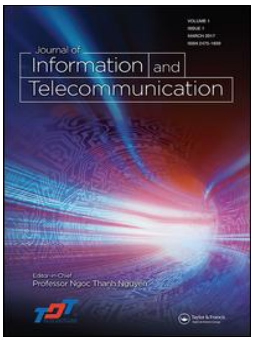



Taylor & Francis
Taylor & Francis Group

Journal of Information and Telecommunication

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjit20

# Social media sentiment analysis based on COVID-19

László Nemes & Attila Kiss

JOURNAL OF INFORMATION AND TELECOMMUNICATION
https://doi.org/10.1080/24751839.2020.1790793



# Social media sentiment analysis based on COVID-19

László Nemes and Attila Kiss

Department of Information Systems, ELTE Eötvös Loránd University, Budapest, Hungary

**ABSTRACT**
In today's world, the social media is everywhere, and everybody come in contact with it every day. With social media datas, we are able to do a lot of analysis and statistics nowdays. Within this scope of article, we conclude and analyse the sentiments and manifestations (comments, hastags, posts, tweets) of the users of the Twitter social media platform, based on the main trends (by keyword, which is mostly the 'covid' and coronavirus theme in this article) with Natural Language Processing and with Sentiment Classification using Recurrent Neural Network. Where we analyse, compile, visualize statistics, and summarize for further processing. The trained model works much more accurately, with a smaller margin of error, in determining emotional polarity in today's 'modern' often with ambiguous tweets. Especially with RNN. We use this fresh scraped data collections (by the keyword's theme) with our RNN model what we have created and trained to determine what emotional manifestations occurred on a given topic in a given time interval.

**ARTICLE HISTORY**
Received 20 May 2020
Accepted 30 June 2020

**KEYWORDS**
natural language processing; recurrent neural network; sentiment analysis; social media; visualization

## 1. Introduction

The main goal is to train a model to sentiment prediction by looking correlations between words and tag it to positive or negative sentiment.

In today's world, social media platforms like twitter are of immense importance to people's everyday lives. We definitely have to deal with the manifestations on these platforms, and as machine learning becomes more and more popular and important just like the natural language processing (NLP), we have to deal with this, and analyse and research the emotions on this platforms.

There are many ways to approach a topic, from 'pure' dictionary-based analysis to 'more serious' deep learning, neural networks. By building learning algorithms and classifiers, we strive to label the relevant tweets with the appropriate emotional polarity.

As we mentioned at the beginning of the introduction, the main objective of this article is to develop a model for predicting emotions by focusing on the relationship between words, thus labelling specific entries, as opposed to the usual 'positive' and 'negative' decomposition, we get a much wider scale for more accurate forecasting. However, at the focus point, there is no larger dataset, but the properly trained model analyses with

**CONTACT** Attila Kiss kiss@inf.elte.hu

a newly mined dataset that matches the current trend (coronavirus themes now) and dataset build number, (number of scraped tweets) which narrowing the circle of a larger amount of data into a narrower topic. In this way, we do not only indicate that the data should be positive or negative, we also provide a more detailed breakdown of the emotional levels. This can provide more accurate data than analysing larger datasets, as fresh mining is always available, so you can get much faster and more accurate results as final result than earlier larger samples and other polls.

We also compare our model with other third-party options to see how small details play a very important in proper categorization using a properly taught a Recurrent Neural Network model for different messages. Thus, as mentioned earlier, focusing on specific topics, by analysing a given number of messages (tweets), and waiting the particular emotional outcomes related to the topic. According to our estimates, we expect a more accurate and detailed analysis and categorization of an emotional analysis related to a current topic, which can provide a more stable and accurate basis for various sociological and other studies. It also provides a different approach to research on the pandemic, focusing on the rapidly changing human mood and opinion. Such as the changes and manifestations of human moods in a given period of the coronavirus on social media.(Twitter)

The model was built and taught using the libraries and capabilities provided by tensorflow. By analysing a Recurrent Neural Network (RNN). The rest of this article contains sections on the structure and use of the encoder, model and results.

## 2. Related works

Emotional analysis of twitter datasets within the article of Balahur (2013) using unigram and bigram (n-gram) and supervised learning with simple Support Vector Machines. Based on the results we can conclude that on the one hand, the best properties to use emotional analysis is the unigram and the bigram together. Second, we can see that generalizations, using unique tags, emotive words and modifiers are strongly improve the performance rating of emotions. (joy, happy, sadness, fear, etc.) Presented in another article, Jianqiang and Xiaolin (2018) introduces a word embedding method implemented, based on unsupervised learning and large twitter corpora, the method uses hidden contextual semantic relationships and co-occurrence statistics between tweets and words. These word embeds are combined with n-gram characteristics and word mood polarity score characteristics form a set of tweet emotional features. Set is integrated into a deep convolutional neural network.

The method, which described by Ortis et al. (2018) uses text extracted from the description of different images instead of classic user entries. Then defines a multimodal embedding space based on the text properties. The emotional examination being performed by a supervised Support Vector Machine.

This study explores techniques of Leskovec (2011) for modelling, analysing, and optimizing social media. First, they show us how to collect large amounts of social media data. Then it will continue to discuss methods for obtaining and tracking information and how to build forecasting models for information dissemination and inclusion. Finally, they discusses methods for monitoring the flow of emotions across the network and the development of polarization.

With the Recurrent Neural Network by Mikolov et al. (2010), which is intentionally run multiple times and the goal with statistical language modelling is to predict the next word in textual data in its given context. Where the experiments show significant reduction of word error rate. In addition, Mikolov et al. (2011) shows that the recurrent neural network language model (RNN LM) significantly outperforms many competitive language modelling techniques. And approaches that result in more than 15-fold acceleration in both the training and testing phases are presented. Finally, they discuss options for reducing the parameters of the models. The resulting RNN model is thus smaller, faster in both training and testing, and may be more accurate than the base. Besides in another article, we can cover up the SummaRuNNer (Nallapati et al., 2017) which is a Recurrent Neural Network (RNN) based sequence model, and interpretable neural sequence model which is proposed to summarize extraction documents. Which shows that, it is better performing than or is comparable to the state-of-the-art deep learning models.

Following this, we were introduced to learning several related tasks together using a multitasking learning framework by Liu et al. (2016). Based on the recurrent neural network, three different mechanisms are proposed sharing information to model text with task-specific and shared layers. Textual classification tasks shows that, the proposed models can improve the task using other related tasks.

In another work, Arras et al. (2017) presented a simple and effective strategy for extending the Layer-wise Relevance Propagation (LRP) process to repetitive architectures such as LSTMs, by proposing a rule for reproducing relevance through multiplicative interactions. The extended LRP version was applied bidirectionally. The LSTM model shows the emotional prediction of sentences to see if the relevance of the resulting words is reliable and what the classifier's decision for or against a particular class is and how they perform better than gradient-based decomposition.

Getting to know a different perspective, we can discover the SmartSA, a lexicon-based sentiment classification system for social media genres by Muhammad et al. (2016), which integrates contextual grasp strategies in two different ways: interaction of terms with their local context and global context. They also present a hybridization method for a general purpose lexicon, SentiWordNet, with genre-specific vocabulary.

Besides, we can focus to describes an emotional analysis study by Neri et al. (2012), which includes more than 1000 Facebook posts based on news summaries of Rai – the Italian public broadcaster service versus the emerging and more dynamic La7 private company. This study maps study results with observations made by the Osservatorio di Pavia, an Italian research institute specializing in theoretical media analysis.

Along with the growth of web content, there is an increasing number of hate speech on various platforms, which provide a suitable filtering tool for natural language processing by Schmidt and Wiegand (2017). It is shown that character-level approaches work better than token-level approaches, and that a lexical list of resources, such a list of slurs, can help rank, but usually only in combination with others.

Additionally, we can also get to introduce a new metaheuristic method (CSK) by Pandey et al. (2017), based on K-means and cuckoo search. The method provides a new way to find optimal cluster heads based on the sentimental content of the Twitter dataset.

