Hochschule für Technik
und Wirtschaft Berlin

University of Applied Sciences

Project Studies

[Semester 1: Pro-ITD]

Project Report

Sentiment Analysis of Social Media Discourses Related to Vaccines

By: Abdul Hadi Aamir & Alexandra Dubovskaya
Mentor: Professor Helena Mihaljevic
Date: 6th May, 2022

Table of contents

1) Abstract

1.1 Background

The COVID-19 pandemic was an extraordinary occurrence in modern human history, and the rapid spread of vaccines against COVID-19 virus was also an exceptional event. During severe stages of pandemic in 2020 and 2021 a number of vaccines against COVID-19 were introduced and were used worldwide in a very short time. For governments it was crucial to gain public trust in these vaccines to ensure collective immunity through vaccination.

Our research interest is focused on the question how populations react to such events? Have the authorities succeeded in gaining public trust?

In order to investigate it we gauged public sentiment regarding vaccines by analyzing social media. Generally, in the 21st century, social media provides the best insights; data collection and its analysis have long replaced traditional survey methods. Thus, collecting data on social media was the preferred choice for this project.

1.2 Project Scope & Objectives

The information that numerous individuals emit every day ranging from certain products, news, occasions to general public and so forth, have a very significant and meaningful amount. The following was taken from Statista that shows such a drastic increase in monthly active users on Twitter from 2010 to 2019. This indicates that we needed to utilize Twitter as the data source in order to acquire the information we needed, since people express themselves quite a lot here.

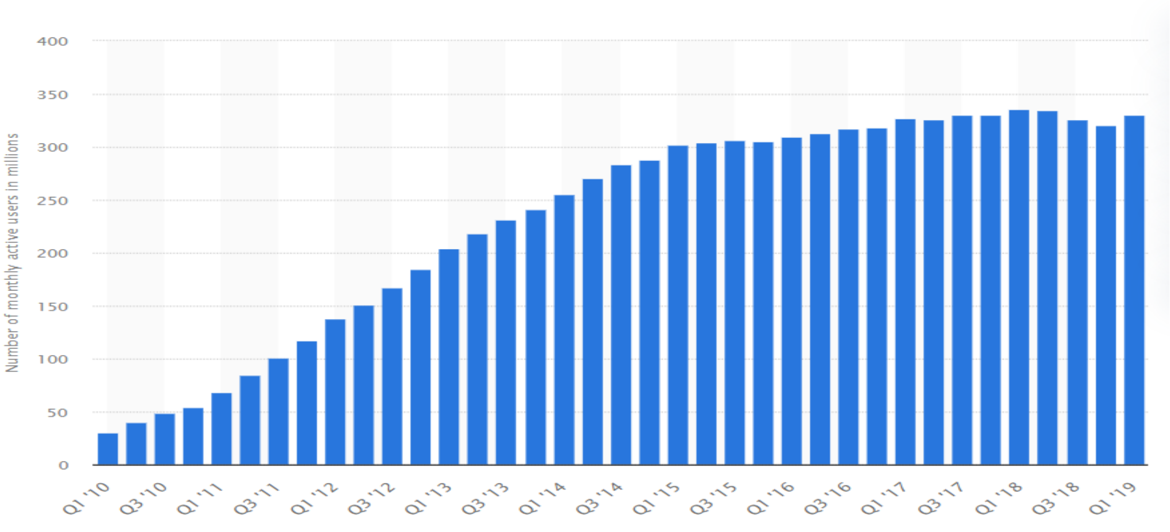


Figure 1.0 Statista Report - Twitter Users (2010 – 2019)

Hence, in this research project, we examined social media discourse on Twitter, in order to understand attitudes and public sentiments towards COVID-19 vaccines. Given that sentiments can be divided into positive, negative and neutral, our aim is to find out which sentiment is dominant.

1.3 Methods

We used social media posts (called Tweets) on Twitter **over 2 different periods (each of 60 days)**; (1) When the vaccines were started and (2) one year later. To prevent our research from being too extensive, we focused on one specific region: **the United Kingdom (UK)** for an English Language Corpus**.** Using the sentiment analysis, we identified different types of sentiments regarding vaccines that were expressed by different people on Twitter. This research report presents all the tools and methods we used to achieve our end goal and assumptions about what the future possibilities could be. From data collection to using VADER for the sentiment Classification, we were able to achieve a lot on the **100,000 tweets** from both timeframes combined. The flow diagram *Figure 1.1* is also presented ahead that portrays all the steps we went through during the research period for this project. We used the *nltk* library for Python for all our Pre-Processing and VADER based functions and procedures as it was a widely popular one with lots of community support and versatility for Sentiment related projects.

1.4 Results

After careful removal of all duplicates and retweets from our datasets during the pre-processing stages, we were able to bring down the total tweets to **67,253** for the **60-day timeframes**. We ended up with polarity scores after applying the Sentiment Analysis library for each tweet on each timeframe, and ultimately after sentiment predictions, were able to aggregate the totals for each sentiment (Positive, Negative and Neutral) to have a clearer view on our results. The polarity scores calculated, were extremely useful in the ultimate label of sentiment being attached to each Tweet.

1.5 Conclusions

We believe that similar studies have been made in the effort to shed more light into the Discourse. But our outputs can have a very influential impact on bringing positivity to Covid Vaccine campaigns and Initiatives that Decision makers may seek to consider in order to drive a solution amidst the pandemic that has been damaging the world since the end of 2019.

1.6 Keywords

The keywords used for our context for Twitter are:

'vaccine', 'vaccines','corona vaccine','corona vaccines', ‘pfizer','biontech', 'moderna', 'Pfizer-BioNTech', 'Pfizer/BioNTech', 'Pfizer BioNTech",'COVAX', 'Sinopharm', 'Sinovac', 'AstraZeneca', 'Sputnik V', 'Gamaleya'.

