



ISSN: 2723-9535

Available online at www.HighTechJournal.org

HighTech and Innovation Journal

Vol. 2, No. 4, December, 2021



Towards a Sentiment Analysis of Tweets from Online Newspapers Regarding the Coronavirus Pandemic

Giulia Pes¹, Angelica Lo Duca^{2*} , Andrea Marchetti²

¹ *University of Pisa, Pisa, Italy.*

² *Institute of Informatics and Telematics, National Research Council, Pisa, Italy.*

Received 02 April 2021; Revised 06 October 2021; Accepted 11 November 2021; Published 01 December 2021

Abstract

In the last year, both offline and online news have had the Coronavirus pandemic as their subject, especially since social networking such as Twitter has significantly increased the news regarding Covid-19. The objectives of the project are: the analysis of news regarding the Coronavirus pandemic was extracted from the Twitter profile of ANSA, a well-known Italian news agency, and the analysis of sentiment and the number of likes for each news extracted. The sentiment analysis has been carried out using the MAL lexicon (Morphologically Affective Lexicon), where the tweet is split into words and each word is associated with a score. Positive (with a score greater than zero), negative (with a score less than zero) and neutral (with a score equal to zero) news were identified. As a result, it emerges that sentiment changes day by day, so it is necessary to use sentiment indicators called indices, but only the positive sentiment index is taken into consideration as the negative one is complementary and the neutral one is almost zero. The positive index is then related to some parameters extrapolated from the Civil Protection site: the number of cases, the number of deaths, and the entry into intensive care. Furthermore, in addition to the parameters listed above, the positivity index is related to the days on which the Prime Minister's Decree (DPCM) was signed. The last relationship analyzed is that between the average number of likes and the number of deaths. The results of the research show that the sentiment of the news from the ANSA Agency contains 62.3% of positive news, 37.3% of negative news, and only 0.3% of neutral news. Furthermore, sentiment is not influenced by the daily parameters: the number of cases, number of deaths, entry into intensive care units, and DPCMs. But there is a relationship between the average of like and the number of deaths.

Keywords: Coronavirus; Covid-19; Pandemic; SWABS; ANSA; Sentiment Analysis; Civil Protection; DPCM; Italy.

1. Introduction

In the modern world, the growth of social data on the web is constantly increasing. Researchers have access to data in real-time for research and information purposes [1]. In the last year, a significant part of the news and information both offline and online available on the web have had as their subject the Coronavirus pandemic (also known COVID-19 outbreak). The consequent actions taken by their respective governments against the disease have produced a series of rapidly evolving sentiments regarding the issue [2]. In fact, during the global COVID-19 pandemic, many companies and people have published and shared their points of view [1]. With the spread of awareness of the discomfort arising from the disease, the messages, videos, posts, and tweets related to COVID-19 have also increased. In fact, messages with negative feelings regarding COVID-19, pandemic, and lockdown have been increasingly frequent [3]. In a very short space of time, Twitter (<https://twitter.com/>) showed a similar effect with the growth of an exponential number of Coronavirus-related tweets [4].

* Corresponding author: angelica.loduca@iit.cnr.it

 <http://dx.doi.org/10.28991/HIJ-2021-02-04-08>

➤ This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

© Authors retain all copyrights.

The aim of this paper is to provide an analysis of the sentiment of the news, day by day, of the ANSA Agency and of people's likes for each piece of news. We wish to point out the criteria that constantly influence the trend of the sentiment of the news over time published by the ANSA Agency and whether this can be correlated with the daily data of the pandemic extracted from the Italian Civil Protection website. The analysis carried out on the data made it possible to trace the profile of the news published by the ANSA Agency on the pandemic and the way in which it manages it.

In detail, the present paper is focused on the analysis and in-depth analysis of news regarding the Coronavirus from the User profile of the ANSA Agency on Twitter*. Tweets were extracted from Tweepy, a software that we implemented for the extraction of tweets written by a single user profile and containing up to five hashtags† [5]. The data extraction period is from 13 October 2020 until 17 January 2021 for a total of 1772 records. These are the data taken into consideration: the text of the news, the number of likes, the number of retweets and the date on which the news was published. Sentiment analysis was then carried out on extracted tweets, this made it possible to assign a sentiment score to each tweet that contained the referenced news. With the data obtained, it was thus possible to calculate the indices of positivity, negativity and neutrality used as indicators of positive, negative and neutral sentiment. The positivity index was subsequently related to the number of cases, the number of deaths and the number of admissions to intensive therapies caused by the COVID-19 outbreak. The data indicated above were acquired from the Civil Protection site which provides total data of the pandemic always updated‡. The three factors were then related to the positive sentiment to determine if the criteria influenced the trend of positivity index of the news of ANSA Agency.

The present study is divided into the following sections: Section 2 describes the state of the art where the analysis of sentiment, in general, is analyzed through the scientific literature and then focuses on the analysis of sentiment in relation to the Coronavirus pandemic. Section 3 focuses on the analysis of the data extrapolated. Finally, in Section 4 we give our conclusions.

2. State of the Art

In this section we give an overview of the current literature regarding the sentiment analysis on Twitter, as well as a description of the Coronavirus pandemic.

2.1. Sentiment Analysis on Twitter

Sentiment analysis is a field of natural language processing that deals with building systems for the identification and extraction of opinions, feelings, attitudes, emotions and evaluations found in the text. It is based on the main methods of computational linguistics and textual analysis. It is fundamental for the classification of sentiment to determine the contexts in which a word can take on different meanings. In fact, the language used to express subjective evaluations is very complex and made up of different components. Along with the development of technology and increasing access to information, a new type of society has emerged, that of interaction and communication [6]. Social media offer the perfect harmony between the two and allow people to connect, share opinions and emotions. They also offer news from around the world in real-time and always up to date. One example is the Twitter microblogging site, where people post messages about their opinions on a variety of topics in real time and express positive sentiment for the products they use [7].

In this context, the emotional role is fundamental [8]. Although natural language remains far beyond the power of machines, sentiment analysis can provide a surprisingly significant sense of how news has a strong impact [9]. Indeed, researchers frequently analyze the opinions and sentiment of the news themselves through supervised and unsupervised methods. Through these it is possible to establish the prevailing sentiment in the news and determine if a news is positive or negative [10]. Of fundamental importance is the assignment of scores to probe the sentiment and determine the degree of positivity, negativity and neutrality of the individual items. Items with a score greater than 0 are considered positive, less than 0 negative, and equal to 0 neutral. News can usually be positive, negative but rarely neutral [9]. To carry out a good sentiment analysis, it is also necessary to take into consideration the space where one operates. If you work on blogs or microblogging like Twitter, you must take into account the presence of emoticons and hashtags that add value to the classifier [7]. The past few years have led to a significant growth in the volume of search in sentiment analysis, mainly on highly subjective types of text such as product or movie reviews. In fact it is essential for the marketing of a company to have a background on a certain product or on what people think about the company itself. This type of sentiment analysis can also be very useful for consumers who are trying to research a product [11]. On the other hand, when you want to analyze news articles, it is necessary to address the topic more specifically. News articles and other reports typically contain less clearly expressed ratings than reviews [8]. In this type of analysis, reference is made to the intentionality of the author and therefore whether the latter wants to convey positive or negative feelings depending on the news and the context that surrounds him [12].

