AJIT-e *Academic Journal of Information Technology*
10.5824/ajite.2022.02.001.x

Check for updates

2022 • Spring – Bahar • Volume – Cilt: 13 • Issue – Sayı: 49

Twitter Sentiment Analysis During Covid-19 Outbreak with VADER

Cihan ÇILGIN, *Abant İzzet Baysal University, Department of Management Information Systems, Research Assistant, cihancilgin@ibu.edu.tr,* *0000-0002-8983-118X*

Metin BAŞ, *Beykent University, Department of Management Information Systems, Research Assistant, metinbas@beykent.edu.tr,* *0000-0002-2783-5513*

Hande BİLGEHAN, *Department of Business Administration, Doctorate Student in Business, handebilgehan@gmail.com,* *0000-0003-0844-8451*

Ceyda ÜNAL, *Dokuz Eylül University, Department of Management Information Systems, Research Assistant, ceyda.unal@deu.edu.tr,* *0000-0002-5503-8124*

ABSTRACT

The Covid-19 outbreak, which has been under the influence of Europe since then, continues to spread rapidly especially in the American continent. Looking at the current data, the virus has affected about 250 million people and has killed more than five million people. Especially with the rapid spread of the outbreak in the European continent, this issue started to be discussed in social media. In particular, Twitter is the most frequently used micro-blogging in this workspace. In this study, it is aimed to analyze the tweets shared by many people, organizations and government agencies through Twitter during the global COVID-19 outbreak with sentiment analysis using the VADER Sentiment Analysis method. The hashtags #covid19, #Covid, #pandemic, #social-distancing, #socialdistance, #covid-19, #corona-virius, #coronavirus, #Chinesevirus, #Chinese-virus were used in this study. With these hashtags, a total of 60,243,040 tweets were collected from Twitter between January 1, 2020 and July 1, 2020. In this study, we use the VADER to classify the sentiments expressed in Twitter data related to Covid-19 and the compound scores of the resulting tweets were divided into five categories: Highly Positive, Positive, Neutral, Negative, Highly Negative. In addition, in the study, the Wordcloud was used to visualize the most frequently collected text data monthly, and N-grams were applied to the tweets to better understand the content of the tweets. When the results obtained in the study are examined, the tweets shared about Covid-19 in different periods of the release reflect different sentimental situations.

Keywords **:** *Covid-19, Coronavirus, Sentiment Analysis, VADER, Twitter*

Araştırma Makalesi / Research Article
Date Received: 13.02.2022
Publishing Date: 31.05.2022

Covid-19 Salgını Esnasında VADER ile Twitter Duygu Analizi

ÖZ

Avrupa'yı etkisi altına aldığından beri Covid-19 salgını, özellikle Amerika kıtasında hızla yayılmaya devam etmektedir. Güncel verilere bakıldığında virüs yaklaşık 250 milyon insanı etkilemiş ve beş milyondan fazla insanın ölümüne neden olmuştur. Özellikle Avrupa kıtasında salgının hızla yayılmasıyla birlikte bu konu sosyal medyada tartışılmaya başlanmıştır. Özellikle Twitter bu çalışma alanında en sık kullanılan mikroblogdur. Bu çalışmada, küresel COVID-19 salgını sırasında Twitter üzerinden birçok kişi, kuruluş ve devlet kurumu tarafından paylaşılan tweetlerin VADER Duygu Analizi yöntemi kullanılarak, duygu analizi gerçekleştirilmesi amaçlanmaktadır. Bu çalışmada #covid19, #Covid, #pandemic, #social-distance, #socialdistance, #covid-19, #corona-virius, #coronavirus, #Chinesevirus, #Chinese-virus hashtagleri kullanılmıştır. Bu hashtag'ler ile 1 Ocak 2020 ile 1 Temmuz 2020 tarihleri arasında Twitter'dan toplam 60.243.040 tweet toplanmıştır. Bu çalışmada, Covid-19 ile ilgili Twitter verilerinde ifade edilen duyguları sınıflandırmak için VADER kullanılmış ve ortaya çıkan tweetlerin bileşik puanları, çok olumlu, olumlu, nötr, olumsuz, çok olumsuz olmak üzere beş kategoriye ayrılmıştır. Ayrıca çalışmada, aylık olarak en sık toplanan metin verilerinin görselleştirilmesi için Wordcloud kullanılmış ve tweetlerin içeriğini daha iyi anlamak için tweetlere N-gram uygulanmıştır. Çalışmada elde edilen sonuçlar incelendiğinde, çıkışın farklı dönemlerinde Covid-19 ile ilgili paylaşılan tweetlerin farklı duygusal durumları yansıtmaktadır.

