

# Sentiment Analysis on the Covid-19 Pandemic from Twitter API

Kenneth I. Dagogo

4/7/2020 (7th of April, 2020)

The Covid-19 pandemic which broke out from the Wuhan Province of South China early this year in January is seemingly giving the rest of the world a reason to be worried.

This analysis aims to see what people's reaction globally to the virus is like from the nature of the tweets they have made with the #Covid19 on Twitter. This is to see if there is an overall positive reaction to the pandemic or an overall negative reaction as a result of fear or pessimism.

Importing some initial libraries, we will need the following:

```
library(rtweet)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(graphics)
```

Downloading about eighteen thousand (18,000) tweets from Twitter (this is the maximum number of tweets allowed by Twitter at every fifteen minutes interval). A Twitter API is needed. This will be requested for by a tab in your default browser. I have initialised my API details, so I will not be requested to do so again.

*#### Downloading the tweets and also giving room for retweets to be added to the data set ####*

```
tweets <- search_tweets("#Covid19", n=18000, include_rts = TRUE)
covid19_tweets = tweets
```

It will be observed that some of these tweets are not in the English Language. We need our tweets to be in the English Language because our codes and analysis are in the English Language. Hence, what we need to do is to extract the tweets that are in English Language.

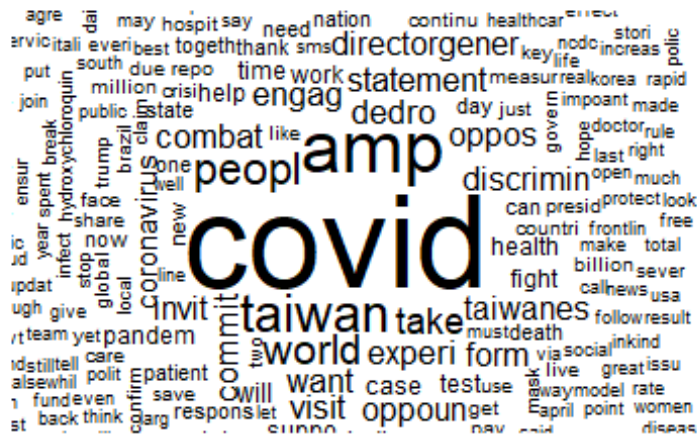
To do this, we query our data and search for the tweets that have their text language in the English Language labelled as “en” in our data. We select our language and also collect all the columns we want to see in a new dataframe.

```
covid19_Eng <- covid19_tweets[ which(covid19_tweets$lang== 'en'),]  
  
### Selecting the columns we want to see in our new data frame ###  
  
covid19_Eng <- covid19_Eng %>%  
  select(user_id, status_id, created_at, screen_name, text, favorite_count, s  
source, is_quote, is_retweet, retweet_text, retweet_created_at)
```

Seeing that the tweets has plenty of contents that will not be useulf for our analysis, we need to cleaning the data. We need to get rid of all the symbols, urls, weblinks, alphanumeric characters etc., from the data. We rely on the “purrr” and “tm” packages to help us get through this. We will also be removing RedHat words like COVID which I am expecting to appear the most as it is the basis for the search hashtag

```
library(purrr)  
library(corpus)  
library(tm)  
  
## Loading required package: NLP  
  
mycorpus <- Corpus(VectorSource(covid19_Eng$text))  
mycorpus <- tm_map(mycorpus, removeWords, stopwords())  
  
  **** Specifying stopwords ****  
myStopWords <- "covid"  
mycorpus <- tm_map(mycorpus, removeWords, myStopWords)  
  
  # Removing urls  
remove_urls <- function(x) gsub("http[^[:space:]]*", "", x)  
mycorpus <- tm_map(mycorpus, content_transformer(remove_urls))  
  
  # Remove anything that is not english letters, words and space  
rnp <- function(x) gsub("[^[:alpha:][:space:]]*", "", x)  
mycorpus <- tm_map(mycorpus, content_transformer(rnp))  
  
mycorpus <- tm_map(mycorpus, removePunctuation)  
  
mycorpus <- tm_map(mycorpus, content_transformer(tolower))  
  
mycorpus <- tm_map(mycorpus, stripWhitespace)  
  
mycorpus <- tm_map(mycorpus, stemDocument)  
  
  # Removing alphanumeric characters  
covid19_Eng$text <- gsub("[^0-9A-Za-z//' ]", "", covid19_Eng$text)
```





Now, The Sentiment Analysis, I chose to do this here with the “sentimentr” package because it is way easier to do, it only takes a single line of code to get the sentiment based on the content of the texts in the tweets. I will run the line of code to get the sentiment and then create a new column for it in our covid19\_Eng data set.

```
library(sentimentr)
SA <- sentiment(covid19_Eng$text)
covid19_Eng$sentiment <- SA$sentiment
```

After this, I will create a new dataframe to carry some details from our covid19\_Eng dataset together with the result of our sentiment analysis.

```
covid19Sentiment <- subset(covid19_Eng, select=c(screen_name, text, source, sentiment))
covid19Sentiment <- as.data.frame(covid19Sentiment)
```

## RESULTS

*The sum*

```
SAsum <- sum(covid19Sentiment$sentiment)
paste("The total sum of all the sentiment scores is", SAsum)

[1] "The total sum of all the sentiment scores is 696.206839593178"
```

This means that there is a tendency that the world is reacting positively to the effect of the pandemic and could be optimistic about beating it.

### *The Mean*

```
SAmear <- mean(covid19Sentiment$sentiment)
paste("The total mean of all the sentiment scores is", SAmear)

[1] "The total mean of all the sentiment scores is 0.0676322945009888"
```

A positive value in the value of the mean here also suggests that the world is mostly reacting positively to the effects of the pandemic.

*The Average Downweighted Zero* Downweight the zeros in a vector for averaging. This is useful in the context of language where we don't want the neutral sentences to have such a strong influence on the general sentiment of the discourse with multiple sentences. Essentially, this means neutral sentences are seen as having less emotional impact than a polarized sentence. (R Documentation)

```
SAadz <- average_downweighted_zero(covid19Sentiment$sentiment)
paste("The total average downweighted zero score of all the sentiment scores is", SAadz)

[1] "The total average downweighted zero score of all the sentiment scores is 0.072396705767282"
```

A positive value here is in support of the sum and the mean values.

Writing the files to CSV format.

```
write.csv(covid19_Eng, file = "covid19_Eng.csv")
write.csv(covid19Sentiment, file = "covid19Sentiments.csv")
```

*CONCLUSION* A positive reaction to the effect of the pandemic could be as a result of various factors. Some of which could be that:

1. People are well aware of the effect of the pandemic.
2. People are well informed on the precautions they need to take to have a chance of beating the virus.
3. Most people are doing what they need to do to stop the virus from spreading further.
4. People are quite optimistic of beating the virus.