Wang and Li (2015) extends significant advances in text-based emotional prediction tasks to a higher level of prediction of emotion behind images. They show that visual and textual features alone are not sufficient for accurate emotional tagging. Experiments with two large datasets show that the proposed method significantly improved the existing state-of-the-art methods.

Finally, Xu et al. (2019) propose a new Hierarchical Deep Fusion (HDF) model for exploring the transverse relationship between images, text, and their social relationships, which, with their complementary features, make emotional analysis more effective. Visual content is combined with various semantic fragments of textual content using three-level hierarchical LSTM (H-LSTM) to learn the inter-modal correlation of image and text at different levels.

## 3. DataSet/DataFrame building for the analysis

### *3.1. Existing dataset usage*

Of course, we also have the option to use data from external sources that was previously built from tweets for specific topics (possibly a huge mixed tweet collection or some more specific collection), but in this case, we have to keep in mind that, these data may not be up to date. So it can also be a previously compiled collection and there are several sources where you can access and download datasets.

Basically, this would not be a problem, but under the circumstances, we try to rely on the most up-to-date datas for test dataset. However, it may be suitable for comparison to what extent the writing trend of a given circle influences the outcome of the analysis.

### *3.2. Build dataSet using Twitter API for scraping*

Using the twitter developer tools, we build a test dataset using a scraping script, which compiles our data collection from tweets into a topic based dataset with the given keyword and a tweet scrape data number. In the state before use for analysis, we have the possibility to submit this data for a completely different non-RNN-based testing, as the dataset construction also supports the performance of a completely different, traditional analysis. For example, Excel-based processing (not deep learning).

About the methods that perform the scraping and cleaning, our main method is the 'datasetbuilding' where according to the parameters we need a keyword for the current scraping, a tweet count limit (how many tweet do we need in this theme) date intervals, which time period where we would like to extract data in this related topic, and of course the language, where we used English in all cases. For the scrape, we have also used the tweetpy library for the Twitter API. Plus we perform the 'extra' cleaning with the 'cleantweet' method (Listing 1).

However, we would like to use the Recurrent Neural Network what we have built, and we also would like to use the test dataset (which we are freshly scraping and mining.) on our already trained model. The scraping script what mentioned above makes this possible, because the dataset has undergone proper formatting and cleaning.

Overall, after compiling the dataset itself, we have the opportunity to use this data in a completely different traditional (Excel) analysis as well. But, these systems and structures are supported by the script in an orderly, uninterrupted manner and also run the analysis. The analysis will focus primarily on a separate specific topic, which will be the Coronavirus. On this Figure 1, we can see there are a lot of another possibility and method to analyse with this dataset.

```
def dataset_building(self, tag, limit, begin_date, end_date, lang):
    with open('result.csv', mode='wt', encoding='UTF-8', newline='') as file:
        w = csv.writer(file)
        w.writerow(['Time', 'UserName', 'Tweet_text', 'All_Hashtags', 'Followers_count'])

        for tweet in tweepy.Cursor(self.api.search, q=tag + ' -filter:retweets', lang=lang,
        tweet_mode='extended', since=begin_date, until=end_date).items(limit):
            w.writerow([tweet.created_at,
                        tweet.user.screen_name,
                        self.clean_tweet(tweet.full_text),
                        [e['text'] for e in tweet._json['entities']['hashtags']],
                        tweet.user.followers_count])

def clean_tweet(self, text):
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t]) | (\w +:\ / \ / \S +)", " ", text).split())

```

**Listing 1.** Part of the Twitter dataset builder

| Time | UserName | Tweet_text | All_Hashtags | Followers_count |
|---|---|---|---|---|
| 2020.04.14 23:59 | SCOTTYSIMM | White House Attacks Voice Of America Over China Coronavirus Coverage https://t.co/qEzZIgC0WW | [] | 2045 |
| 2020.04.14 23:59 | unlewis | Trump says he doesn't think Michigan Dem will vote for 'Sleepy Joe after surviving coronavirus Fox News h | [] | 3359 |
| 2020.04.14 23:59 | sdakshin | Word of the year https://t.co/mi0pzuITyb via | [] | 62 |
| 2020.04.14 23:59 | PeepsRide | Elderly people who have suspended in-home services during the coronavirus crisis will soon receive welfa | [] | 66 |
| 2020.04.14 23:59 | SeattleMet | Dig around a little You probably already have something mask-worthy at home https://t.co/NOpObfEwIs | [] | 191429 |
| 2020.04.14 23:59 | rep4bettergovt | Coronavirus Live Updates Trump Halts U.S Funding of World Health Organization Perhaps funding of the Re | [] | 186 |
| 2020.04.14 23:59 | Milesandmiles30 | Why have so many people in this thread stopped reading after coronavirus You know that's not the end of | [] | 34 |
| 2020.04.14 23:59 | muhammad_fall | Haunting photos of empty airports and planes at the height of the #COVID-19 pandemic show the airline in | ['COVID'] | 208 |
| 2020.04.14 23:59 | AllOnMedicare | US for-profit healthcare sector cuts thousands of jobs as pandemic rages https://t.co/eGL63ujnlR | [] | 13825 |
| 2020.04.14 23:59 | Jsurfer0730 | Trump campaign still bragging about 'lowest unemployment rate in years https://t.co/CHBHDBfD3A | [] | 5293 |
| 2020.04.14 23:59 | kbaseballump | One Arrested as Raleigh NC Police Suspend First Amendment Declare Coronavirus Lockdown Protest "Non- | [] | 2022 |
| 2020.04.14 23:59 | brisuelou | He thinks that the WHO mismanaged this? Are you fucking kidding me? https://t.co/IzU2YG6ssx | [] | 55 |
| 2020.04.14 23:59 | Verda_777 | _camera _defender Here is the chronological order provided by the Hudson Institute China and WHO are di | [] | 453 |
| 2020.04.14 23:59 | oracleofliberty | wow out of control for federal funds. #CoronavirusPandemic #coronavirus #newyorkcoronavirus #AndrewC | ['CoronavirusPandemic', 'coronavirus', 'newyorkcoronavirus', 'AndrewCuomo'] | 1395 |
| 2020.04.14 23:59 | hocott_david | . I urge you to #cancelstudentdebt in the next #coronavirus package A #StudentDebtStimulus will help the | ['cancelstudentdebt', 'coronavirus', 'StudentDebtStimulus'] | 33 |
| 2020.04.14 23:59 | aprilhenrybooks | The Art (and Awkwardness of a Virtual Haircut https://t.co/WIQn3MHN0b | [] | 4054 |
| 2020.04.14 23:59 | WarWraith | This implies there will be an "after" If there's no effective vaccine and this coronavirus remains with us the | [] | 2712 |
| 2020.04.14 23:59 | AdvocaatBroeke | U.S will halt funding to World Health Organization while it investigates groupâ€™s response to Coronaviru | ['Trump', 'China'] | 1068 |
| 2020.04.14 23:59 | FulviaCalcagni | [... there is little reason to expect the coronavirus crisis to accelerate climate-friendly decoupling UNLESS p | [] | 59 |
| 2020.04.14 23:59 | nisar__adil | The study found that the genetic sequence of a coronavirus discovered in lung samples of pangolins was hi | [] | 417 |
| 2020.04.14 23:59 | manatelugumovie | Corona In India 644 In Telangana 52 Just From Today! The number of Coronavirus in https://t.co/acLVNgeAI | [] | 4351 |
| 2020.04.14 23:59 | ChiTownLionPSU | Data from https://t.co/MywM5iAcVh wrt Coronavirus/Covid19. Comparing the United States to some of ou | [] | 3246 |
| 2020.04.14 23:59 | MickSPrice | #WhereIsNana Crime Family Pelosi extorting taxpayer funds for decades I see her dad was a corrupt politic | ['WhereIsNana'] | 523 |
| 2020.04.14 23:59 | jjbigcity | TRUMP KNEW ABOUT CHINA'S OUTBREAKS OF CORONAVIRUS IN NOVEMBER 2019 IN FACT OF THIS PRESIDEI | [] | 6 |
| 2020.04.14 23:59 | RonRonkmfa | There are those among us who are chomping at the bit to congregate and mingle without regard for the eff | ['TheInvincibles', 'COVID19', 'coronavirus', 'StayHome', 'SocialDistancing', 'Covid_19', 'COVID' | 12 |
| 2020.04.14 23:59 | RichinWriterss | Trump Coronavirus Task Force hold press briefing at White House 4/14/20 https://t.co/dwvPGr9DaK | [] | 1929 |
| 2020.04.14 23:59 | diimsa | The Word Challenge High School (7:00 PM 7:05 PM CST). #DIIMSAVBOARD #DIIMSACHALLENGE #DIIMSA #Cli | ['DIIMSAVBOARD', 'DIIMSACHALLENGE', 'DIIMSA', 'ClimateChange', 'COVID19', 'science', 'rem | 565 |
| 2020.04.14 23:59 | Reuters | As virus tears through reservation Navajos give lifeline to elders and families https://t.co/g72TcleIZE https: | [] | 21747669 |
| 2020.04.14 23:59 | steviekim222 | What Does Coronavirus Mean For Brands on Social Media /feed/what-does-coronavirus-mean-for-brands- | ['ogilvy'] | 4218 |
| 2020.04.14 23:59 | BagalueSunab | https://t.co/tfezR73pbt Coronavirus mutation could threaten the race to develop vaccine A strain found in | ['India'] | 2615 |
| 2020.04.14 23:59 | mharisman | (4) Dog of a Covid-19 patient in Hong Kong has tested â€śweak-positive for Covid19 but officials say there i | [] | 224 |
| 2020.04.14 23:59 | Ronnie_Zen | Together we can stop the spread of Coronavirus Disease (COVID-19) #AyoBersamaLawanKorona #DiRumah | ['AyoBersamaLawanKorona', 'DiRumahAja', 'JagaJarak'] | 295 |
| 2020.04.14 23:59 | The_Real_Fly | Boeing Co on Tuesday reported another 75 cancellations for its 737 MAX jetliner in March as the coronaviru | [] | 47858 |
| 2020.04.14 23:59 | DonDoncarey4 | Student files class-action lawsuit against Liberty University over coronavirus response TheHill https://t.co/ | [] | 82 |
| 2020.04.14 23:59 | ParhamMansor | Chinaâ€™s â€Donation Diplomacy Raises Tensions With U.S https://t.co/7rgzDduAQr | [] | 1342 |
| 2020.04.14 23:59 | khw_lker | Trying not to be angry about this coronavirus is hard Talking listening and even seeing the effects of it is tir | [] | 559 |
| 2020.04.14 23:59 | HausSante | encouraging people to get Coronavirus In the name of Virtue https://t.co/vqI8uIkhVf | [] | 117 |