Anahtar Kelimeler **:** *Covid-19, Koronavirüs, Duygu Analizi, VADER, Twitter*

INTRODUCTION

Coronavirus is one of the major pathogens that primarily targets the human respiratory system. Coronary virus outbreaks, previously characterized as agents with a major public health threat, include severe acute respiratory syndrome (SAR – CoV) and Middle East respiratory syndrome (MERS-CoV). Coronavirus, which can cause disease in humans and animals and has many species, is thought to first appear in Wuhan, China (Wang et al., 2020, p. 1062) at the end of 2019, and from there it began to spread to the world as a Covid19 (SARS-CoV-2) outbreak (Zhou et al., 2020: p. 271). On March 11, 2020, the Covid19 virus was announced as a pandemic by the World Health Organization (World Health Organization, 2020). SARS-CoV-2 has a stronger infectious capacity compared to SARS-CoV, which caused the SARS outbreak that affected countries such as Hong-Kong, Taiwan, Canada, Singapore in 2003.

The rapid increase in confirmed cases makes the prevention and control of COVID-19 extremely serious (Zheng et al., 2020: p. 259). The Covid19 outbreak, which has been under the influence of Europe since then, continues to spread rapidly especially in the American continent. Looking at the current data, the virus has affected about 33 million people and has killed more than one million people. The outbreak brought along many social and economic problems, along with health problems. Especially with the rapid spread of the outbreak in the European continent, this issue started to be discussed in social media. Many posts about both the course of the outbreak and people's thoughts about the epidemic appeared intensely on social media platforms. Beyond the discussions and opinions, social media platforms such as Twitter played an important role in sharing and acquiring important and striking information about the Covid19 outbreak. In addition, the lockdown and social distance rules have increased the rate of people using social media platforms within the framework of the epidemic measures implemented by country governments.

Today, the use of various platforms such as social media, blog and the rapid spread of online shopping causes users to share their many interpretable thoughts on the virtual environment. The need to analyze these texts automatically increases in the same way due to the rapid increase of data in the digital environment, where millions bytes of data are produced every day. Therefore, the concept of Sentiment analysis or Opinion mining emerged as a subfield of Natural Language Processing. Sentiment analysis or Opinion Mining can be expressed as a classification process by using various Machine Learning algorithms or Deep Learning networks to evaluate a text or author's attitude according to how positive, neutral or negative it is.

Sentiment analysis is commonly used with data from social media to be used in classifying consumers' attitudes by marketing and customer service teams. In addition to these business practices, it can be used to classify mass ideas on financial, social and political issues. Sentiment analysis in the micro-blogging area is a current and intense research topic. In particular, Twitter is the most frequently used micro-blogging in this workspace. In this study, it is aimed to analyze the tweets shared by many people, organizations and government agencies through Twitter during the global COVID-19 outbreak with sentiment analysis. It is certain that a longer-term investigation will be made on the Covid19 outbreak, which can be considered the longest epidemic of the current century. When the studies in the literature are analyzed, then the Sentiment analysis on the Covid19 outbreak is examined and it is seen that they generally include the data sets obtained at short date intervals. In this context, the Covid19 outbreak, which has been going on for about 6 months and is likely to continue, examines the tweets obtained between January 01, 2020 and July 01, 2020, shared in English from all over the world. Especially considering that the emotional state of people during the outbreak varies from day to day, the analysis of all the tweets obtained may not reflect the correct results. For this reason, in this study, we conduct Sentiment analysis for each day for the tweets about Covid19.

The structure of this article is organized as follows. Chapter 2 presents a literature review and some thoughts from the previous study on sentiment analysis of public health. Chapter 3 consists of two parts, the first part describes the data set obtained from Twitter and other data sources and the second part refers to the method used in the study. Chapter 4 presents an analysis of results and findings.

1. LITERATURE REVIEW

One of the important issues related to public health is to follow the issue of public concern in epidemic situations. In this sense, it is very advantageous to use social media platforms published in real time to analyze public health concerns and community ideas in any epidemic situation. When we examine the past studies, a wide variety of Sentiment analysis or Opinion Mining studies have been carried out on the data obtained from the Twitter in the MERS-CoV outbreak (Fung et al., 2013; Shin et al., 2016), H1N1 and swine flue outbreak (Chew and Eysenbach, 2010; McNeill et al., 2016), and ebola outbreak (Kim et al., 2016; Van Lent et al., 2017).

In the study conducted by Dubey (2020) between the dates of 11 March 2020 and 31 March 2020, he conducted emotional analysis using the 50000 tweets he obtained from the Twitter related to "Covid19". In the study, 8 emotions were obtained for tweets belonging to each country by applying NRC Emotion Lexicon from 12 different countries such as Belgium, India, Australia, Netherlands, Spain, United Kingdom, Italy, Germany, France, USA, Switzerland, China. The results of the study revealed that countries such as Belgium, India and Australia tweeted with more positive feelings about COVID19 and people in China also had negative feelings about the same. Similarly, while analyzing the word clouds of different countries, it was concluded that people are tweeting words like Pandemic, Death, Quarantine, Hope, Stay Safe, Government, Political, Fight and Masks with different emotions. The name of the USA President, Donald Trump was amongst one of the most tweeted words not only in USA, but across all the twelve countries considered for the study.