2) Underlying Flow

2.1 Process Diagram

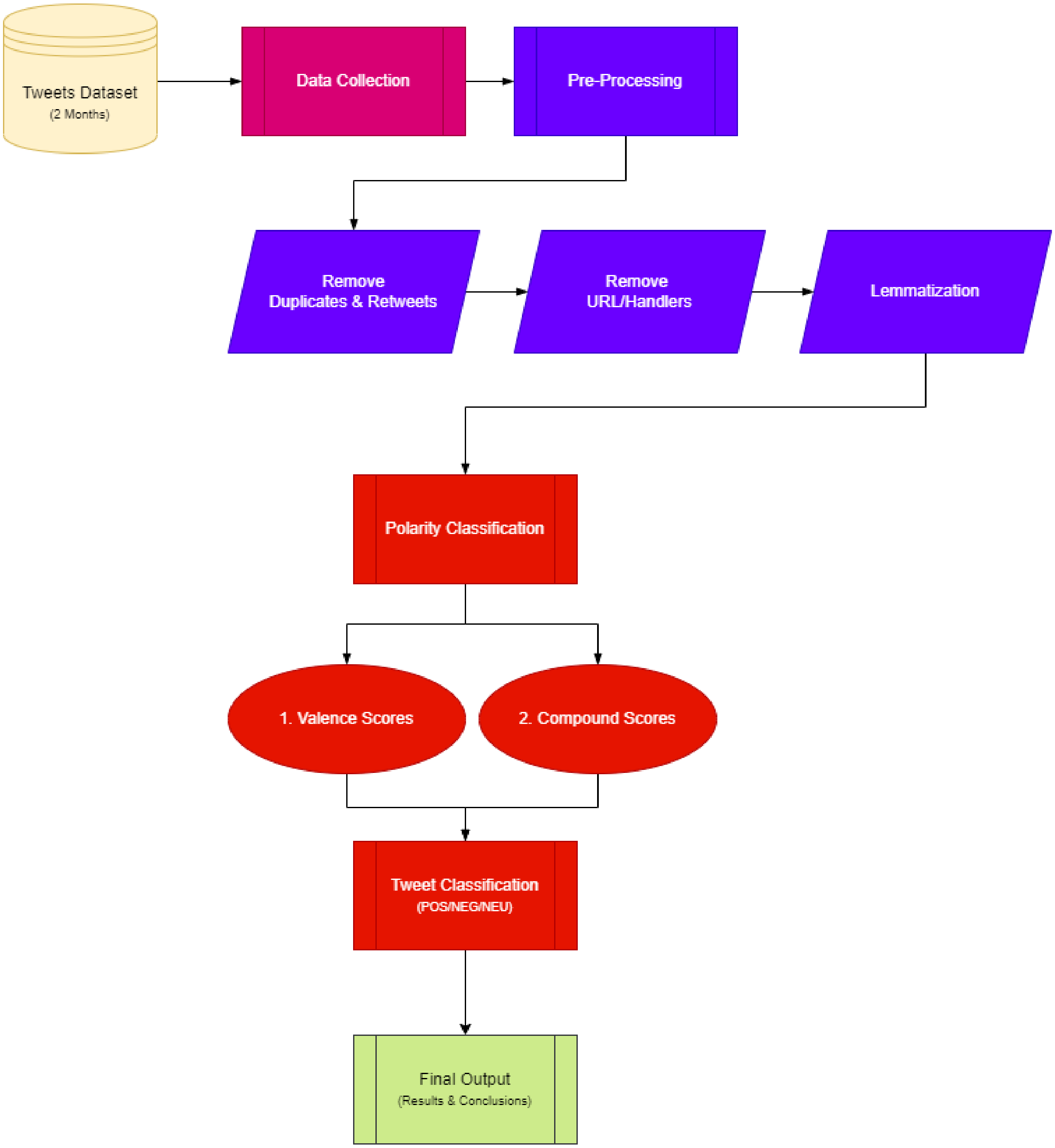
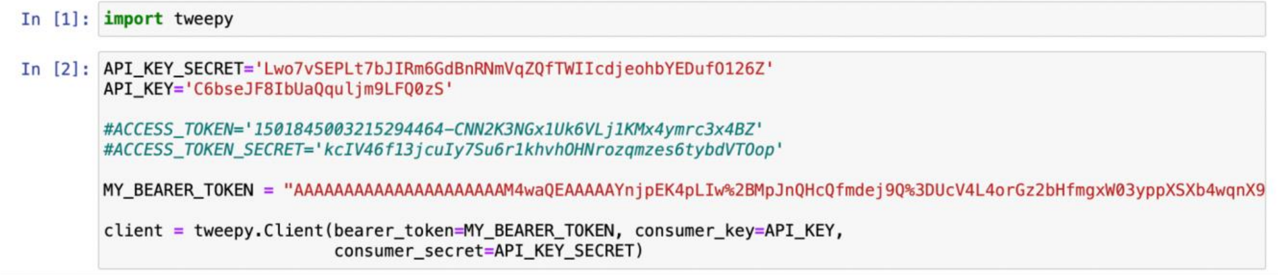