**Figure 1.** Part from the fresh mined DataSet.

## 4. Different ways of sentiment analysis

As we mentioned there are several different possibility to the Natural Language processing and Sentiment analysis. If we would like to separate that into two categories, first, the classic dictionary style, which is not the most modern way as opposed to the Deep Learning possibilities. What we also use in this analysis, is the Recurrent Neural Network.

In classical dictionary-based analysis, we have a pre-set vocabulary where each word has a value, whether the effect of the word is positive or rather negative. Accordingly, the sentences are decomposed so that each word is identified, and then, according to our dictionary, we assign the given value to the effect what that word also has. The sum of these values would give the emotional value of our particular sentence in the most general case. Of course, we can run into a lot of problems here, as denials, double denials, word turns, word combinations that can affect emotion which cannot be detected. This is why it has shifted this topic towards Deep Learning, using properly trained models.

### 4.1. Deep learning – RNN

We use and build Recurrent Neural Network (RNN).

What is Recurrent Neural Network (RNN)[1] – A neural network that is intentionally run multiple times, where parts of each run feed into the next run. Specifically, hidden layers from the previous run provide part of the input to the same hidden layer in the next run. Recurrent neural networks are particularly useful for evaluating sequences, so that the hidden layers can learn from previous runs of the neural network on earlier parts of the sequence.

For example, one recurrent neural network that runs four times. Notice that the values learned in the hidden layers from the first run become part of the input to the same hidden layers in the second run. Similarly, the values learned in the hidden layer on the second run become part of the input to the same hidden layer in the third run. In this way, the recurrent neural network gradually trains and predicts the meaning of the entire sequence rather than just the meaning of individual.

In addition to the RNN, the advantages are that it is possible to process inputs of any length. The size of the model does not increase with the size of the input. The calculation takes into account historical information. The weights are distributed as a function of time. Of course, it should be noted that some general counter-arguments are mainly that the calculation is slow.

### 4.2. RNN model build and train

The tools provided by Keras and Tensorflow were used to build the model. Where we created a Sequential model by passing a list of layer instances to the constructor and the first layer is the Embedding layer, which can be used for neural networks on text data. It requires that the input data be integer encoded, so that each word is represented by a unique integer. The embedding layer is initialized with random weights and will learn an embedding for all of the words in the training dataset. Then we used Bidirectional wrapper for RNNs. Next is the Dense and Dropout layers. A dense layer is a classic fully connected neural network layer, each input node is connected to each output node. A dropout layer is similar except that when the layer is used, the activations are set to zero for some random nodes. This is a way to prevent overfitting.

Then we also save our trained models in .h5 format with the actual training date, to reuse that, if we need to. Also we have a another possibility to load this trained models and use it on the new scraped datas. (There is a separate menu option to use a re-trained model or a previous model where we use the name of this .h5 file to refer this.)

There are numerous way to use train and test datasets before you use the trained model in a real dataset. Tensorflow gives us numerous datasets for example: 'imdb reviews/subwords8k' and 'civil comemments' etc. We can split it up to train and test dataset and use it for 'compile' and 'fit' model calls, and of course we can use our own datasets for this train and test phase as well. In the case of models trained by external datasets, we can talk about 'continuous learning', since another dataset is made for the model and we use the result for our own actual datasets.

For display, we use the matplotlib.pyplot package, where our model walk through the given dataset and use the predict method. Accordingly, we categorize how positive and negative the emotional value of the tweet or sentence, plus visualize this results with a colourful plot.

### 4.3. RNN analysis – themes and results

To examine and compare the model, the coronavirus topic (which is the most prominent and up-to-date topic of our time, the recent data mining results have a lot of potential, as there is no pre-compiled dataset here and rapid changes can be topical here) and different numbers of fresh datas, what we mine. In addition, comparisons are made with several third-party applications and we also compare with traditional, classical analysis what differences and conclusions can be drawn about efficiency, accuracy, and speed in different cases.

We expect that the model we have trained and developed and taught in detail can provide more accurate results for today's online communication formulas, difficult multi-meaning sentences and unique topics than a traditional or a third-party application that also works with accurate but larger error ranges than our more accurately prepared model.

#### 4.3.1. Compare to the oldfashion research work

Traditional polling or purely human work, tracking, data collection, analysis, these processes are time consuming. The result would be very accurate, but by the time the report is completed, the conclusion may be outdated. The result would no longer be relevant. Thus, in the case of any human labour trigger, especially data mining, scraping can be a huge step forward as a test dataset. In this way, the process takes less time and we are also able to use a number of other third-party tools to speed up our processes.

In essence, we can discover incomparably large differences between traditionally supported analyses and analyses which supported by different scraping and other dataset compilation options, as the difference is found in time and accuracy. In addition to the traditional research process, it should also be mentioned because people do the analysis, so the accuracy of the tweet polarity would be really good, but it cannot cover such a large sample, i.e. not in such a short time, so the results may be much more relevant for analyses despite perhaps a larger error factor, as up-to-date and fast results as well as partial results can be obtained. Not to mention the special well-trained neural networks, the results and speed of which cannot be measured by the speed of human work.

#### 4.3.2. Compare to some third-party sentiment analyser like TextBlob

As mentioned earlier for the analyses, we will use coronavirus theme, which dominates social media platforms.

The pre-measurement expectations are as follows: We would like more accurate, less or even zero neutral expression in the results, as these data would greatly distort the real picture, and basically we would like to minimize the neutral category as much as possible. Based on the small details, the twisted manifestations, we expect some cases move to the negative or positive direction from the neutral space. For both TextBlob and RNN, we apply a same appropriate categorical distribution to different levels of feel.