* https://twitter.com/Agenzia_ANSA

† <https://github.com/giulipes/Tweepy>

‡ <https://github.com/pcm-dpc/COVID-19>

2.2. Coronavirus Pandemic

The Coronavirus pandemic has triggered an unprecedented crisis. Coronavirus is a semi-flu virus whose epicenter was in Wuhan, a city in China, in December 2019. The causes that led to the emergence of the virus are not yet clear, but the theories are different. People around the world have been forced to stay at home for their safety, limit contact with strangers, and comply with safety measures. In fact, various measures have been implemented to fight the pandemic such as blocking and social distancing which can also lead to mental health problems such as depression, anxiety and sadness [13]. Since the declaration of the first Coronavirus case, the pandemic has been on the offline and online news headlines. This triggered positive or negative responses from readers. The analysis of the sentiment linked to the Coronavirus is one of the most in-depth topics in the last year. In this context, the data provided by social media can be a very important source of information. User-generated messages provide a window into people's minds, allowing us to understand their moods and opinions [14]. Social media has always been widely used as a means of posting and sharing one's views. Large-scale tweets provide an ideal source of data, and sentiment dynamics provide the means to analyze the data.

A study conducted in Bangladesh has highlighted an increasing use of social platform, as people spend most of their time at home due to the virus. News and articles on Coronavirus are read and commented on through social media. Sentiment analysis on article comments categorized some audiences that turn out to be: Analytical, Depressed, and Angry. In this way, public psychology towards the pandemic is traced [15]. In South Korea, research results also suggest a negative predictor of civilization when citizens comment or tweet about COVID-19. However, we must take into account the factor by which it is estimated that Twitter users, having built larger networks and obtained positive responses from others, are more likely to use uncivilized language [16]. However, when sentiment analysis is no longer examined at the level of individual news but at the level of the topic, various aspects of the pandemic are captured. In fact, it is possible that there are more topics with a positive feeling than a negative one. This is because topics such as "staying safe at home" are categorized as positive while "people's deaths" are negative [17]. How different cultures react and respond to a crisis is predominant in the norms and political will of a society to combat the situation. Often the decisions made are necessitated by events, social pressures or needs of the moment, which may not represent the will of the nation. Coronavirus has led to a mix of similar emotions in nations where governments have made similar decisions [3].

In this tense climate, the rise and fall in the number of cases or deaths have become a constant headline in world news. A recent study revealed an approximately 57% increase in viewing news on a TV or smartphone due to lockdowns. During the pandemic, the changing statistics of those affected formed a focus of the news published by the different channels. The result inevitably features a lack of positivity in world news, and there is only a small number of news items delivered on a positive note. The connection between the dependence on the number of cases and deaths and the negative sentiment of the news is therefore evident, even if the situation can change from country to country due to regional socio-political factors [10].

3. Analysis

This paper focuses on the analysis of news regarding the Coronavirus pandemic extracted from the Twitter profile of ANSA, a well-known Italian news agency. The aim of this paper is to analyze the sentiment and the number of likes for each news extracted. The result is in the selection of some keywords for the peaks of sentiment, both positive and negative, which is the focus of the news on the pandemic. Furthermore, we want to understand if the news trend may vary based on some data on the pandemic offered by the results of the Civil Protection: the number of cases, deaths and admissions for intensive care. For each extracted tweet, the following information is stored: the text, the number of likes, the number of retweets and the date of publication of the tweet.

With regard to the text of each tweet, the sentiment was extracted* and subsequently a score greater than zero, less than zero or equal to zero. On the basis of the score, the daily positivity, negativity and neutrality indices were calculated. The indices obtained were compared with those available from the Civil Protection website which offers daily real-time data on the pandemic.

3.1. Data Collection

The data were collected from two main sources: the Twitter profile of the ANSA Agency and the data provided by the Civil Protection regarding daily infections from COVID-19 in Italy.

* The sentiment has been extracted through a free software available at this address: <https://github.com/stepthom/lexicon-sentiment-analysis>

ANSA Agency

The National Associated Press Agency is commonly known by the acronym ANSA. It is the first multimedia information agency in Italy and the fifth in the world. It was founded in Rome in 1945 to succeed the dissolved Stefani agency. The ANSA is a cooperative made up of 36 publishing members of the main Italian newspapers and has the aim of collecting and transmitting news on the main Italian and world events. Nowadays, almost all news agencies have a free site on which ends up only a small part of the content they daily produce [18-20]. The ANSA agency is known for its principles of rigorous independence, impartiality and objectivity enshrined in its Statute and compliance with national and international laws. ANSA's main customers are private TV channels and local newspapers. The main difference between ANSA and the other news agencies is the physical presence on the territory. In fact, the agency has a center in almost every region. The ANSA political desk is in Montecitorio. This guarantees great influence and recognition for journalists who are often, also, among the most expert in politics. The other important part of the reporters' work is that of summarizing laws approved in the courtroom, which are then the same shot and commented on by the journalists and by the news programs. Table 1 shows the fields extracted for each record with their description.

Table 1. Description of the extracted fields

Field	Description
Description	Contains the text of the tweet, it may contain additional hashtags than those specified, links to other users and the news link that connects directly to the ANSA Agency website
Likes	Contains the number of likes of other users in a given news
Retweets	Contains the number of shares of a tweet. How many times that tweet has been shared by other users
Data	Contains the date of publication of the tweet on the user profile

The period examined is between 13 October 2020 and 17 January 2021. The data collected there are a total of 1772 records, relating to all the news regarding the pandemic. Table 2 shows an extract of the obtained dataset.