Figure 1.1 Sentiment Analysis Process

3) Phase 1: Data Collection

3.1 Tweepy & the Twitter Developer Account

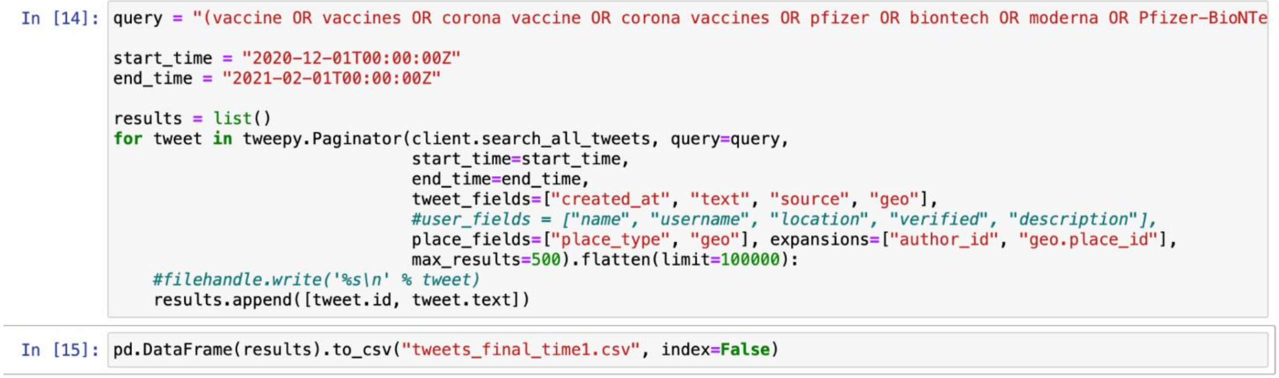
In order to gain access to the vast datasets for our whole project, we needed to use some API that allowed us to interface easily with Twitter and help us easily extract and figure out the relevant dataset spanning the targeted timeframes for all tweets by different users. Tweepy helped us achieve that functionality. And therefore, we signed up for a Developer Account for Twitter API and configured an authorization. In Particular, we used Academic Research Access.



```
In [1]: import tweepy

In [2]: API_KEY_SECRET='Lwo7vSEPLt7bJIRm6GdBnRNmVqZQfTWIIcdjeohbYEDuf0126Z'
        API_KEY='C6bseJF8IbUaQquljm9LFQ0zS'

        #ACCESS_TOKEN='1501845003215294464-CNN2K3NGx1Uk6VLj1KMx4ymrc3x4BZ'
        #ACCESS_TOKEN_SECRET='kcIV46f13jcuIy7Su6r1khvhOHNrozqmzes6tybdVTOop'

        MY_BEARER_TOKEN = "AAAAAAAAAAAAAAAAAAAAAM4waQEAAAAAYnjpEK4pLIw%2BMpJnQHcQfmdej9Q%3DUcV4L4orGz2bHfmgxW03yppXSXb4wqnX9

        client = tweepy.Client(bearer_token=MY_BEARER_TOKEN, consumer_key=API_KEY,
                               consumer_secret=API_KEY_SECRET)
```

Figure 1.2 Tweepy Auth Credentials

As the above screenshot suggests, making an import and then setting up the API Key credentials for the Twitter Developer Account helped us connect the missing dots after which we initialized our client that was used in connecting with the configured credentials.



```
In [14]: query = "(vaccine OR vaccines OR corona vaccine OR corona vaccines OR pfizer OR biontech OR moderna OR Pfizer-BioNTe

         start_time = "2020-12-01T00:00:00Z"
         end_time = "2021-02-01T00:00:00Z"

         results = list()
         for tweet in tweepy.Paginator(client.search_all_tweets, query=query,
                                       start_time=start_time,
                                       end_time=end_time,
                                       tweet_fields=["created_at", "text", "source", "geo"],
                                       #user_fields = ["name", "username", "location", "verified", "description"],
                                       place_fields=["place_type", "geo"], expansions=["author_id", "geo.place_id"],
                                       max_results=500).flatten(limit=100000):
             #filehandle.write('%s\n' % tweet)
             results.append([tweet.id, tweet.text])

In [15]: pd.DataFrame(results).to_csv("tweets_final_time1.csv", index=False)
```

Figure 1.3 Querying Tweets with intended timeframes

3.2 Query Parameters & Timeframes

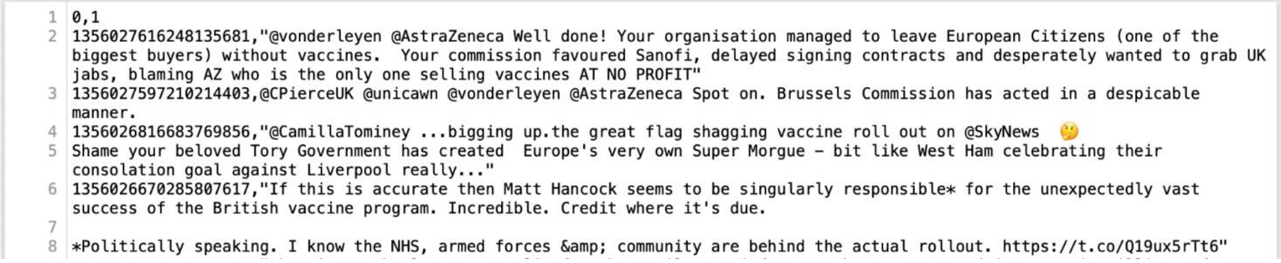
In order to make sure we were going to get the relevant tweets inside our dataset, we made sure to refine our query string parameter as much as we possibly could so that we could cover a broad spectrum of related context for the Discourse. The exact keywords we targeted are mentioned on *Page 4* above.

The timeframes we chose for this period were:

1. **December 2020 — February 2021**
2. **December 2021 — February 2022**

Both being extremely crucial as the first timeframe helped highlight some time into the pandemic to see how it suddenly changed the tables of the world, and the next one being crucial as it depicted the overall gradual inclination of the sentiments of people towards these Covid vaccines in the prolonged stages. Although the code in the screenshots above only shows us for one timeframe, we were later able to use the same code snippet for generating a dataset for the other mentioned time frame.