Thus, we scraped a different number of tweet data in each analysis and compared the results of the test datasets. (Fresh Scraped Tweets dataset what we use on the trained RNN model and TextBlob as well.)

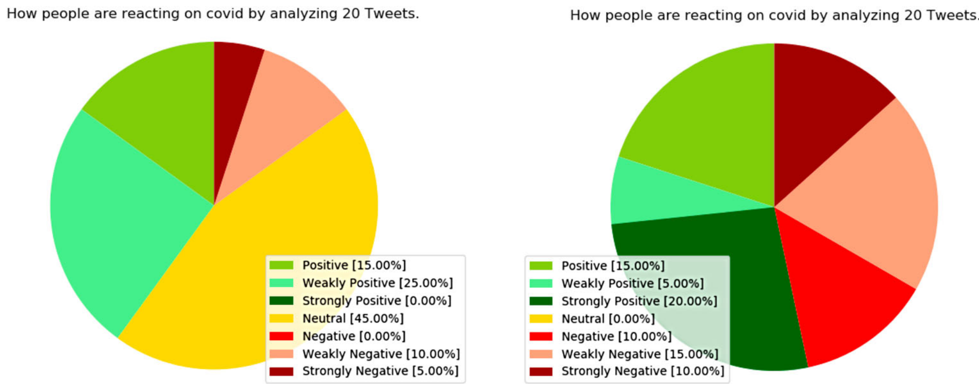
The large-scale presence of the given topic on Twitter was already visible in the first rounds, it greatly influences the results. Initially, the first difference between the trained models came out on a smaller sample of 10 and 20 pieces. Using the functions of TextBlob, you can see how many different and cluttering tweets direct the end result of the analysis to the neutral topic, and we often get a smaller but positive end result, which of course was also the case in our own model. (With a smaller or zero neutral segment and a better distributed area.)

Primarily against the background of this phenomenon, looking at the test datasets, which data currently analysed, it was noticeable that the age group currently on twitter who is mostly active is young/younger. Thus, school closes appear as a positive phenomenon in smaller samples and with a small positive and neutral direction for the end result. In addition, the hospital donations also moves the end result in a positive direction. There is a trend in addition to negative deaths, tweets about these donations and cohesion are much more present even in small samples, of course here the influence of the current scrape is great on what data it collects. Plus, the factuality of newscasts also reinforces the neutral or weakly positive or weakly negative slices. One cannot emotionally shift the simple statement in any direction in most cases.

Other third-party models will not be mentioned in detail, as an analyst based on a simple dictionary has already given completely misleading results on tweets that have reported positive or negative disease of the virus outcomes on a given topic. Like (Figure 2), the textBlob and our own well-trained model were able to filter out these word turns and manifestations really accurately. (Maybe, the RNN looks more significant, but now, we cannot prove it 100%, but the RNN has not have a Neutral section most of the time, which gives us more improvement to the analysis.) Mainly the amount of test data will be the influencing factor.

Note: The RNN model was trained based on an imdb review dataset (In test and train dataset sections using shuffle method as well. Then we use the fresh scraped dataset as test dataset with this trained model.)

We can see on the figures (For this run, the keyword was the 'covid'.), the RNN managed to categorization on all tweets without giving a neutral result, so we conclude that the model 'was better' defined in the smaller details and categorized it based on the small details. Our model stands out in the strongly positive and the weakly negative sections, which is a good indicator of the division of the topic and the abundance of interactions on the topic. Of course, it can be noticed that on social media platforms, positive manifestations continue to dominate which also driven by partial results, but it is also realistic that there are also calls for negative and different perspectives. TextBlob also deviates in the positive direction as our model, both results tipped in the same direction, but a larger neutral value can also be noticed in this case in addition to the negative manifestations. Overall, the categorization of both models can be realistic, the difference is to be found primarily in the detail handling of the models, which hopefully our model handled better even with so little test data. Figure 2 worked from this DataSet (Figure 3).




(a) (b)

**Figure 2.** Analysis of sample of 20 tweets by TextBlob and RNN, using 'covid' keyword. (a) TextBlob result, (b) RNN result.

| Time | UserName | Tweet_text | All_Hashtags | Followers_count |
|---|---|---|---|---|
| 2020.04.15 23:59 | John_R_Amelia | CenturyLink's adaptive networks flourish during COVID-19 crisis https://t.co/iy9zDQcxrz https://t.co/da8amB4nOR | [] | 794 |
| 2020.04.15 23:59 | TheRecipe6 | If not for covid 19 you were just enjoying your rich celebrity life Don't act like you care now please Lol But just for the record the answer is "Nothing | [] | 10822 |
| 2020.04.15 23:59 | BitcoinBiology | _salad It literally says "COVID-19 originated in a Wuhan laboratory not as a bioweapon" Are you assuming every biosafety lab leak is a bioweapon ? | [] | 215 |
| 2020.04.15 23:59 | HayleySnider2 | Not COVID but Iâ€™ve literally had people tell me if I donâ€™t believe in god my CF will be much worse and that I can be saved if I pray or whateve | [] | 174 |
| 2020.04.15 23:59 | meganbushway15 | _Boom _mckraken _tony _SecondFleet Then there's this: https://t.co/wTg4lbeuul | [] | 1400 |
| 2020.04.15 23:59 | theladralph | #EXCLUSIVE: Fox News is reporting that Covid-19 originated in a Wuhan laboratory (China not as a bioweapon but as part of China's effort to demon | ['EXCLUSIVE'] | 44 |
| 2020.04.15 23:59 | RadWagon | It's good to see some politicians making sure they let everyone know that they are in this as well Isn't it I'm mean isn't it #Tories I mean isn't it ? htt | ['Tories'] | 1136 |
| 2020.04.15 23:59 | imleftcoast | "People should know the facts. Blames WHO &amp claims Covid-19 is the 19th strain of the virus #TrumpPlague | ['TrumpPlague'] | 425 |
| 2020.04.15 23:59 | jimgoldstein | UW team illustrates the adverse impact of visiting â€just one friend during COVID-19 lockdown UW News https://t.co/eoqid0nSLz | [] | 25347 |
| 2020.04.15 23:59 | ArrighiOrosz | But they do seem to have the best resistance to the Covid virus Maybe that's the key eat like a toddler đź¤Łđź¤Łđź¤Ł | [] | 72 |
| 2020.04.15 23:59 | VestaviaVoice | Rachel Brockwell has been selling handmade scrunchies at The Clotheshorse since last fall but when the COVID-19 pandemic hit she created a patt | [] | 1610 |
| 2020.04.15 23:59 | stockx | Hey there our team just followed up through email Thank you for your patience at this time we appreciate it! https://t.co/Xkqzr9lmL8 | [] | 241983 |
| 2020.04.15 23:59 | DavidACohen_MD | Inpatient thoughts: I canâ€™t count how many hundreds of times Iâ€™ve told a patient they have cancer. But telling them they have COVID seems | [] | 2084 |
| 2020.04.15 23:59 | KaitMarieox | One day after CNN ran Chinese propaganda its being confirmed that Covid-19 was created in a Chinese laboratory &amp their government covered | [] | 345239 |
| 2020.04.15 23:59 | daniialonzo | COVID-19 Sonora https://t.co/UvGWa51HbW | [] | 208 |
| 2020.04.15 23:59 | deluded_jim | Ok so those deaths are COVID-19 but ERs and ICUs that are suddenly so full Trump brings a hospital ship up to New York is probably because of an u | [] | 178 |
| 2020.04.15 23:59 | phifedogg1 | over 65,000 people died from the flu in 2019 a lot of those were children no child in the US has died from covid-19 under the age of 17 The joke is yo | [] | 10 |
| 2020.04.15 23:59 | Khelrauko | The UK is currently contributing approximately 9 of the total deaths from COVID-19 in the world A country with a world population share of 0.87% I' | [] | 28 |
| 2020.04.15 23:59 | Nanjimo | Maybe itâ€™s a new symptom of COVID-19 (For Kellyanne thatâ€™s 2019) | [] | 81 |
| 2020.04.15 23:59 | bluetreepoint | With the spread and impact of COVID-19 have we now entered a Bearish Market period #bearish #trading #market #investing #markets #covid19 ht | ['bearish', 'trading', 'market', 'investing', 'markets', 'covid19'] | 3 |

**Figure 3.** 15 April 2020 DataSet.

We continue to compare TextBlob and our own RNN model, how it performs on larger and larger test datasets, and how accurate it gives less erroneous results, with double denials and other, 'sleng' and general manifestations, reports.