Table 2. Extracted dataset

Description	Likes	Retweets	Date
La #Scala, A Unique #Covid Premiere. The Stellar Cast, Set Design And The Anthem Sung By The Workers #ANSA @Teatroallascala https://t.co/Lqp3i2utgx	14	4	2020-12-06 08:50:02
# COVID-19 #Usa, Over A Million Cases In 5 #ANSA Days https://t.co/Y7mwdsyhkl	16	10	2020-12-06 09:50:02
#Covid: 18,887 New Cases And 564 Victims In Italy In The Last 24 Hours #ANSA https://t.co/Cysbib309p	31	9	2020-12-06 19:39:50
#Covid, #Abruzzo Ordinance: From Tomorrow Out Of The Red Zone. Marsilio: 'It Turns Orange, I Informed Speranza'. But Government Sources: 'It Has To Wait For Wednesday. The Deadline Of 21 Days Must Be Respected. No Endorsement Of The Decision On Tomorrow ' #ANSA https://t.co/Jbqnuixyth	71	41	2020-12-06 21:30:28
#Covid Bavaria Declares A State Of Calamity , New Restrictions 'We Must Do More'. Lockdown Even More Severe #ANSA https://t.co/Noucqm5m0p	42	11	2020-12-06 22:17:46

Five main hashtags have been identified related to the “Coronavirus” topic: #Covid, # COVID-19, #Coronavirus, #pandemic and #babes. Table 3 shows some statistics regarding each extracted hashtag. Table 3 shows that the hashtag covid produced a greater number of tweets, followed by Coronavirus, COVID-19, pandemic and finally swabs with 20 values. The analysis of the number of likes, on the other hand, identified the maximum, minimum and average values. Through the number of likes extracted, it was possible to determine the average total likes for each tweet which is 40.98. The standard deviation, on the other hand, which is an estimate of the variability of a population of data or of a random variable for the number of likes, is zero.

Table 3. Statistics related to each extracted hashtag

Hashtag	Frequencies	Average number of like	Like minimum value of	The maximum value of like	Maximum length of tweet	Minimum length of the tweet
record #Covid#	1432	39.26	0	947	295	61
# COVID-19	102	38.45	0	415	280	74
#Coronavirus	334	36.86	0	276	280	77
#pandemic	92	42.33	5	185	280	90
#bampons	29	32.0	4	249	275	84
Total	1989					

Data from Italian Civil Protection

The data available on the progress of the Coronavirus pandemic in Italy are available on the Civil Protection website*. The file has a csv extension updated every day and contains: the date, status, hospitalized with symptoms, intensive care, total hospitalized, home isolation, total positives, total positive variation, new positives, discharged healed, deceased, cases of suspected diagnosis, cases from screening, total cases, number of swabs and, finally, the cases tested. The reference period is from October 13, 2020 until January 17, 2021. Table 4 illustrates extracted data.

Table 4. Description of the parameters extrapolated from the Civil Protection Github

Parameters	Description
Date	date information, provided in column number 0
Intensive therapy	total inputs in ICUs , available in the third column of
Deaths	total number of deceased, available in the tenth column
Total_cases	total number of cases, available in column number 13

Table 5 illustrates an example of the dataset of the Protection site Civil.

Table 5. Example of Civil Protection dataset

Date	Intensive Therapy	Death	Total_Cases
2020-10-13T17: 00: 00	514	36246	365467
2020-10-14T17: 00: 00	539	36289	372799
2020-10-15T17: 00: 00	47	83	8803
2020-10-16T17: 00: 00	52	55	10009
2020-10-17T17: 00: 00	67	47	10925

3.2. Analysis of Tweets

Once the tweets have been extracted, an analysis of the news sentiment for each tweet [21]. A score called score was associated with each tweet. The MAL (Morphologically-inflected Affective Lexicon) lexicon was used for sentiment analysis, a Natural Language Processing resource that associates each word in the lexicon with a certain score according to the context in which it occurs [22, 23]. The tweet was first divided into words and subsequently on each word the score was calculated based on the MAL lexicon. The tweet with a score greater than zero was evaluated as a positive tweet, less than negative zero and equal to zero neutral. In this way, the sentiment was obtained for each tweet. The sentiment results showed 1101 positive news (62.3%), 659 negative news (37.3%) and 6 neutral news (0.3%). Sentiment analysis shows that each news article has a different sentiment category from day to day. Normalization is therefore necessary before carrying out comparative studies. For this purpose, daily values called indices are calculated and are used as indicators of the overall negative or positive or neutral sentiment in the news of that day. The various indices are calculated as follows:

Positivity index for day *i*:

$$P(i) = \frac{\text{number of positive tweets} \in \text{day } i}{\text{total number of tweets} \in \text{day } i} \tag{1}$$

Negativity index for day *i*:

* <https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv> .

$$N(i) = \frac{\text{numberofnegativetweets} \in \text{day}i}{\text{totalnumberoftweets} \in \text{day}i} \tag{2}$$

Neutrality index for day i :

$$Neu(i) = \frac{\text{numberofneutraltweets} \in \text{day}i}{\text{totalnumberoftweets} \in \text{day}i} \tag{3}$$

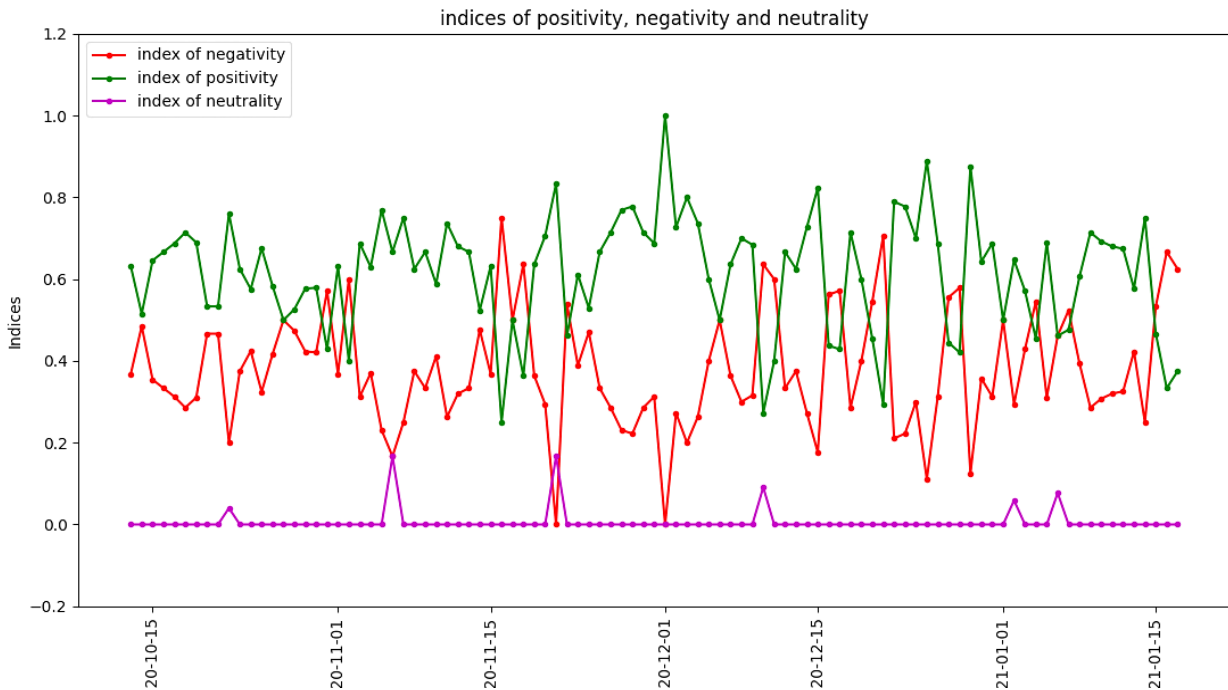


Figure 1. Indices of positivity, negativity and neutrality

Figure 1 shows the values of the positive, negative and neutral indices. The negativity index is indicated with red, positive green and neutral in magenta. The neutrality index rarely deviates from zero. In the positive and negative indices, on the other hand, we find various variations in value, in fact, during the analyzed period; there were some events that led the negative index, positive to the change. To highlight how much the positive and negative indices have changed, they have been related to their average in the following Figure 2.