One thing to be noted is that we converted the whole imported datasets (initially CSVs) into Python Pandas data frames to be able to perform the various operations and Python libraries on.



```
0,1
1356027616248135681,"@vonderleyen @AstraZeneca Well done! Your organisation managed to leave European Citizens (one of the biggest buyers) without vaccines.  Your commission favoured Sanofi, delayed signing contracts and desperately wanted to grab UK jabs, blaming AZ who is the only one selling vaccines AT NO PROFIT"
1356027597210214403,@CPierceUK @unicawn @vonderleyen @AstraZeneca Spot on. Brussels Commission has acted in a despicable manner.
1356026816683769856,"@CamillaTominey ...bigging up.the great flag shagging vaccine roll out on @SkyNews  🤔
Shame your beloved Tory Government has created  Europe's very own Super Morgue - bit like West Ham celebrating their consolation goal against Liverpool really..."
1356026670285807617,"If this is accurate then Matt Hancock seems to be singularly responsible* for the unexpectedly vast success of the British vaccine program. Incredible. Credit where it's due.

*Politically speaking. I know the NHS, armed forces &amp; community are behind the actual rollout. https://t.co/Q19ux5rTt6"
```

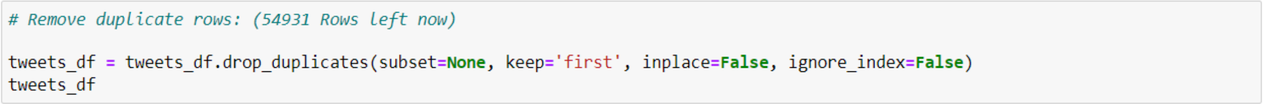
Figure 1.4 Data as dataframe -only EN language with retweets filtered out

4) Phase 2: Pre-Processing & Lemmatization

4.1 Removing Duplicate Records & Retweets

Let’s start with the possibility of having thousands of unrelated and redundant data that may have been gathered during the data collection phase. This data was raw and was of no direct use to us as there were many issues that could further hinder our Sentiment Analysis process. We started off with first checking and filtering for duplicate records in each dataset. We then checked to see if there were any retweets that could be taken care of.

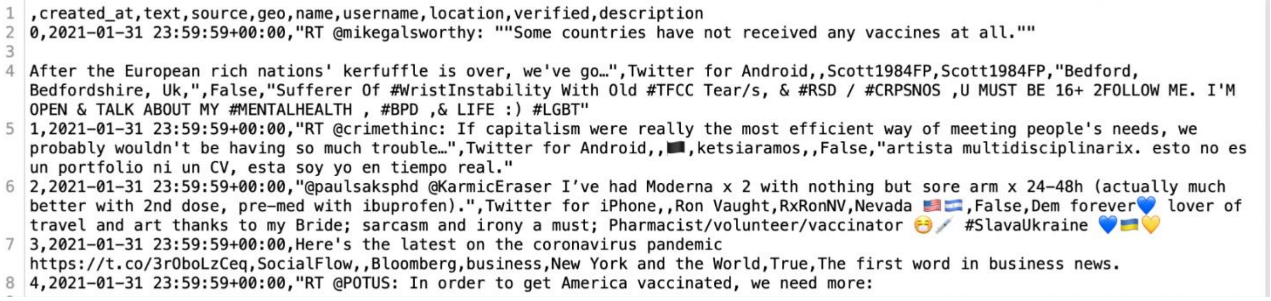
We considered removing single characters and stop words too but, we found out that they hardly or rarely even affected the operations as no significant changes were made after doing so. For example, in one of our datasets, we noticed that the count after doing the pre-processing with other steps reduced the number of tweets to ***19,569 rows*** which remained unaltered even after applying removal mechanisms for single characters and stop words. Therefore, we decided to exclude the pre-processing stages for these. The following code allowed us to handle duplicates on our Pandas data frame.



```
# Remove duplicate rows: (54931 Rows left now)

tweets_df = tweets_df.drop_duplicates(subset=None, keep='first', inplace=False, ignore_index=False)
tweets_df
```

Figure 1.5 Removal of Duplicates

For handling the retweets, we did this earlier on while collecting the data, as the Tweepy API already had built-in methods we could use to do so. We removed Retweets because we wanted to have a guarantee that we were using the original text of the tweets and that viral messages could be retweeted a number of times which was actually not so useful for our research.



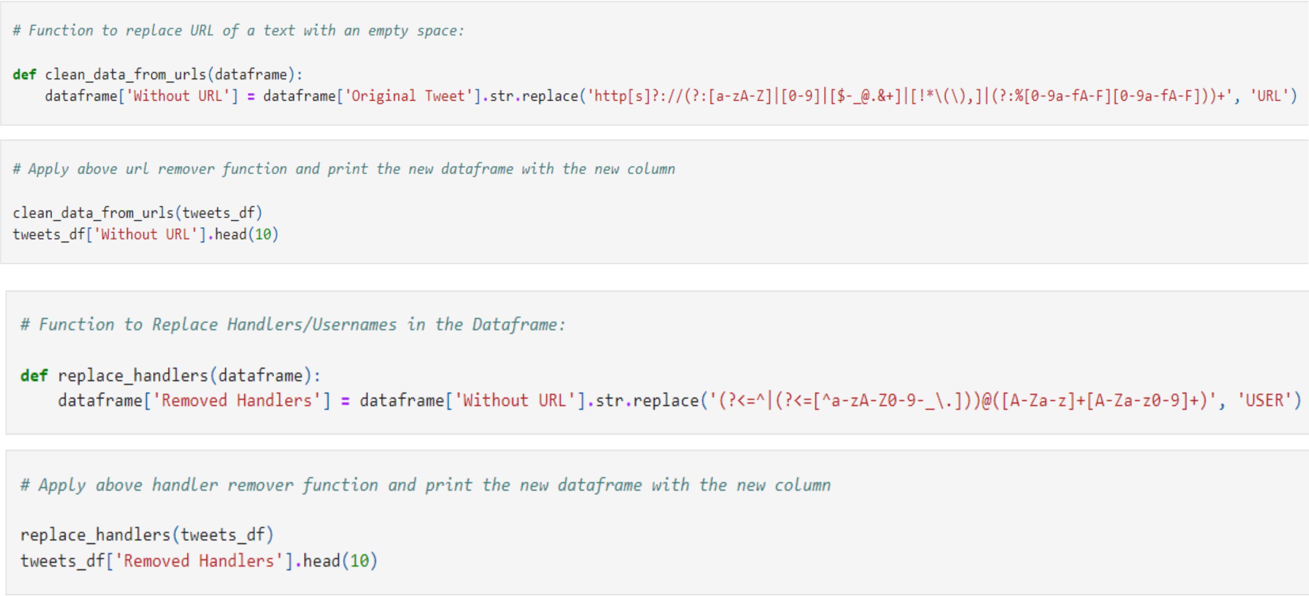
```
,created_at,text,source,geo,name,username,location,verified,description
0,2021-01-31 23:59:59+00:00,"RT @mikegalsworthy: ""Some countries have not received any vaccines at all.""

After the European rich nations' kerfuffle is over, we've go…",Twitter for Android,,Scott1984FP,Scott1984FP,"Bedford,
Bedfordshire, Uk,",False,"Sufferer Of #WristInstability With Old #TFCC Tear/s, & #RSD / #CRPSNOS ,U MUST BE 16+ 2FOLLOW ME. I'M
OPEN & TALK ABOUT MY #MENTALHEALTH , #BPD ,& LIFE :) #LGBT"
1,2021-01-31 23:59:59+00:00,"RT @crimethinc: If capitalism were really the most efficient way of meeting people's needs, we
probably wouldn't be having so much trouble…",Twitter for Android,,🏴,ketsiaramos,,False,"artista multidisciplinarix. esto no es
un portfolio ni un CV, esta soy yo en tiempo real."
2,2021-01-31 23:59:59+00:00,"@paulsaksphd @KarmicEraser I’ve had Moderna x 2 with nothing but sore arm x 24-48h (actually much
better with 2nd dose, pre-med with ibuprofen).",Twitter for iPhone,,Ron Vaught,RxRonNV,Nevada 🇺🇸🇳🇮,False,Dem forever💙 lover of
travel and art thanks to my Bride; sarcasm and irony a must; Pharmacist/volunteer/vaccinator 😷💉 #SlavaUkraine 💙🇺🇦💛
3,2021-01-31 23:59:59+00:00,Here's the latest on the coronavirus pandemic
https://t.co/3rOboLzCeq,SocialFlow,,Bloomberg,business,New York and the World,True,The first word in business news.
4,2021-01-31 23:59:59+00:00,"RT @POTUS: In order to get America vaccinated, we need more:
```