Between 24 April 2020 and 25 April 2020 on the sample of 50, we can see the increasing distance towards the two extremes. In the case of the RNN (b) (Figure 4) model, again, tweets did not fall into the neutral category, they were subdivided into weakly negative and weakly positive parts, as opposed to textBlob (a) (Figure 4), where there is a more significant neutral unit. In addition, there is a kind of progress towards extremism, which can be concluded that people are already starting to 'get bored' of this whole topic, and the daily numbers and the situation itself. Of course, the high divisions can be inferred from the different policy reactions and announcements and the tweets that respond to them, which either express a sympathetic opinion or a dissenting opinion about the situation. (Looking at the dataset, we can see a strong wave of manifestations about the decisions and political influence of the WHO – which amplified positive negative opinions probably.)

Increasing (200 tweets) the dataset but still using the keyword 'covid', we can see that the division is still similar. A kind of increase in the positive direction can be detected,but this increase in the amount of tweets can be explained in this case of both models.(There is

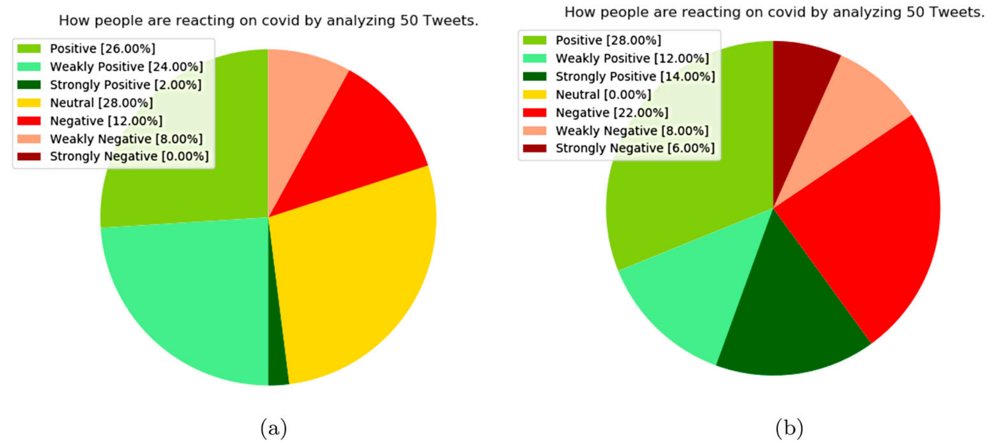



**Figure 4.** Analysis of sample of 50 tweets by TextBlob and RNN, using 'covid' keyword. (a) TextBlob result, (b) RNN result.

a difference in the strength of positivity between the two results.) The increased number of tweets shows that mostly the 'positive expression', the support, and the 'hope' – greater extent of positivity is still highly present in social media – which was expected, but negative messages are also present in significant amounts on the subject (Figure 5).

Using the keyword 'coronavirus' and a much larger dataset, the result is very similar to the trends so far. Smaller increase in both positive and negative directions, we can see only smaller movements in the strength of positivity or negativity (Figure 6).

Our model did not place a tweet in a neutral section, which makes it easier to see differences of opinion. It should also be mentioned that our model evaluates tweets between 0 and 1, while textBlob between −1 and 1. The categories would be defined accordingly, so that a few small details of the tweet are able to move that into another category. Because

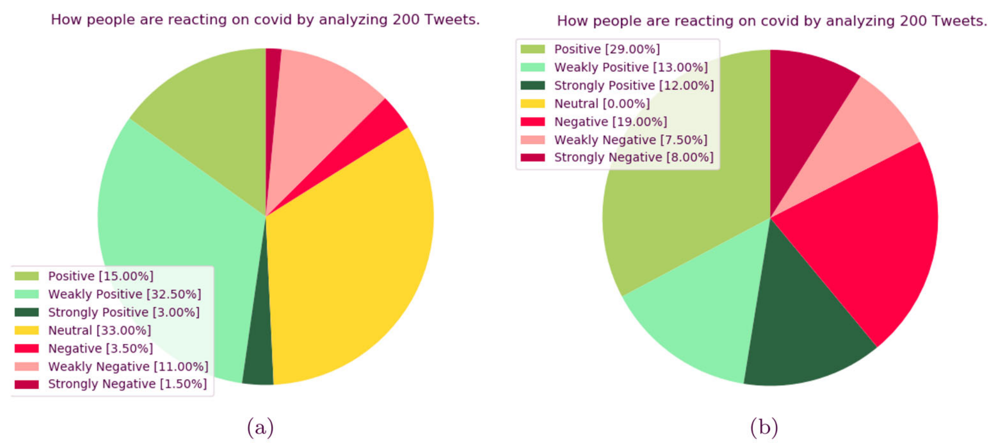



**Figure 5.** Analysis of sample of 200 tweets by TextBlob and RNN, using 'covid' keyword. The time period stands between 24 April 2020 and 25 April 2020. (a) TextBlob result, (b) RNN result.

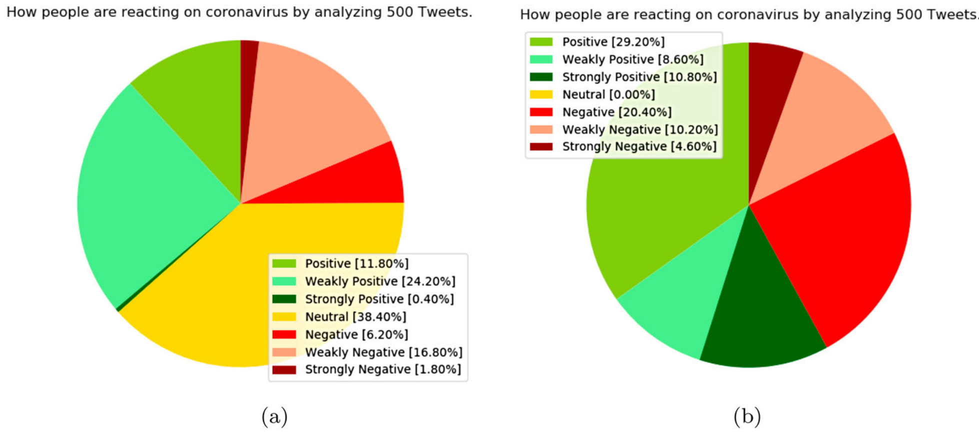



**Figure 6.** Analysis of sample of 500 tweets by TextBlob and RNN, using 'coronavirus' keyword. The time period stands between 24 April 2020 and 25 April 2020. (a) TextBlob result, (b) RNN result.

of these small details, we can say that perhaps we can get a more comprehensive picture in order to avoid neutrality (Figure 7).

People are divided on this topic, just like in most other cases nowadays. They clash arguments on different topics, try to draw conclusions in this place, convince others about the support. In most cases, strong negativity or positivity is a bit, meaning most people are not biased completely towards one side, but there is a visible percentage who is totally biased – that is normal today. For smaller DataSets, these are strengthened a little better, which is also normal.