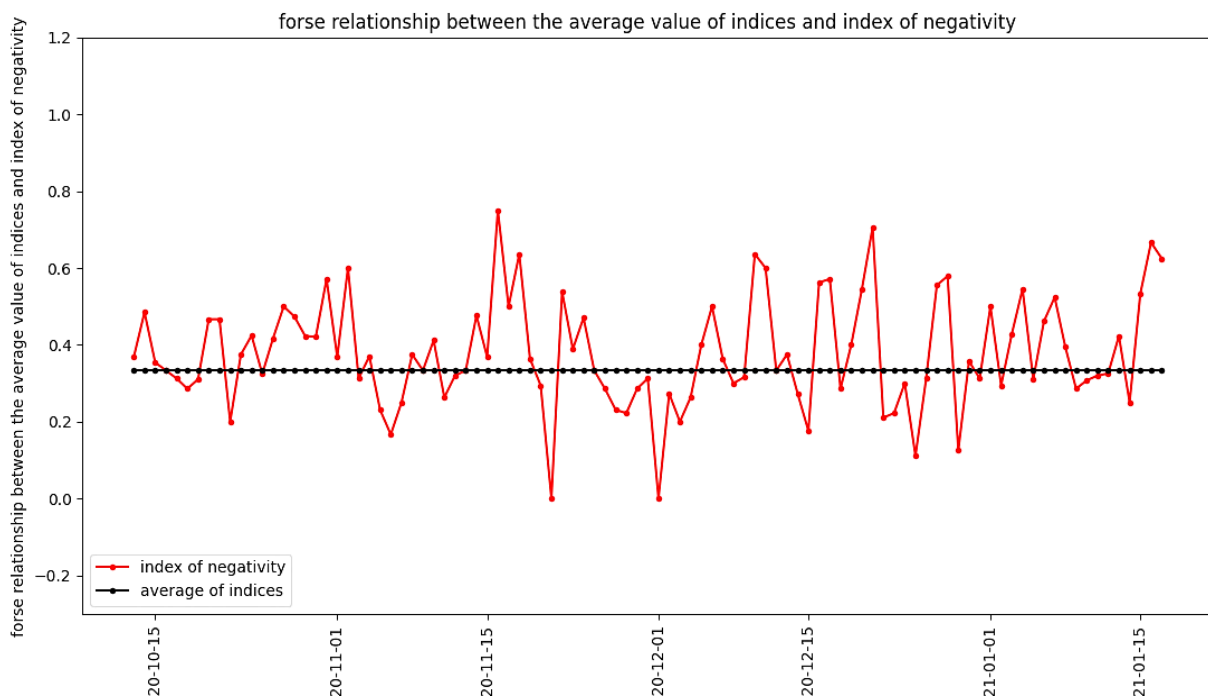


Figure 2. Average of the indices and index of negativity

For the subsequent analyses the neutrality index was discarded, since almost always a zero, taking into account only the positivity index, since the negativity index is complementary. Table 6 shows some statistics regarding the described indices.

Table 6. Index, mean and standard deviation for positive and negative index

Index	Mean	Standard Deviation
positive	0.3	0.13560179225400495
negative	0.3	0.13007428660130818.

Figure 2 shows how the negative index differs, in a few days, from the average value. The days where there are high negative sentiment scores are: 31 October, 16 November, 11 December and 20 December 2020 and 16 January 2021. Table 7 shows three main news items for each day where a peak is negative.

Table 7. Main news related to negative peaks

Data	Text	Keyword
2020-10-31	<ul style="list-style-type: none"> Covid: harder for female workers, 470,000 jobs lost economy ANSA; Masks survey: 2 and a half years to an entrepreneur covid hangs; England in lockdown, tightened in Austria and Greece.Johnson: 'if we do not act there will be thousands of deaths a day' covid ANSA. 	Covid, ANSA, Lockdown
2020-11-17	<ul style="list-style-type: none"> Covid 232 inspections of the nas in the rsa, irregularities in 37 ANSA; In 24 hours 120 hospitalizations in intensive care, 731 covid deaths -19 ANSA; Covid: milan trouble after positive pegs also the deputy giacomo Murelli football seriea ANSA. 	Covid, Died
12.11.2020	<ul style="list-style-type: none"> And 'died of covid the Korean director kimki-duk, golden lion in Venice in 2012, had 59 years.ANSA; Covid exceeded in the world 70 million cases ANSA; In 24 hours covid positive 18,727, 761 victims. Positive-swab rate stable at 9.8%. Overswabs carried out ANSA. 	190 thousand Covid, ANSA, SWABS
2020-12-20	<ul style="list-style-type: none"> Covid: a case of the new variant also in the Netherlands, the Netherlands suspends flights from the united kingdom; Exceptions, second homes, bans: what will the Christmas covid be like ANSA; Covid, England: johnson announces new lockdown in London. "residents of those areas will have to stay at home from Sunday" video. 	ANSA Covid, Natale, ANSA, England, Lockdown, Video
2021-01-16	<ul style="list-style-type: none"> #Covid, Italy turns orange. Red Lombardy announcesAppeal; Turin Localremains open, the holder will be sanctioned # COVID-19 VIDEO #ANSA. #Covid, 718 positive inmates, surge in prisons in Milan. In one week the cases have grown by more than three times, while in all of Italy by 25%. 	Covid, COVID-19

Figure 3, on the other hand, relates the average of the indices, which has a value of 0.3, with the positivity index.

