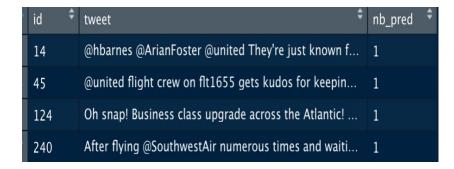
## Packages used: Snowball, tm, twitter, syuzhet, e1071

- In order to understand the non-negative tweets from the given corpus I employed multiple methods to ensure the accuracy of my data, considering context.
- With that said, I created an evaluation function which I would later use to make a comparison between my true positive tweets, and false positive tweets.
  - a. In addition to this, I used the Naïve Bayes classifier to train my data set and remove sparse terms, only wanting 0.5% of important words.

 Loaded actual airline tweet dataset to be tested, removing 99.5% of the sparse words, and finally converting it to a matrix in order to predict non-negative tweets.

```
#predict with Naive bayes
nb_pred <- predict(nb.model, X_test)
airline_tweets$nb_pred=nb_pred</pre>
```

Sorting the data based on non-negatives, extrapolating 1975 tweets, from the initial
 4555.



- Then used sentiment analysis to distinguish between positive and non-positive responses and getting sentiment scores i.e. emotion.
- Set sentiment valuation using sent.value <- get\_sentiment(word.df)</li>
- Merged non-negative tweets and airline tweets by tweets, giving me 768 tweets combined
- Built final evaluation, indicating true positives setting them equal to '2' and adding the variable to the data frame
- Found 607 true positive tweets
- Built evaluation, indicating false positives setting them equal to 'o'
- Found 101 false positives
- Precision = (607/768)

PRECISION = 79%