Examining the period between 13 May 2020 and 14 May 2020 again using the keyword 'covid', we obtained a result (with 200 tweets) very similar to the previous month's 200 tweets. Overall, movement can be observed in the categories delimiting strength (weakly, strongly) within the positive and negative sides. So, it can be concluded that the RNN model (b) (Figure 8) continues to deliver significantly results with small

| Time | UserName | Tweet_text | All_Hashtags | Followers_count |
|---|---|---|---|---|
| 2020.04.24 23:59 | GoReactforASL | As colleges and schools scramble to take their teaching online in response to the spread of the Coronavirus dise: | ['deaf', 'hardofhearing'] | 597 |
| 2020.04.24 23:59 | BillVanSarasota | Why would WHO Dir Tedros Adhanom Ghebreyesus applaud China allowing CoVid-19 to spread worldwide? Dee | [] | 334 |
| 2020.04.24 23:59 | l78lancer | Actually watch it again He specifically turned and asked Birks if there's some way to "bring the light inside the bc | ['coronavirus'] | 1120 |
| 2020.04.24 23:59 | ComradeArthur | More on those English (and European excess mortality numbers. https://t.co/NtGvcPg4dF https://t.co/1MJAurfd | [] | 3051 |
| 2020.04.24 23:59 | INTELNESIA | #coronavirus update Middle East: ďź‡ą Turkey 3,122 new cases ďź‡¸ Saudi 1,172 new cases ďź‡® Iran 1,168 new ca | ['coronavirus'] | 257 |
| 2020.04.24 23:59 | JayneTu79859110 | He has abused the lockdown by imposing oppressive and unrelated rules and regulations He has not even achie\ | [] | 64 |
| 2020.04.24 23:59 | PhilBo35 | How can you not love David Dobrik Cars 10gâ€™s Xboxâ€™s PS4â€™s #coronavirus #DavidDobrik https://t.co/tiLx | ['coronavirus', 'DavidDobrik'] | 53 |
| 2020.04.24 23:59 | AngelAzpeitia | Coronavirus Email From Stanford Professorâ€™s Wife Claimed His Antibody Study Would Prove If You Were Imm | [] | 1402 |
| 2020.04.24 23:59 | BigDaddyKasyoka | said we should inject lysol as a possible way to treat coronavirus. This is the president of the USA | [] | 1210 |
| 2020.04.24 23:59 | besthealthyou | A major blow for #football aka #soccer fans #Bundesliga is not coming back on May 23 Due to #COVID19 #pandem | ['football', 'soccer', 'Bundesliga', 'COVID19', 'pandemic', 'Germany', 'coronavirus'] | 361 |
| 2020.04.24 23:59 | HernanfotoLive | Top story Trump plans to cut daily coronavirus briefings Axios https://t.co/l1twdIZRzu see more https://t.co/Mo\ | [] | 698 |
| 2020.04.24 23:59 | Jim_ItsNotRight | The Republicans who were once so pro-life they fought over one woman on life support now want to sacrifice gı | ['GOPCorruptionOverCountry'] | 3088 |
| 2020.04.24 23:59 | RockinRobin2012 | Hi _Beauvau _Bund thousands of refugees are at risk of #COVID19 on Greek islands due to crowded unsanitary cc | ['COVID19', 'LeaveNoOneBehind'] | 118 |
| 2020.04.24 23:59 | WaNonprofits | Educational nonprofits Nuestra Casa and translate state and local emergency information into Spanish https://t. | [] | 2152 |
| 2020.04.24 23:59 | WnCSohn | Apparently this person is the "Secretary of Health Dr Rachel Levine" https://t.co/KcectR2PeO | [] | 344 |
| 2020.04.24 23:59 | laughinghyena13 | SNHU to cut tuition from $31,000 to $10,000 revamp on-campus learning https://t.co/UBhNt2Vb1O Putting educa | [] | 1020 |
| 2020.04.24 23:59 | DonaldJCarlin | Getting FUCK bad reviews of the job the Unitary Globes is doing off CoronaVirus Ventilators blind eye hospitals l | [] | 2147 |
| 2020.04.24 23:59 | RepMarkTakano | I had a great time chatting with and answering questions sent in by YOU Check out the Coronavirus Daily podcast | [] | 55934 |
| 2020.04.24 23:59 | HEDGEenergy | New Article U.S CDC reports 865,585 coronavirus cases 48,816 deaths https://t.co/wAqd9V4KSN #Politics April 24 | ['Politics'] | 8926 |
| 2020.04.24 23:59 | kr3at | #CoronaVirus Outside of China 2,743,870 cases and 192,340 deaths. To date a total of 196,972 deaths and 2,826,67 | ['CoronaVirus', 'covid19', 'CoronaVirusOutbreak'] | 18442 |
| 2020.04.24 23:59 | TheBookTweeters | Espionage was nothing new to Captain Larry McGraw and the crew of the USS Charlotte Orders from the White H | ['Thriller', 'China', 'Coronavirus'] | 235264 |
| 2020.04.24 23:59 | IntegrityTeam1 | Please donate to 501(c)(3 charity https://t.co/DfPGUGaXlp Donation will fund COVID-19 [#Coronavirus recovery a | ['Coronavirus'] | 19493 |
| 2020.04.24 23:59 | ragnasun8 | Revealed 'former Vodafone executive in 5G conspiracy video is UK pastor https://t.co/nzwbFj92qB | [] | 17386 |
| 2020.04.24 23:59 | alixtersa | Young and middle-aged people barely sick with covid-19 are dying from strokes https://t.co/LANbB0W2En | [] | 325 |
| 2020.04.24 23:59 | EliSananes | _regan But the Coronavirus a hoax?? Stay out of this one.... | [] | 96 |
| 2020.04.24 23:59 | MalibuEOC | UPDATE 4/24 4PM 430 tested in Malibu Friday Out of area public urged not to visit Malibu Malibu residents shoul | [] | 1886 |
| 2020.04.24 23:59 | damedic7276 | Young and middle-aged people barely sick with covid-19 are dying from strokes https://t.co/Rt44OZmBU1 | [] | 802 |
| 2020.04.24 23:59 | ExciteMagazine | _Jordan Dr.J.Trump Prescriptions injects disinfectant to kill Coronavirus Hahahaha GOPs disasters #GOPsDisaster | ['GOPsDisasters'] | 52 |
| 2020.04.24 23:59 | Too5moothTim | Hell yeah love my boys They don't know shit about #coronavirus https://t.co/AFAcBElCLh | ['coronavirus'] | 59 |
| 2020.04.24 23:59 | MarcDFroese | Justin Trudeau confirms China blocking consular visits to detained Canadians who have been held for 500 days h | [] | 1298 |
| 2020.04.24 23:59 | mellafinger | New York City to distribute 500,000 free halal meals during Ramadan https://t.co/cnZHO7mKfA | [] | 822 |
| 2020.04.24 23:59 | FloridaPost1 | German lives â€valued five times higher than British in coronavirus treatment https://t.co/YUavsAYeAj https://t. | [] | 15 |
| 2020.04.24 23:59 | nrmiller01 | Which would do nothing to combat the coronavirus in patients. | [] | 10 |
| 2020.04.24 23:59 | FactsMathLogic | why are you bragging about having 22 million tests done than you promised weâ€™d have almost a month ago... | [] | 1604 |
| 2020.04.24 23:59 | nickcunningham1 | Fantastic look at how the collapse of oil will destabilize oil-producing countries https://t.co/wCL6Fb8XNa | [] | 2196 |
| 2020.04.24 23:59 | rozina_atiq | _z000 _HFQ _se_hu I think she is bored because of #coronavirus #lockdown and also canâ€™t watch #Ertugrul for | ['coronavirus', 'lockdown', 'Ertugrul'] | 27 |
| 2020.04.24 23:59 | CBSNews | Michael Avenatti released from prison over coronavirus fears https://t.co/u3WDQWaUjK https://t.co/AefjvMXYf | [] | 7524435 |

**Figure 7.** 24 April 2020 Part of the dataSet of the 'coronavirus' keyword.

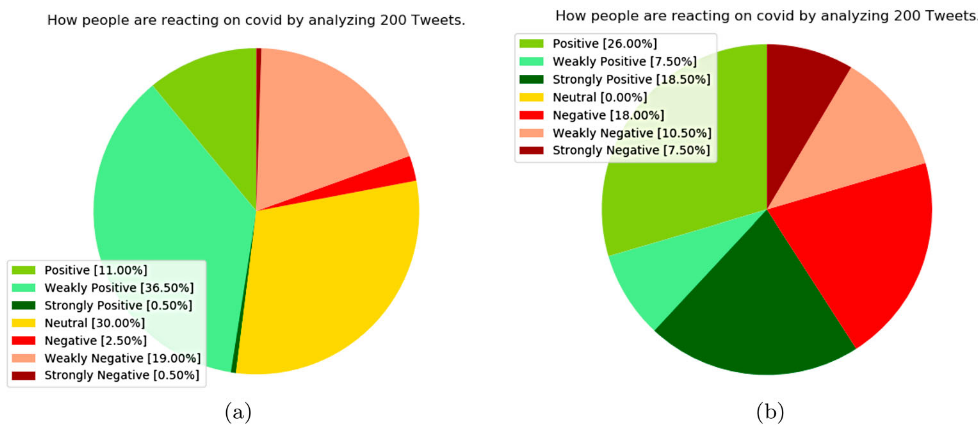



**Figure 8.** Analysis of sample of 200 tweets by TextBlob and RNN, using 'covid' keyword. The time period stands between 13 May 2020 and 14 May 2020. (a) TextBlob result, (b) RNN result.