Figure 1.6 Data collection handling retweets

4.2 Removing URL & Handlers/Usernames

Now, predicting the sentiments do not require us to have URLs in our vocabulary as they are not words or sentences that can have meaning and therefore, do not give us anything when we analyze the text from words. But completely removing the same would also not be a wise idea as we can lose a somewhat valuable feature. So, we simply replaced them with the string ‘URL’. We also considered taking care of Handlers and Usernames and replaced them with ‘USER’ in a similar fashion as they themselves had no useful contribution to the ultimate dictionary of words we may have required for the Sentiment Analysis procedures. Functions were specifically created to take in each row and process them as necessary.

The following code snippets show how were used Regular Expressions to pick and replace all instances of URLs and Handlers from our acquired datasets:



```
# Function to replace URL of a text with an empty space:

def clean_data_from_urls(dataframe):
    dataframe['Without URL'] = dataframe['Original Tweet'].str.replace('http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+]|[!*\(\),]|(?:%[0-9a-fA-F][0-9a-fA-F]))+', 'URL')
```

```
# Apply above url remover function and print the new dataframe with the new column

clean_data_from_urls(tweets_df)
tweets_df['Without URL'].head(10)
```

```
# Function to Replace Handlers/Usernames in the Dataframe:

def replace_handlers(dataframe):
    dataframe['Removed Handlers'] = dataframe['Without URL'].str.replace('(?<=^|(?<=[^a-zA-Z0-9-_\.]))@([A-Za-z]+[A-Za-z0-9]+)', 'USER')
```

```
# Apply above handler remover function and print the new dataframe with the new column

replace_handlers(tweets_df)
tweets_df['Removed Handlers'].head(10)
```

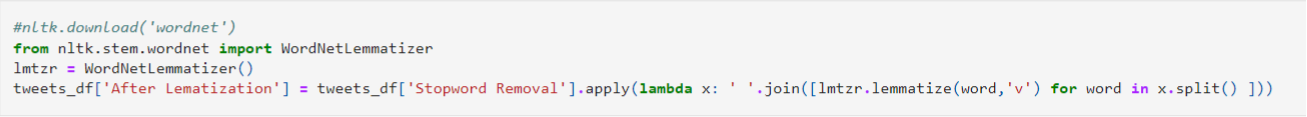
Figure 1.7 URL and Handlers Removal

4.3 Lemmatization

The most important part for us to consider was whether or not to consider and prioritize Lemmatization over Stemming. Both are important for text normalizing as they reduce the words to root forms and is thus used to simplify search queries. Here is what we concluded:

> Since stemming completely devastates the meaning of the word, we end up not having what we need to fuel the Sentiment Analysis process where meaning is of immense context for each and every single word.

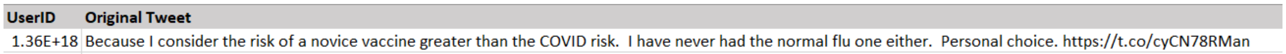
So, Lemmatization seemed like a much better way than stemming to obtain the original form of any given text rather than stemming and lemmatization returns the actual word that has some meaning in the dictionary and also reduces dimensionality of the dataset. We hence proceed to use the *WordNetLemmatizer* provided by the *nltk* library in Python:



```
#nltk.download('wordnet')
from nltk.stem.wordnet import WordNetLemmatizer
lmtzr = WordNetLemmatizer()
tweets_df['After Lematization'] = tweets_df['Stopword Removal'].apply(lambda x: ' '.join([lmtzr.lemmatize(word,'v') for word in x.split() ]))
```

Figure 1.8 Lemmatization

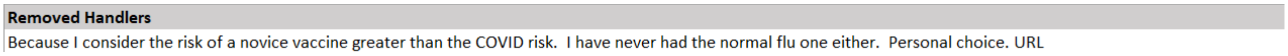
Let's examine what really happens behind the scenes. We pick the following row from our Twitter Dataset from Timeframe 1:



UserID	Original Tweet
1.36E+18	Because I consider the risk of a novice vaccine greater than the COVID risk. I have never had the normal flu one either. Personal choice. https://t.co/cyCN78RMan

Figure 1.8 Sample Tweet in Original form

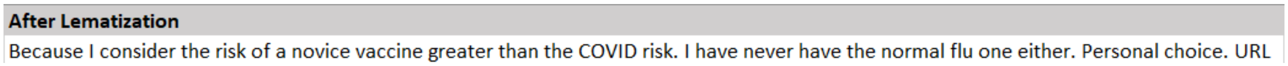
After Applying all the filters, we get the following:



Removed Handlers
Because I consider the risk of a novice vaccine greater than the COVID risk. I have never had the normal flu one either. Personal choice. URL

Figure 1.9 Sample Tweet after URL and Handler Removals

And after ultimately performing Lemmatization, we are returned with the following:



After Lematization
Because I consider the risk of a novice vaccine greater than the COVID risk. I have never have the normal flu one either. Personal choice. URL

Figure 2.0 Sample Tweet after Lemmatization

Here in all the screenshots, we see that the URL ‘’ https://t.co/cyCN78RMan” in the original tweet got replaced by “URL” and the WordNetLemmatizer reduced the word “had” to “have”. We also omitted the usage of Part-of-Speech (POS) Tagging as it had no useful impact on the dataset and VADER did not require such an input to be so.

5) Polarity Scoring & Sentiment Analysis

5.1 Working with VADER

In the final stages, we detected and analyzed the sentimental content of tweets. Therefore, in order to detect the sentiment conveyed, we computed the polarity score of our tweets first, and based on this score, we classified the tweets as either positive, neutral, or negative. Our main tool for identifying emotional content is Valence Aware Dictionary (VADER). It is a rule-based model for general sentiment analysis and is currently considered as a gold standard in social media lexicons.

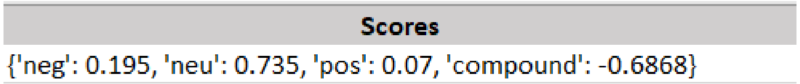
5.2 Polarity Scoring

To first understand polarity scores, we need to understand that each sentiment, whether positive, negative or neutral, needs to be assigned some value to it which would allow us to measure the right sentiment with the most dominating score. And VADER already helps do that by assigning polarity scores to each tweet and thus, returning the results that include the score values for all sentiments. The *overall sentiment* is often inferred as *positive*, *neutral* or *negative* from the sign of the polarity score.

Let's take an example of the following Tweet:

"@CarrDutton @drstevejames Why are all double-barrelled nomenclatures Covid Hysterics?? A vaccine is designed to stop you catching a disease and spreading it. You are aware this does neither so I can't understand your flawed rationale. And you're happy for them to risk their lives for your stupidity. "

After all the necessary pre-processing and Lemmatization steps, we apply the polarity scores method to this text and the result returned is:



Scores
{'neg': 0.195, 'neu': 0.735, 'pos': 0.07, 'compound': -0.6868}

Figure 2.1 Sample Tweet after Polarity Scoring

The result of the scoring method applied also returns us a value for a 'Compound Score' which is basically the aggregate of the positive, negative & neutral scores and is thus normalized between -1(most extreme negative) and +1 (most extreme positive). The above score can also be interpreted as percentages which means it is 7.0% Positive, 19.5% Negative, 73.5% Neutral. While the compound score is 68.68%.

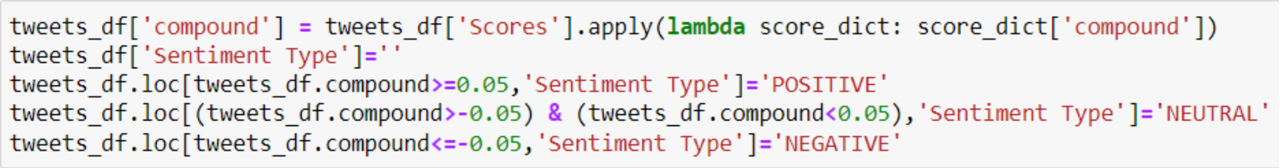
5.3 Compound Scoring & Sentiment Label Classification

Now, we define a threshold that can help us make the classification in a more normalized manner. Therefore, based on the classification thresholds determined by the developers of the library, model assigned sentiments by the following logic:

- **Negative sentiment: compound score <= –0.05**
- **Positive sentiment: compound score >= 0.05**
- **Neutral sentiment: compound score between –0.05 and 0.05**

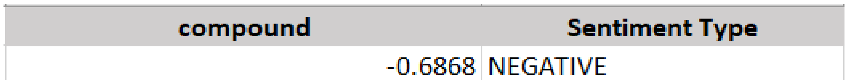
Therefore, in the example above, the tweet fell in the range **<= –0.05** and hence, was assigned a ‘Negative’ label.

More details of how this label was assigned can be analyzed by reviewing the following code for the conditional label procedure:



```
tweets_df['compound'] = tweets_df['Scores'].apply(lambda score_dict: score_dict['compound'])
tweets_df['Sentiment Type']=''
tweets_df.loc[tweets_df.compound>=0.05,'Sentiment Type']='POSITIVE'
tweets_df.loc[(tweets_df.compound>-0.05) & (tweets_df.compound<0.05),'Sentiment Type']='NEUTRAL'
tweets_df.loc[tweets_df.compound<=-0.05,'Sentiment Type']='NEGATIVE'
```

Figure 2.2 Code for Label Classification Based on Compound Score



compound	Sentiment Type
-0.6868	NEGATIVE

Figure 2.3 Sample Tweet after Sentiment Assigning

Thus, we use the code to process each and every tweet which we got after Lemmatization, and acquire the compound scores along with the right Sentiment Label for each.