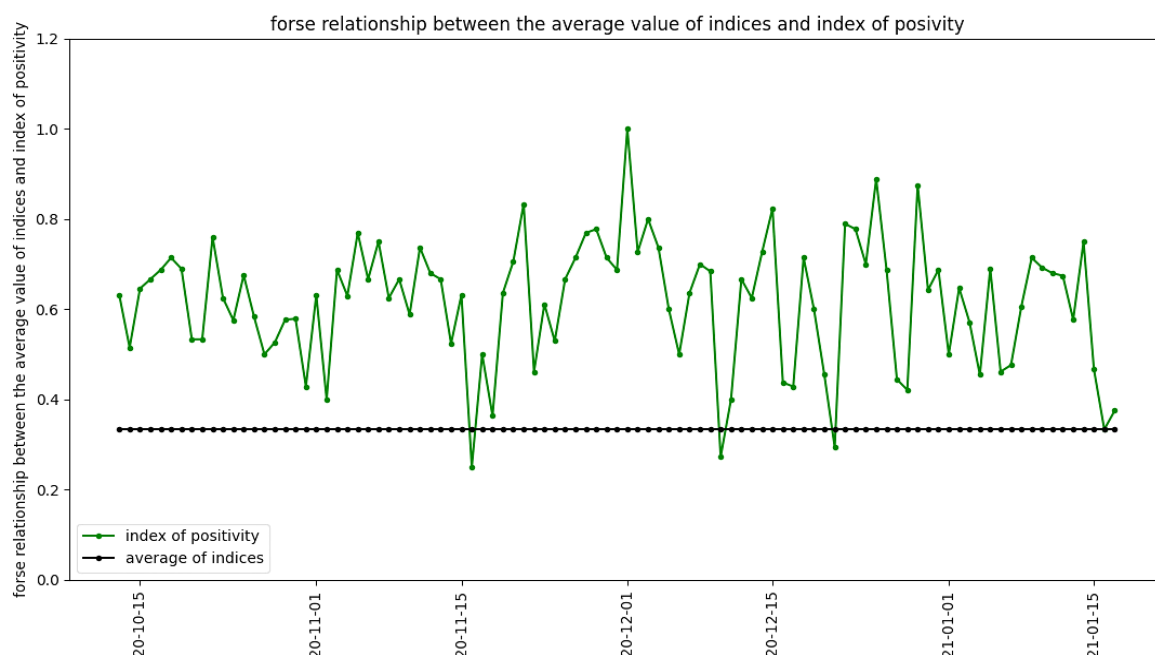


Figure 3. Average indices and positivity index

Table 9 shows for all the news extracted, therefore, the 10 words that are most frequent. Where positive and negative peaks were identified in the previous graphs concerning the index of positivity and negativity in relation to the mean, the words that were most frequent for the respective indices were identified. Figure 5 shows the words found to be most frequent in the dates identified as most significant for the negativity index.

Table 9. Frequency of the most common words

Words	Frequency
ANSA	1667
covid	1354
coronavirus	338
cases	199
dpcm	137
hours	154
infections	130
video	147
vaccine	150



Figure 5. Word Cloud of the most frequent words for negative spikes.

The variant name refers to the latest reports according to which in England and the Netherlands it was discovered that people fell ill with a variant of COVID-19, with following further restrictions and controls. Table 10 indicates the frequencies of words that appear multiple times within negative peaks. The most frequent words on the days in which positive peaks have been identified are shown in Figure 6.

Table 10. Frequency of the most common words for negative peaks

Words	Frequency
ANSA	64
covid	50
variant	16
christmas	10
coronavirus	8
COVID-19	8
lockdown	6
great britain	5



Figure 6. Word Cloud of the most frequent words for positive.

Peaks: In the positive peaks, the most frequent words identified are shown in Table 11.

Table 11. Frequency of the most common words for positive peaks

Words	Frequency
covid	75
ANSA	62
vaccine	12
vaccines	11
coronavirus	8
health ministry	6
robersperanza	6
swabs	6

3.3. Comparison with data from the Civil Protection

The numbers of cases, admissions to intensive care and deaths report the overall numbers of the pandemic. Subtraction operations were therefore carried out, relative to the day analyzed; in fact, the number of the previous day was subtracted from the daily number in order to obtain the number of daily cases. The data were normalized, where they are resized following a fixed interval in order to be able to compare them with the positivity index. Figure 7 relates the positivity index to the number of daily cases.

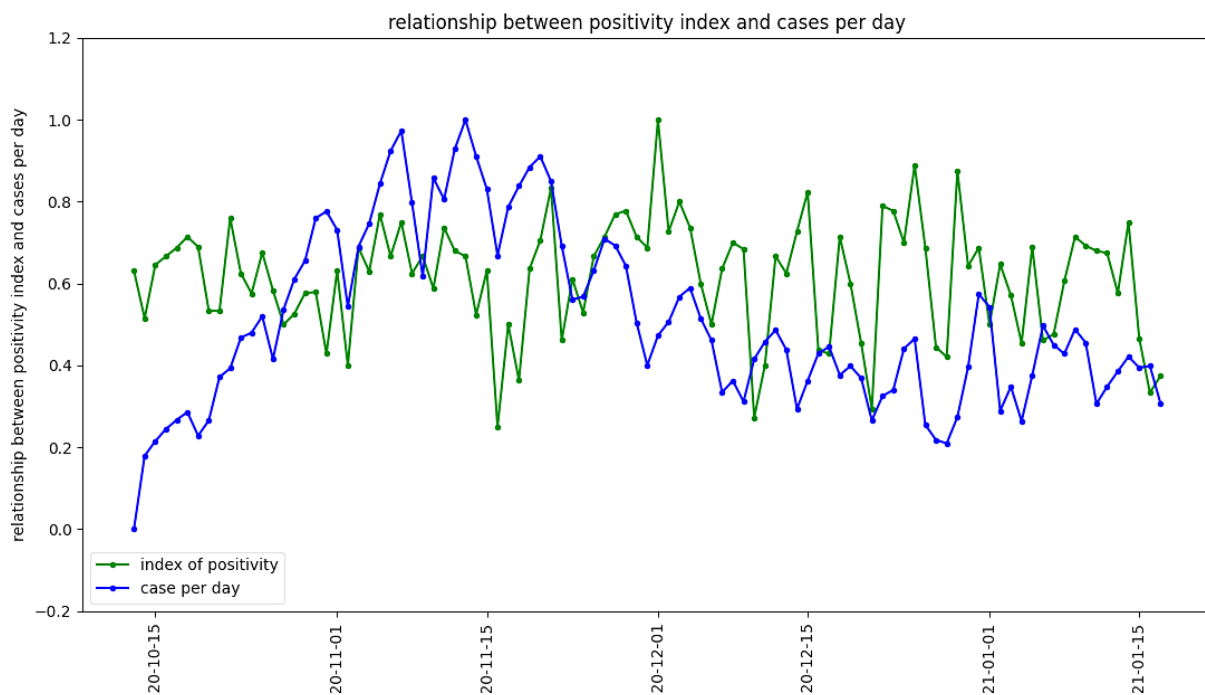


Figure 7. Positivity index and number of daily cases.