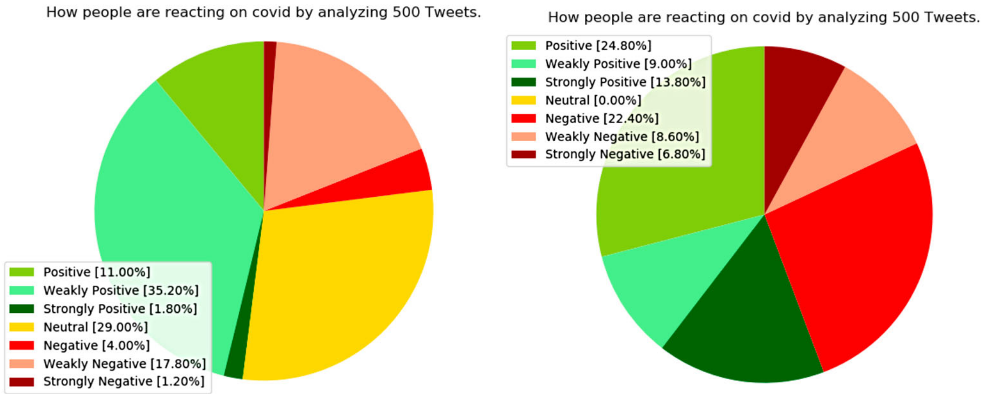
changes in time and still without categorization into the neutral section. Looking at the program logs we saw that in some cases decimal values have decided a particular writing to be weakly positive or negative and not neutral. The weakly positive in this case is 7.50% and the weakly negative is 10.50%. Based on the results of the RNN model, it can be said that positivity is still more present in social media in the case of pandemic-related manifestations. Based on the result of TextBlob (a) (Figure 8), we see a similar result in the positive direction, but with a significant 30% neutral data, and the weakly positive section is 36.50% against the RNN's 7.50%.

Overall, the RNN chart provides a much more realistic and thorough picture of current emotional levels (for us) with minimal or even zero neutral results.

If we increase the number of tweets to 500 in the same time period. In the case of the RNN model, we can observe a strengthening in the negative section (simple negative not together with the strongly and weakly negatives), which can also be said for the result of TextBlob.

In our textBlob (a) (Figure 9) analysis, we can see again 29% of neutral value, in addition to a weakly negative value of 17.80%. For the RNN model (b) (Figure 9), again, the neutral result is 0% and only 8.60% is weakly negative. Overall, comparing the categorical values of the two analyses, the positive displacement can be said again, but the division of this end result is reflected in a completely different way in the two models. In the case of RNN, a positive value of 24.80% can be observed, in addition to the negative value, which is 22.40%, which is a proportionate division and the positive manifestation in the sample of 500 tweets are a little more. In contrast, in a TextBlob analysis, weakly positive value is 35.20%, which is dominate. The positive value is 11% and a negative value is 4%.

The reactions and evaluations of various political announcements and decisions, after the announcement, provoke significant activity from the people who argue and talk about the effects in the social media. Thus drastically increasing the number of tweets related to the topic. A similar reaction has been shown by various international events on this subjects, especially after the details have been described. (There is a visible shift into the positive and negative directions, sometimes from the neutral, but also there are some changes in the strength distribution of positivity and negativity itself.)




(a) (b)

**Figure 9.** Analysis of sample of 500 tweets by TextBlob and RNN, using 'covid' keyword. The time period stands between 13 May 2020 and 14 May 2020. (a) TextBlob result, (b) RNN result.

| Time | UserName | Tweet_text | All_Hashtags | Followers_count |
|---|---|---|---|---|
| 2020.05.13 23:59 | CarePageAU | By gathering regular feedback #agedcare providers can be responsive to the changing mood | ['agedcare', 'COVID', 'feedback'] | 1358 |
| 2020.05.13 23:59 | NewsMedical | Should school closures related to COVID-19 be continued long-term https://t.co/C8Tc8NdWI | ['Coronavirus', 'Education', 'SARSCoV2', 'COVID19', 'CV19', 'Virus', 'Pandemic', 'Schools', 'Scho | 13122 |
| 2020.05.13 23:59 | lexi_kenney | In the past 3 months weâ€™ve laughed together mourned drank danced and debated To me | [] | 1711 |
| 2020.05.13 23:59 | ByrdAdatto | Episode 43 ByrdAdatto discusses updates to advertising rules and regulations for businesses | [] | 45 |
| 2020.05.13 23:59 | ActonInstitute | . on the Dan Proft Show COVID-19 lockdown orders are the state-mandated â€marginalizatio | [] | 22583 |
| 2020.05.13 23:59 | AprilGarbusjuk | Empathy is a critical component in communications and relationship-building both now and | ['SalesStrategy'] | 830 |
| 2020.05.13 23:59 | oliviadeng1 | Fuck anyone who says Charlie Baker is handling the COVID-19 pandemic well It took him a m | [] | 1041 |
| 2020.05.13 23:59 | bettycjung | US formally accuses China of hacking US entities working on COVID-19 research https://t.co/ | ['COVID19hacking'] | 712 |
| 2020.05.13 23:59 | guy123_g | Blocked for thinking all teens are ignorant of COVID-19. | [] | 16 |
| 2020.05.13 23:59 | MairiMMcInnes | _UK's new analysis on the â€return to work strategy and what a phased approach to lifting se | ['PwC', 'futureofwork'] | 133 |
| 2020.05.13 23:59 | JasonClark829 | If someone has a heart condition and wasnâ€™t going to die but contracts covid and dies yes | [] | 44 |
| 2020.05.13 23:59 | Coffee_2222 | #CCPvirus #Wuhan #Coronavirus China Removes All Traces of â€Thousand Talents Program C | ['CCPvirus', 'Wuhan', 'Coronavirus'] | 397 |
| 2020.05.13 23:59 | LightTherapy_ | Wow this is a Success for Wisconsin Stay at home order is Unconstitutional from Court đź‡đ | [] | 103 |
| 2020.05.13 23:59 | q_abolitionist | Yo â¦ donâ€™t get any ideas This wonâ€™t go the way you think it will https://t.co/yqsXKwF | [] | 1406 |
| 2020.05.13 23:59 | Sucksie_Sucks | Almonds are high in Zinc which is Trump's covid's home remedy I will get cuz Jaz to ban then | [] | 82 |
| 2020.05.13 23:59 | TUDigitalMedia | With many concerned with how the pandemicâ€™s trend line will impact future media mark | [] | 52 |
| 2020.05.13 23:59 | TheChandlerDude | @__ListenUp_ Very few jurisdictions have been reporting COVID-19 mortality properly "Exc | [] | 1841 |
| 2020.05.13 23:59 | cannonhillpark | A Wings and Scrubs Angel in Lightwoods Park on Hagley Rd by sculptor Luke Perry shines a li | [] | 3428 |
| 2020.05.13 23:59 | michele_kay_ | The end of an eráđź™Źđź Fuck you Covid for pillaging the remainder of my youth | [] | 257 |
| 2020.05.13 23:59 | 903KAZU | On our COVID-19 blog today: -No restrictions on who can get tested in Monterey County.. ev | [] | 1615 |
| 2020.05.13 23:59 | StephenNee2 | Which figures More than 30,000 excess deaths so far this year in UK v any previous average F | [] | 76 |
| 2020.05.13 23:59 | compass_housing | Compass Professor Adamson is a leading advocate for #HousingForAll and recently participa | ['HousingForAll'] | 693 |
| 2020.05.13 23:59 | AamirKhan9feb97 | Day 208 19th&20th Ramadan Mubarak Allah ap &amp apki family ko Salaamat rakhe Aar | [] | 16 |
| 2020.05.13 23:59 | weact2 | Vitamin D determines severity in COVID-19 so government advice needs to change https://t | [] | 3078 |
| 2020.05.13 23:59 | Feminismandfre1 | What if it isnâ€™t covid and is Kawasaki Serious question. | [] | 448 |
| 2020.05.13 23:59 | oranglaut | 'This virus may never go away, WHO says https://t.co/C4cfHwc7YM | [] | 10183 |
| 2020.05.13 23:59 | frankkimmel4 | With hospitals incented by Medicare payments to report all deaths as CoVid-related regardl | [] | 3 |
| 2020.05.13 23:59 | jeffjameslee | John Ivison Ottawa's COVID-19 debt binge runs very real risk of ruining the next generation I | [] | 187 |
| 2020.05.13 23:59 | Degeniusmedia | BREAKING COVID-19 Cases In Nigeria Near 5000 As Lagos Record Fresh 51 Cases #NoLagosLoc | ['NoLagosLockDown'] | 2887 |
| 2020.05.13 23:59 | shortfamly | There are now reports showing that children from 1 to 18 are developing different effects fr | [] | 74 |
| 2020.05.13 23:59 | Akamai49 | One more thing since you are personally responsible for the evening news content how do j | [] | 1 |
| 2020.05.13 23:59 | dyreerawrs | Yeah.. So this is why the US is never going to be rid of covid. CNN Wisconsin Supreme Court : | [] | 985 |
| 2020.05.13 23:59 | azfamily | Maricopa County Public Health hires trains more contact tracers to track COVID-19 to help de | [] | 256444 |
| 2020.05.13 23:59 | USEmbassySeoul | The gathers #COVID19 data and makes it accessible to the public so that practitioners arounc | ['COVID19', 'WeAreInThisTogether'] | 49202 |