An even better understanding of this can also be demonstrated by 3 further examples and their analysis as follows:

Tweet, first time period, line 7
@BainsHarjyot It’s progress it’s great to see. I see the daily vaccine figures today it really gave me a much needed lift. B.
Scores: {'neg': 0.0, 'neu': 0.67, 'pos': 0.33, 'compound': 0.7845} **Compound score**: 0.7845 **Sentiment Type**: Positive

Tweet, first time period, line 200:
@SandieBlickem No evidence for that. In fact no evidence that ANY of the vaccines have lasting efficacy.
Scores: {'neg': 0.18, 'neu': 0.82, 'pos': 0.0, 'compound': -0.296} **Compound score**: -296 **Sentiment Type**: Negative

Tweet, first time period, line 498:
@lesserspottedH And why are they trying to gag the Scottish govt from saying how much vaccine there getting by stating national security.
Scores: {'neg': 0.152, 'neu': 0.696, 'pos': 0.152, 'compound': 0.0} **Compound score**: 0 **Sentiment Type**: Neutral

Referenced from: Model Built(Initial - Timeframe1).csv

6) Results, Conclusion & Future Work

6.1 Results & Interpretation

Now coming down to what we were above to obtain after the whole Sentiment Analysis procedure. We wanted to learn more about how the data changed and inclined when we decided to compare the outputs of both Time Frames together.

And the findings were that the sentiment analysis process showed that for the first timeframe, 48,8% (23 282) of tweets were positive, 28,8% (13 710) of the tweets were negative, and 22,4% (10 691) of tweets were neutral.

Total (Positive)	Total (Negative)	Total (Neutral)
23 282	13 710	10 691

Figure 2.4 Sentiment Count for First Timeframe

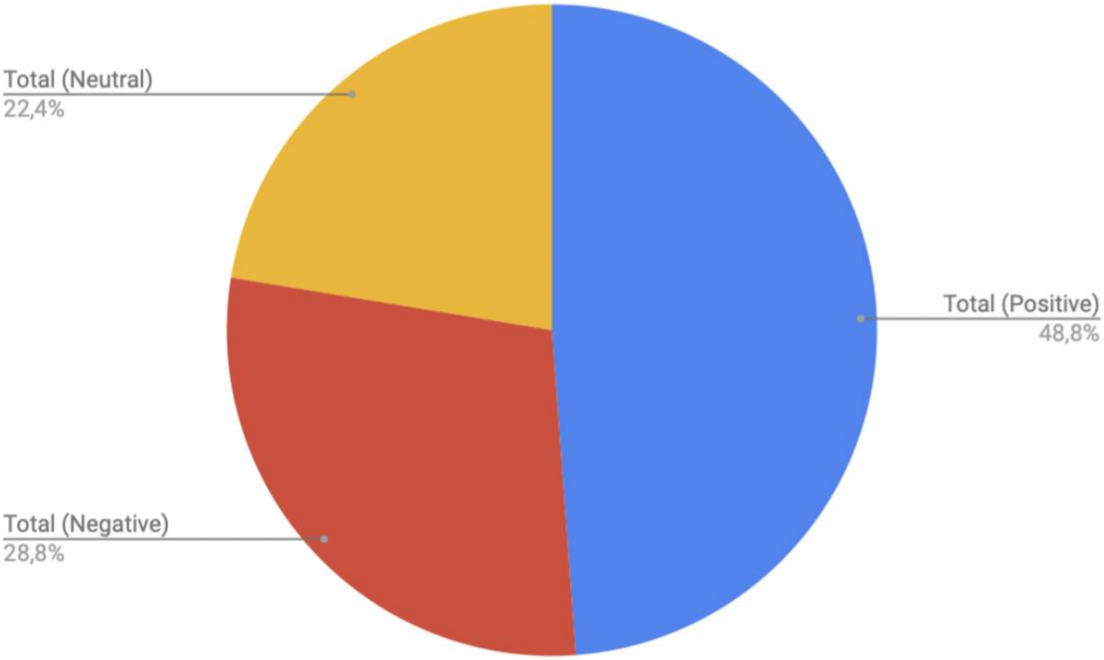


Figure 2.5 Pi Chart for First Timeframe

And, for the second time period we have 44,9% (8 785) of tweets positive, 35,6% (6 971) of the tweets negative, and 19,5% (3 813) of tweets neutral.