The graph shows the positivity index and the number of daily Coronavirus cases. The positivity index remains more or less constant, except for a few days where it undergoes a peak that is determined by some positive news, while the number of cases first grows more or less constant and then decreases. When the number of cases begins to drop, the positivity index goes up. In order to evaluate a possible correlation between the number of daily cases and the positivity index, the Pearson coefficient was calculated. The result of this index consists of two values: the first is the actual coefficient, which, if between 0.5 and 1, implies that the type of correlation that exists is strong. If instead, the coefficient is between -0.5 and -1, there is a strong inverse correlation, otherwise if it is between 0 and 0.5 or 0 and -0.5 the correlation is weak. A coefficient equal to zero implies no kind of relationship between the values. The second result, on the other hand, is called p value and determines when the values can be taken into account; this p value must be less than 0.005. The Pearson index revealed that there is no correlation between the two variables, as the p value exceeds the correlation threshold which is 0.005. Figure 8, on the other hand, relates the positivity index to the number of daily deaths.

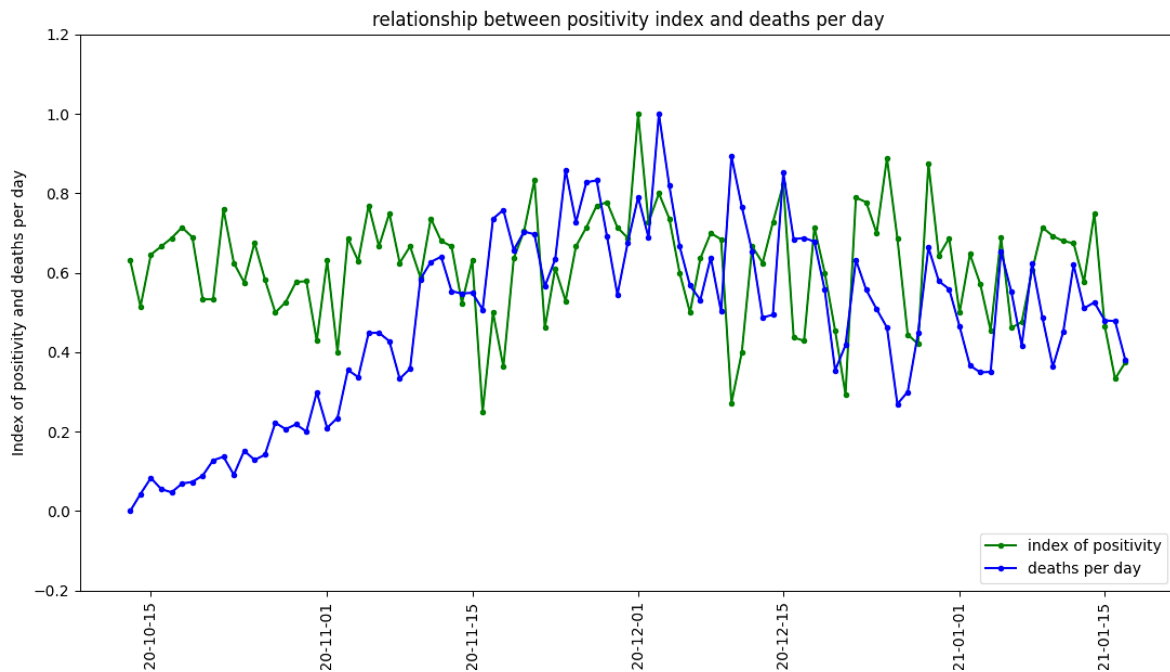


Figure 8. Positive index and daily deaths

The graph shows how both indices have an almost constant trend; however, they are not related since, also in this case, the p value of the Pearson index is greater than 0.005. Figure 9 takes into account the index of positivity with the entry into intensive care.

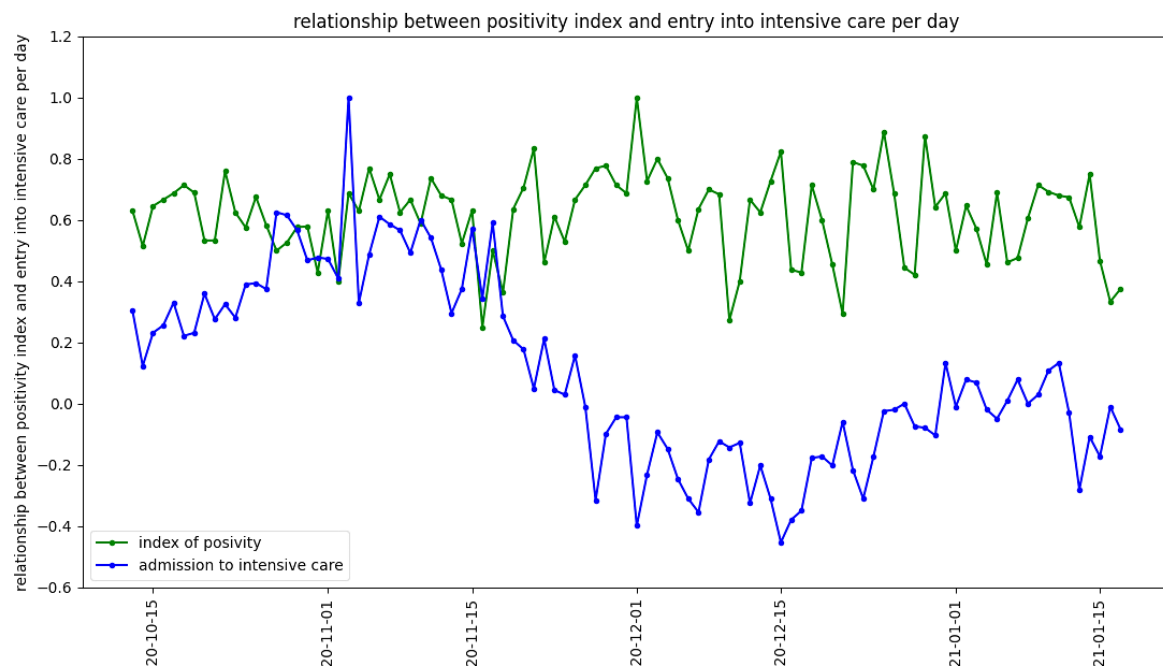


Figure 9. Index of positivity and admissions to daily intensive care

The graph has as values the index of positivity and the number of daily accesses to intensive care. Therapy numbers have been normalized in order to be able to compare them with the index. The Pearson index was also calculated for the graph in question, which is again greater than 0.005. This implies that the two values have no correlation. In addition to the criteria identified above, another parameter was analyzed: the dates on which the Decrees of the President of the Council of Ministers (DPCM) were signed, shown in Figure 10.

