

# Sentiment analysis of tweets on #COVID-19

## Data collection and preprocessing

In order to get tweets with certain hashtags the first thing we had to do was to set up a Twitter account. Then we had to apply for developer access. To get it there is a need to answer a few questions concerning things like: why we want it, who we are, what kind of data we will get and what we will do with this data. After successful authorization we had to create a special Twitter application that generates API credentials which can be used to access Twitter from Python.

In Python we needed to provide 4 special keys from our Twitter app (consumer key, consumer secret key, access token key and access token secret key). After that we could use a special *.Cursor()* method from *tweepy* library to search Twitter for tweets containing some terms like for example *#COVID-19*. The *.Cursor()* method returns an object that we can iterate or loop over to access the data. We can also add a special command like *-filter: retweets* and in consequence remove all retweets from our data. There are many other special commands like this. We can for example search for tweets with a given attitude. We can also customize our search query in many ways by for example changing date (from which we want to look for tweets), changing number of tweets we get or searching for some special combination of words and not just one word.

Since 'raw' tweets usually contain some URLs, during preprocessing we removed those URLs in order to get 'clear' text.

## Data analysis

Our goal was to check sentiment of tweets containing *#COVID-19* hashtag and in order to do this we decided to use the *flair* library. We collected 1000 tweets and obtained the sentiments of those.

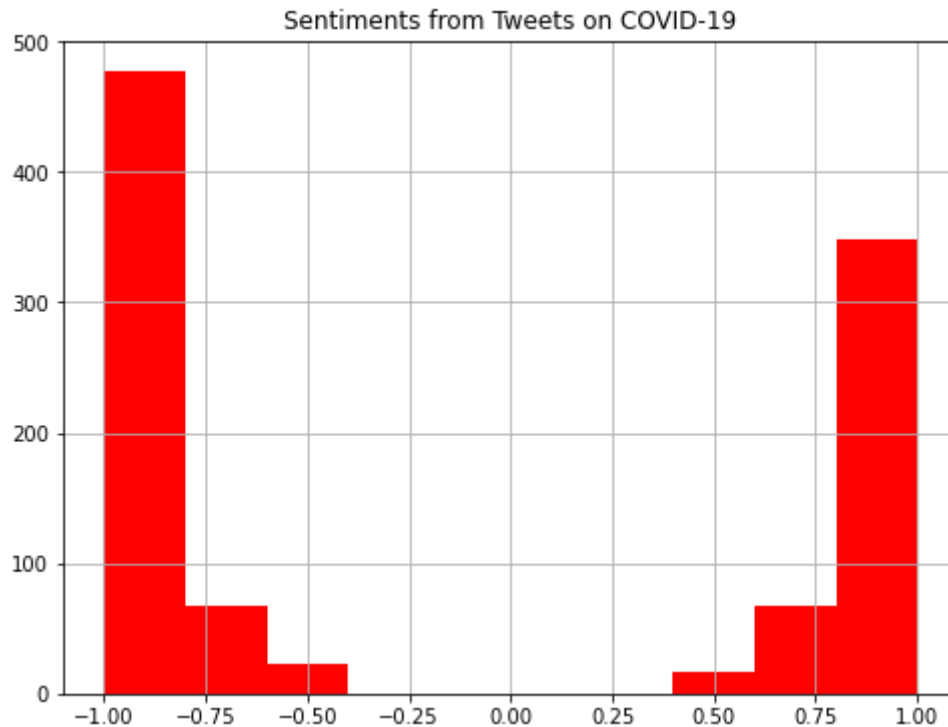


Figure 1.

As we can see in Fig. 1. tweets were very polarized. Most of them had extreme values of sentiment. There were more tweets with negative sentiment (667) than with positive (443).

The automatic system of classification of tweets did quite well in our opinion. To inspect if the classification was right or wrong we selected 30 negative and positive sentiment tweets and decided whether they are positive or negative. After our inspection of most polarized tweets the ratio of correctly classified tweets to incorrectly was 23:1 in positive sentiment tweets and 24:1 in negative sentiment tweets. The system made a mistake only in the case of sentences in which the sentiment depended strongly on the understanding of the context. Like in this example:

“Ohio Attorney General Dave Yost is taking an Athens couple to court for allegedly hoarding hundreds of bottles of hand sanitizer during the COVID19 pandemic and price gouging consumers” - the things those people did are negative but the fact that they are going to court and answer for them is a positive information.

Some of the tweets were hard to understand, and made no sense to us, examples: “Me to Covid19 Look what you made me do Comment nice for superpowers” ; “Question 9 How much revenue does Promega invest in RampD Hint You may need a corporate mind to

find this answer Find amp submit it at the link below for a chance to win a Promega water bottle or a 200 donation to a COVID19 charity US amp Canada only”. These kinds of tweets were not taken into account in our inspection.

The following analysis of content of tweets was based on 100 most polarized positive and negative tweets (in sum 200 tweets). The tweets with positive sentiment mainly concern the subject of good things that are happening during the pandemic. They are referring to outside sources of information about the new features and advancements provided by many companies. Those somehow play a role of advertisement. Others are private amusement about people getting closer in this tough time. There are also tweets that state how many people recover and express sympathy and support for the medical staff. There are also tweets expressing gratitude for workers in restaurants like McDonald’s. Another thing worth mentioning is that the subject of mental health is often brought up, either as a set of advice, a reminder to take care of it or just a fact of a growing need for mental health help. Also users tend to give advice about things related to the life during the pandemic.

The words that were commonly used besides COVID19, pandemic and virus were slightly different for negative and positive sentiment. For example in positive sentiment were:

- **for** (as in medicine for, solutions for, test for, power yoga for),
- **how** (as in some advices for how to stay safe etc.),
- **our** (join our, keep our branches, keeping our staff, tips for our community),
- **during** (life during and after covid, hiring and recruiting during covid, cope with stress during covid, needs of the community during covid, continue learning during covid),
- **help** (things help boost your immune system, help landlords understand, equipment to help our children, launched to help retail sector),
- **new** (adjust to new normal, new career opportunities, trying new restaurants).

While in the negative sentiment the words were:

- **for** (as in waiting for, tested positive for),
- **people** (as in negative things about people who do something),
- **realDonaldTrump** (as people are tagging him and criticising his decisions and lack of knowledge),
- **you** (let me correct this for you, sorry to inform you, if you refuse to something, who the hell are you to say anything)

Comparison of the word 'you' in both sentiments:

- **Positive 'you'** (as in: you could walk daily, solutions you need, thank you salt yoga, who has entertained you most),
- **Negative 'you'** (let me correct this for you, sorry to inform you, if you refuse to something, who the hell are you to say anything)