**Figure 10.** 13 May 2020 Part of the dataSet of the 'covid' keyword.

How long the 'covid' and 'coronavirus' topics will be dominant on the entire Internet no one knows. If the vaccine will be available, the topic is still expected to stay with us for a significant period of time and it will still to dominate the various community platforms with its subsequent effects (Figure 10).

## 5. Conclusion and future work

In this work, we use a Recurrent Neural Network (RNN) to classify emotions on tweets. We developed a model to analyse the emotional nature of various tweets, using the recurrent neural network for emotional prediction, searching for connections between words, and marking them with positive or negative emotions. Where instead of simple positive and negative extremes, we have classified the various texts into a much more articulated class of emotional strength (weakly positive/negative, strongly positive/negative). This has been combined with a keyword-based special data scraper, so we can apply our taught RNN model with these specific freshly scraped datasets. As a result, we get an emotional classification related to specific topics. What kind of tweets they were and what emotional class they belong to, what is the distribution on that topic at the emotional level within the given start interval. In the article, we focused most on the coronavirus and related emotional changes and fluctuations, and it was shown that the overall positive manifestation and presence on the social platform remained on social media surfaces during this pandemic. Of course, in addition to negative and other manifestations. Over time, positivity has strengthened, but there is also a stronger negative array that is natural. According to our expectations this topic remain positive manifestations, sometimes with a higher and sometimes with a smaller percentage. It can be seen that the recurrent neural network provides good performance and prediction in text classification. Where the RNN model brought a smaller amount of data in neutral result or completely reduced to zero that. Which proves that our model is 'able to make' a decision and categorize in some direction even on the basis of small details. Our comparisons were made mainly against TextBlob, which also worked very well and delivered stable results, but there were many times when the neutral results were above 30% compared to our RNN model, which we cannot use as usefully for further evaluations as for our RNN model. The classification of emotions for both models (TextBlob, RNN) was properly segmented.

For future work and further development, it may be advisable to create an interface that better visualizes and interacts with users, which can be supplemented with sophisticated database management for archiving, tracking, and exploring datas to other areas. We can further expand the analysis by introducing various classifications and clusters as well as other data analyses. Allowing examinations and comparisons from a new perspective, in addition to emotional analyses, may even provide an opportunity to further support current results and compare the conclusions. In addition, implementing or refactoring future potential tensorflow features and keeping it up to date.

## Note

1. https://developers.google.com/machine-learning/glossary/#recurrent_neural_network

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The project has been supported by the European Union, co-financed by the European Social Fund [grant number EFOP-3.6.3-VEKOP-16-2017-00002].

## Notes on contributors

*László Nemes* received the B.Sc. degree in computer science from Eötvös Loránd University in 2020 and currently pursuing a M.Sc. degree. He is a Demonstrator with the Department of Media and Educational Technology, Eötvös Loránd University.

***Attila Kiss*** was born in 1960. In 1985 he graduated (MSc) as mathematician at Eötvös Loránd University, in Budapest, Hungary. He defended his PhD in the field of database theory in 1991. Since 2010 he is working as the head of Information Systems Department at Eötvös Loránd University. His scientific research is focusing on database theory and practice, security, semantic web, big data, data mining, artificial intelligence and bioinformatics. He was the supervisor of seven PhD students. He has more than 145 scientific publications.

## ORCID

*László Nemes* https://orcid.org/0000-0001-6167-9369
*Attila Kiss* https://orcid.org/0000-0001-8174-6194

## References

Arras, L., Montavon, G., Müller, K. R.., & Samek, W.2017). *Explaining recurrent neural network predictions in sentiment analysis*. Preprint arXiv:1706.07206.

Balahur, A. (2013). Sentiment analysis in social media texts. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 120–128).

Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep convolution neural networks for twitter sentiment analysis. *IEEE Access*, *6*, 23253–23260. https://doi.org/10.1109/ACCESS.2017.2776930

Leskovec, J. (2011). Social media analytics: Tracking, modeling and predicting the flow of information through networks. In *Proceedings of the 20th international conference companion on world wide web* (pp. 277–278).

Liu, P., Qiu, X., & Huang, X. (2016). *Recurrent neural network for text classification with multi-task learning*. Preprint arXiv:1605.05101.

Mikolov, T., Karafiát, M., Burget, L., Černockỳ, J., & Khudanpur, S. (2010). Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*.

Mikolov, T., Kombrink, S., Burget, L., Černockỳ, J., & Khudanpur, S. (2011). Extensions of recurrent neural network language model. In *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 5528–5531).

Muhammad, A., Wiratunga, N., & Lothian, R. (2016). Contextual sentiment analysis for social media genres. *Knowledge-Based Systems*, *108*, 92–101. https://doi.org/10.1016/j.knosys.2016.05.032

Nallapati, R., Zhai, F., & Zhou, B. (2017). Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Thirty-first AAAI conference on artificial intelligence*.

Neri, F., Aliprandi, C., Capeci, F., Cuadros, M., & By, T. (2012). Sentiment analysis on social media. In *2012 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 919–926).

Ortis, A., Farinella, G. M., Torrisi, G., & Battiato, S. (2018). Visual sentiment analysis based on on objective text description of images. In *2018 international conference on content-based multimedia indexing (CBMI)* (pp. 1–6).

Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management 53*(4), 764–779. https://doi.org/10.1016/j.ipm.2017.02.004

Schmidt, A., & Wiegand, M. (2017). A survey on hate speech detection using natural language processing. In *Proceedings of the fifth international workshop on natural language processing for social media* (pp. 1–10).

Wang, Y., & Li, B. (2015). Sentiment analysis for social media images. In *2015 IEEE international conference on data mining workshop (ICDMW)* (pp. 1584–1591).

Xu, J., Huang, F., Zhang, X., Wang, S., Li, C., Li, Z., & He, Y. (2019). Sentiment analysis of social images via hierarchical deep fusion of content and links. *Applied Soft Computing*, *80*, 387–399. https://doi.org/10.1016/j.asoc.2019.04.010