The graph relates the positivity index with the days in which the DPCMs were signed announcing different measures such as the closure of shops and bars or the measures implemented during the Christmas period. The index of positive sentiment is not correlated to the precise days of the release of the decrees. In the following Figure 11, on the other hand, the average number of likes and the number of daily deaths are related.

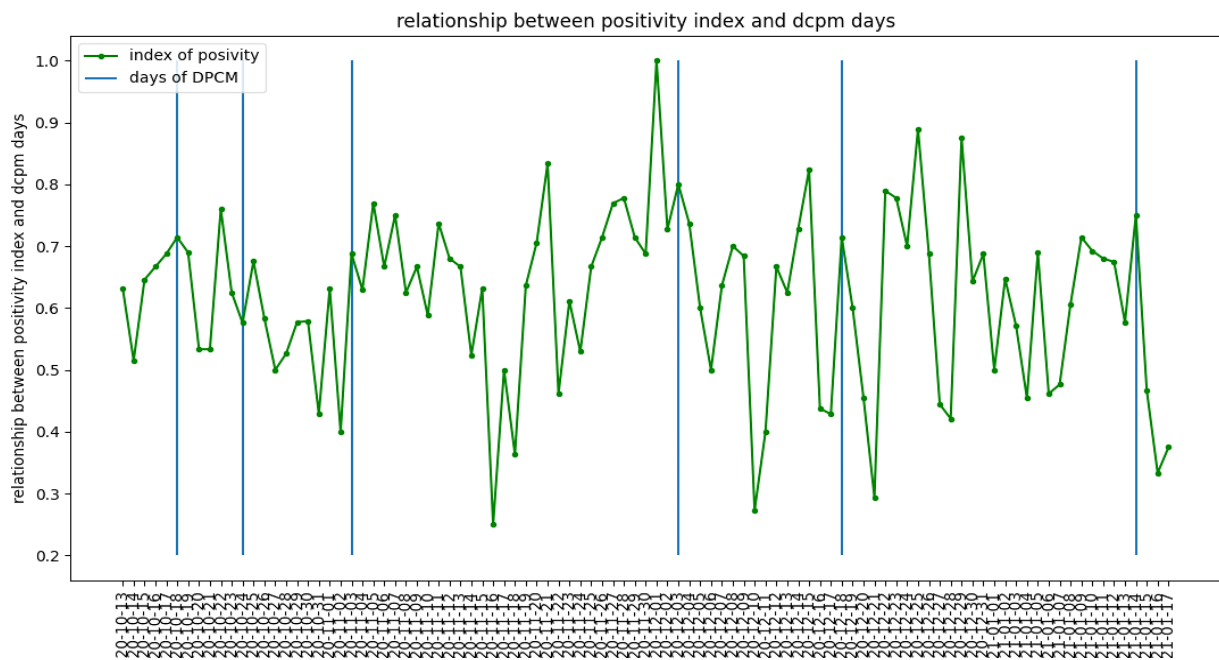


Figure 10. Positive index and days in which the DPCM were signed.

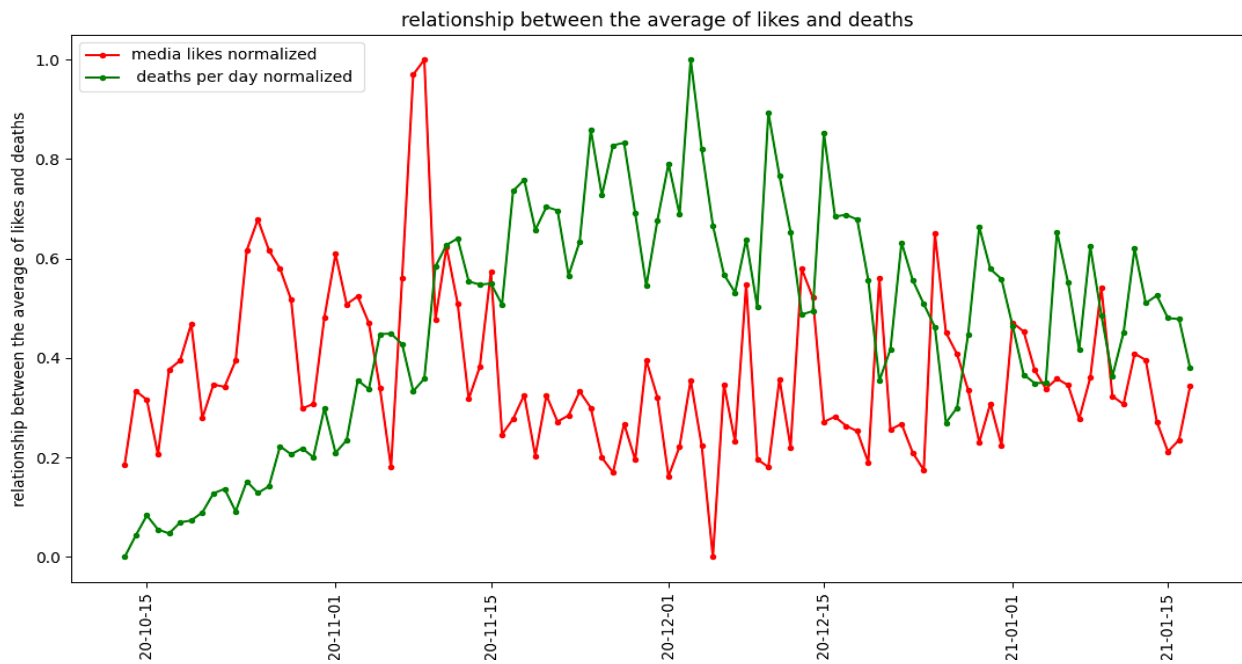


Figure 11. Average likes and number of deaths

The graph shows how both the normalized number of deaths and the average likes have constant values, with the exception of the average likes for November 8, 2020. The calculation made on the data previous has identified a Pearson index equal to -0.35, while the p-value, which determines the correlation of the two values, was 0.002. In fact, as deaths increase, the number of likes per news decreases, even if the value of the Pearson index denotes a weak correlation between the two variables. Table 12 summarizes the described comparisons.

Table 12. Summary of the values of the comparisons made: Pearson's coefficient, p-value and the result

Type of comparison	Pearson's coefficient	p-value	Result
Positive index-number of cases	0.035	0.75 (greater than 0.005)	not correlated
Positive index-number of deaths	0.12	0.26 (greater than 0.005)	not correlated
Positive index-entry into intensive care	-0.13	0.24 (greater than 0.005)	not correlated
Index of positivity-dpcm	0.14	0.29 (greater than 0.005)	not correlated
Mean like-deaths	-0.35	0.002	correlated (weakly)

4. Conclusion

Due to social distancing policies as a consequence of the Covid-19 pandemic, a lot of people rely on social platforms for news consulting. Therefore, it is crucial to identify the trend in news sentiment over time (Karishma Sharma et al., 2020). The results of the research showed that the news extracted through the Tweepy software has a greater number of positive news than negative ones. In fact, the positive news is 1101 (62.3%) while the negative news is 659 (37.3%). Neutral news, on the other hand, occupies only 0.3% of the entire dataset with only 6 news items.