Total (Positive)	Total (Negative)	Total (Neutral)
8 785	6 971	3 813

Figure 2.6 Sentiment Count for Second Timeframe

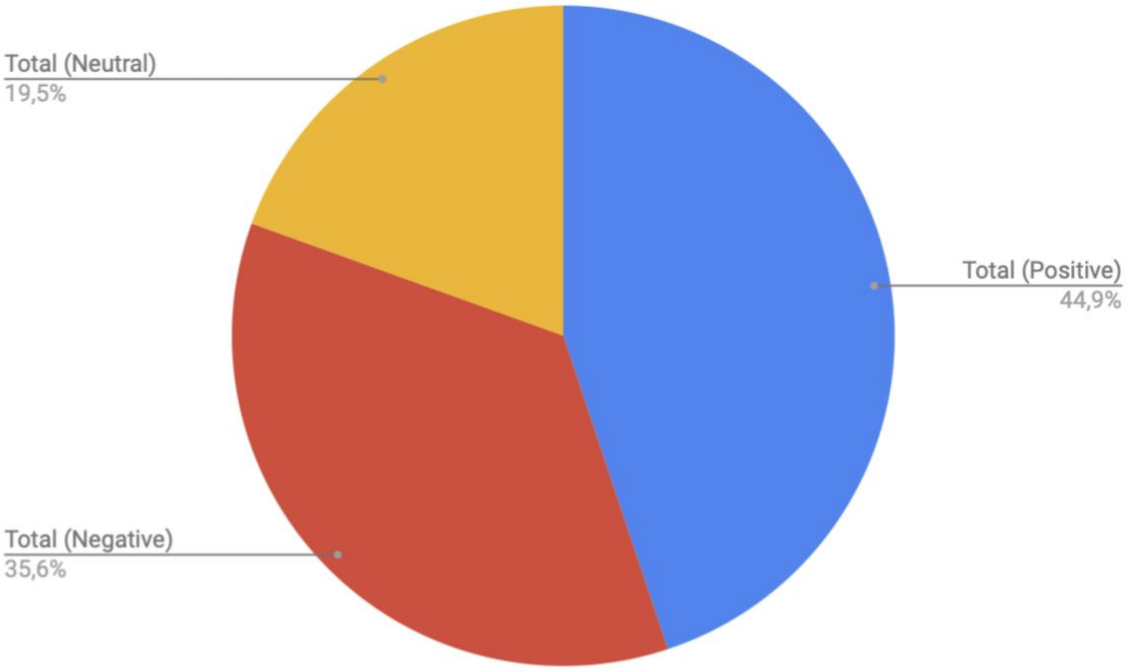


Figure 2.7 Pi Chart for Second Timeframe

We can see that positive tweets dominate in both periods, with a higher proportion in the first period relative to the total. While the proportion of negative tweets in the second period is higher by 6.8% compared to the first period. Thus, generally, we can say that people expressed more positive sentiments than negative, while neutral sentiments are 19-22% for both periods. Additionally, for the second period, the total number of tweets is lower than for the first period. This allows us to suggest that interest in expressing an opinion on vaccines decreased by the second period.

6.2 Conclusion & Future Work

So, what we were able to build so far, we think such a solution can play a vital role in similar events whether or not this has to be for Vaccines necessarily. It can provide key insights in other domains too such as Political Voting, Market and Competitor Researches, Customer Support Ticket Analysis etc. Therefore, similar methodologies can be used to predict sentiments in various other dataset contexts to better fit the respective needs.

As we mentioned, we had been able to build up a decent model with a very high accuracy rate for this Project. To summarize everything, both the positive and neutral sentiments together dominate over negative. That means that authorities succeeded in gaining public trust.

This leaves us with limitless opportunities about the future research and methodologies and how more scientific researchers can dive deeper into the domain for figuring out whether vaccines are really the solution to counter-acting massive-scale pandemics in the future.

As an example of future work, which could be based on the open-ended questions of the current research, we could investigate which sentiment-related words occur most frequently in each sentiment group.

7) Related Research Works

Before starting off with the research, we investigated the related research we could come across for getting the relevant domain and expertise.

Several researches, for example, *building public trust: a response to COVID-19 vaccine hesitancy predicament* [10], are focused on researching public trust and particularly on different social media. Researchers have been able to identify dominant sentiments in different demographic groups over different time periods.

Among others, we highlighted the research named Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis [1], where authors used topic detection and sentiment analysis as social media trend analysis to better understand the discourse on COVID-19 vaccines tweets. The authors identified the trending topics that reflected the public concerns on COVID-19 vaccines and their responses to the topics indicated by the polarity and emotions on the sentiments. Among other outcomes, the researchers found that the administration and access to vaccines were some of the major concerns.

Unlike the above-mentioned study, in our research we did not use topic detection and did not classify emotions using BERT or any similar methods for computing Cosine Similarities to associate fine-grained emotions as this was beyond the scope of our project.

8) Acknowledgement

Without the constant support of our supervisor, Professor Helena Mihaljevic, this paper and the research that went into it would not have been feasible. From our initial interaction with the Data Collection of the different timeframes from Twitter to the final draft of this dissertation, her zeal, knowledge, and meticulous attention to detail have been an inspiration and kept our work on track. This report corresponds with the Project Studies module of the Degree Programme of the “Professional IT Business and Digitalization” and therefore is intended to shed light on the various phases we went through in order to achieve the desired results.

8.1 Source Files

I also hereby acknowledge that the work done for all this research and model creation is entirely the effort and investment of both team members and external sources were thus, only used to guarantee that both members were equipped with the right domain and toolset in order to help proceed the project to successful completion.

All source files along with the result CSV files can be found on the Google Drive link:

https://drive.google.com/drive/folders/1s0dmV-3-o91ImNyJa8u7LCdciG5IO1ii?usp=sharing

The order of execution of files are as follows:

1. Data-Collection.ipynb
2. Pre-Processing.ipynb
3. Lemmatization.ipynb
4. VADER.ipynb

9) Conflict of Interest

None Declared.

10) References:

[1] Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis https://www.jmir.org/2021/10/e30765

[2] Using VADER https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664

[3] VADER Documentation https://github.com/cjhutto/vaderSentiment

[4] nltk Python Package for Sentiment Analysis https://www.nltk.org/api/nltk.sentiment.html

[5] Calculating Polarity Scores https://www.analyticsvidhya.com/blog/2021/12/different-methods-for-calculating-sentiment-score-of-text/

[6] Sentiment Analysis Domain Study => HTW Berlin Pro-ITD Data Science Lectures

[7] Tweepy Documentation https://docs.tweepy.org/en/stable/client.html

[8] Twitter API Documentation https://developer.twitter.com/en/docs/twitter-api

[9] Prabowo R, Thelwall M. Sentiment analysis: A combined approach. Journal of Informetrics 2009 https://www.sciencedirect.com/science/article/abs/pii/S1751157709000108?via%3Dihub

[10] Vergara R, Sarmiento P, Lagman J. Building public trust: a response to COVID-19 vaccine hesitancy predicament. J Public Health http://europepmc.org/article/MED/33454769

[11] COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis https://pubmed.ncbi.nlm.nih.gov/34115608/

[12] Tool for building Regex expressions https://regex101.com/

[13] Documentation to facilitate communication between dataset creators and consumers https://dl.acm.org/action/downloadSupplement?doi=10.1145%2F3458723&file=gebruappendix.pdf

GitHub Url: https://github.com/abdulhadi25/SentimentAnalysis