The calculation of the indices made it possible to consider them as general indicators of the positive, negative, and neutral sentiment offered by the ANSA Agency and showed that the neutrality index is practically zero, as the values are almost always close to zero, contrary to the other two indices. The indices obtained from the sentiment analysis were compared with the data chosen by the Civil Protection website. The results of the research then established how the elements extracted from the Civil Protection website (number of daily cases, number of deaths and admissions to intensive care) do not influence the news regarding the ANSA Agency pandemic, which, in fact, is more positive despite the argument. On the other hand, the average day-to-day likes with the number of deaths per day were inversely correlated, although with a weak link. It would be interesting to expand the extraction of data to different Twitter profiles of news agencies or newspapers, to see how the topic of Coronavirus has been treated and if the sentiment of the news is different. Furthermore, the sentiment could be correlated with the factors obtained from the Civil Protection website: number of cases, intensive care admissions, and number of deaths to verify that other agencies have been influenced by these factors. Moreover, through a Name Entity Recognizer (NER) study, the relationship between the peaks of the positive and negative indices and their keywords taken from the extracted news could be deepened.

5. Declarations

5.1. Author Contributions

Conceptualization, G.P., A.L.D. and A.M.; methodology, A.L.D.; software, G.P.; validation, G.P., A.L.D. and A.M.; formal analysis, A.L.D.; investigation, G.P.; resources, G.P.; data curation, G.P.; writing—original draft preparation, G.P. and A.L.D.; writing—review and editing, G.P., A.L.D. and A.M.; visualization, G.P.; supervision, A.M.; project administration, A.M. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

Data released by the Italian Civil Protection are available in a publicly accessible repository: The data presented in this study are openly available in <https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv>.

Data extracted from Twitter are available on request due to restrictions privacy. The data presented in this study are available on request from the corresponding author.

5.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6. References

- [1] Pokharel, B. P. (2020). Twitter Sentiment Analysis during Covid-19 Outbreak in Nepal. SSRN Electronic Journal. doi:10.2139/ssrn.3624719.
- [2] Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends (Preprint). doi:10.2196/preprints.19447.
- [3] Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets. IEEE Access, 8, 181074–181090. doi:10.1109/access.2020.3027350.

- [4] Rajput, N. K., Grover, B. A., & Rathi, V. K. (2020). Word frequency and sentiment analysis of twitter messages during coronavirus pandemic. arXiv preprint arXiv:2004.03925.
- [5] Pes G., Lo Duca A., Marchetti A., (2020), Tweepy: a software to extract tweets related to a single Twitter user. "CNR", 1-11.
- [6] Wiberg, M. (2005). *The Interaction Society: Practice, Theories and Supportive Technologies*. IGI Global Publishing, Pennsylvania, United States
- [7] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data. *Proceedings of the Workshop on Languages in Social Media, LSM'11*, Oregon, United States.
- [8] Balahur, A., & Steinberger, R. (2009). Rethinking Sentiment Analysis in the News: from Theory to Practice and back. *Proceeding of WOMSA*, 1-12.
- [9] Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *ICWSM*, 7(21), 219-222, Colorado, United States.
- [10] Chakraborty, A., & Bose, S. (2020). Around the world in 60 days: an exploratory study of impact of COVID-19 on online global news sentiment. *Journal of Computational Social Science*, 3(2), 367–400. doi:10.1007/s42001-020-00088-3.
- [11] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report*, 1(12), Stanford, California, United States.
- [12] Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., ... & Belyaeva, J. (2013). Sentiment analysis in the news. arXiv preprint arXiv:1309.6202.
- [13] Chen, H., Zhu, Z., Qi, F., Ye, Y., Liu, Z., Sun, M., & Jin, J. (2021). Country Image in COVID-19 Pandemic: A Case Study of China. *IEEE Transactions on Big Data*, 7(1), 81–92. doi:10.1109/tbdata.2020.3023459.
- [14] Kruspe, A., Häberle, M., Kuhn, I., & Zhu, X. X. (2020). Cross-language sentiment analysis of european twitter messages during the covid-19 pandemic. arXiv preprint arXiv:2008.12172.
- [15] Pran, M. S. A., Bhuiyan, M. R., Hossain, S. A., & Abujar, S. (2020, July). Analysis of Bangladeshi People's Emotion during Covid-19 In Social Media Using Deep Learning. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1-6. IEEE. doi:10.1109/ICCCNT49239.2020.9225500.
- [16] Kim, B. (2020). Effects of social grooming on incivility in COVID-19. In *Cyberpsychology, Behavior, and Social Networking* 23(8), 519-525. doi:10.1089/cyber.2020.0201.
- [17] Ho, Y. F. W., Hiu, Y. S. L., Ambrose, H. T. F., Siu, T. L., Thomas, W. Y. C., Christine, S. Y. L., ... & Ming-Yen, N. (2020). Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology*, 201160.
- [18] Padmaja, S., S, P., & Bandu, S. (2014). Evaluating Sentiment Analysis Methods and Identifying Scope of Negation in Newspaper Articles. *International Journal of Advanced Research in Artificial Intelligence*, 3(11), 1–58. doi:10.14569/ijarai.2014.031101.
- [19] Nurlaila, I., Rahutomo, R., Purwandari, K., & Pardamean, B. (2020). Provoking Tweets by Indonesia Media Twitter in the Initial Month of Coronavirus Disease Hit. *2020 International Conference on Information Management and Technology (ICIMTech)*. doi:10.1109/icimtech50083.2020.9211179.
- [20] Chun, S. A., Li, A. C.-Y., Toliyat, A., & Geller, J. (2020). Tracking Citizen's Concerns during COVID-19 Pandemic. *The 21st Annual International Conference on Digital Government Research*. Digitranscope Spring Institute, Seoul National University, Seoul, Korea. doi:10.1145/3396956.3397000.
- [21] Marco, V., Giuliano, G., Basile, V., & Bosco, C. (2019). The Tenuousness of lemmatization in lexicon-based sentiment analysis. In *Sixth Italian Conference on Computational Linguistics*, 2481, 1-6.
- [22] Ferro, M., Pezzulo, G., & Pirrelli, V. (2010). Morphology, memory and the mental lexicon. *Lingue e Linguaggio*, 9(2), 199-0.
- [23] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, Association for Computational Linguistics, University of Pennsylvania, Philadelphia, United States.