Pokharel (2020) worked on sentiment analysis on Twitter data on COVID-19 outbreaks in Nepal. The data used in the study were collected between 21 May 2020 and 31 May 2020 by using the Twitter API and Tweepy Python library from people who indicated their location in Nepal. TextBlob Library, one of the Python Sentiment Analysis techniques, was used for the 615 tweets obtained. According to the results, the tweets used in the study include feelings of 56% Neutral and comfortable, 19% calm, 8% hopeful, 10% relaxed, 4% pessimistic, 2% optimistic, and less than 1% self-confident. Kaila and Prasad (2020) examined the information flow on Twitter in the Covid19 outbreak. The model created with 18000 tweets was investigated the Covid19 outbreak using sentiment analysis and subject modeling using Latent Dirichlet Allocation. With the LDA analysis, Covid19 identified the most appropriate and accurate issues related to the outbreak. In addition, with the analysis of emotions, the prevalence of negative emotions such as fear and the prevalence of positive emotions such as trust were confirmed by this study. In this context, the authors concluded that Governments and Health authorities effectively use Twitter to disseminate accurate and reliable information.

10.5824/ajite.2022.02.001.x

Andreade et al. (2020) analyzed the production of discourses on Covid19 in the face of political tensions between Brazil and China using tweets in Portuguese. They used 1.6 million tweets from March 19 to April 1, 2020. thematic and sentiment analysis was carried out in this dataset created after the tweets obtained in this study passed various filter processes. The findings reveal the potential of social media to understand the discourses that occurred during the epidemic and to reveal the weaknesses of Twitter management. In addition, the results of the study are revealed in the racism underlying the tweets using the term “Chinese virus” and the negative emotions that arise with the current tensions between Brazil and China.

Kaur and Sharma (2020) obtained 3000 English tweets from various countries and conducted emotional analysis with the remaining 2058 tweets after the pretreatment step of these tweets. According to the results of the research conducted with the TextBlob library of Python, the application performed with these 2058 tweets yielded 24.0% positive, 32.1% negative and 43.9% neutral results.

In their study, Ahmed et al. (2020) investigated both the feelings and emotions of people in the USA about reopening. Between 3 May 2020 and 15 May 2020, they used “# covid19”, “#covid”, “#corona”, “#coronaviras”, “# corona-virus”, “# covid19-virus” and “# sarscov2” hashtags for the 5,703,590 tweets they collected from the Twitter. Findings obtained from the data set in the study; Emotion Analysis results, “Analytical” (34.7%) of the highest percentage of emotional tone, the second highest tone was “Joy” (17.35%), the next few tones are “Temporary”, “Sadness” and “Confident”, respectively, also “Anger” and “Fear” had the lowest percentage in the dataset and in the Sentiment Analysis results, most of the tweets have a neutral sentiment (43.66%) followed by a positive (39.89%) sentiment.

In their papers, Kruspe et al. (2020) analyze Twitter messages (tweets) collected in the first months of the COVID-19 outbreak in Europe in terms of their sentiment. Data sets consist of 4.6 million geotagged Twitter messages collected from December 2019 to April 2020. In the study, it was applied with a neural network for sentiment analysis of multilingual sentences. They analyzed the results by separating them by country and associating them with the temporal development of the outbreak.

This research includes a much longer time period and a large data set compared to other studies in the literature. Thus, it is aimed to analyze the thoughts of the society on COVID-19 and the pandemic more easily in a longer period of time. The results support this situation, revealing different results in different periods and presenting more remarkable findings than existing studies. In addition, the research did not focus on a single country, but included tweets in English from all countries.

2. MATERIAL-METHOD

2.1. Data

The data used in this study were obtained from more than one source. First of all, Python's Scrapy library was used to obtain tweets that are the focus of the study. Considering the size of the data set used in the study, we have developed a Scraping Bot for collecting tweets because the Twitter devolopers API has certain limitations. We used #covid19, #Covid, #pandemic, #social-distancing, #socialdistance, #covid-19, #corona-virius, #coronavirus, #Chinesevirus, #Chinese-virus hashtags in collecting English tweets. With these hashtags, a total of 60,243,040 tweets were collected from Twitter between January 1, 2020 and July 1, 2020. We also used the World Health Organization web page as the second data source. The WHO (2020) website only provides daily confirmed case data from January 11, 2020 and July 1, 2020, and we used this range in our study.

2.2. Methodology

The first step after collecting the data is the data preprocessing phase. First of all, the tweets were removed from the fields such as "usernameTweet" and "ID" that will not be used in the data set with the feature selection. All fields except "Text" and "Datetime" fields to be used within the scope of this study have been removed from the data set. Subsequently, all the uppercase letters were converted into lowercase characters and numbers characters to string expressions was performed for all text characters. It was also removed in stopwords using Python NLTK, and all URL and Email related words, Noisy words, Newlines and Whitespaces and Punctuations were also deleted from all tweets. In addition, all tweets in the dataset have been applied Lemmatization with Python NLTK. Duplicate tweets have been removed because there are too many duplicate tweets in the data set. And after this removal, a total of 52,671,376 tweets remained.

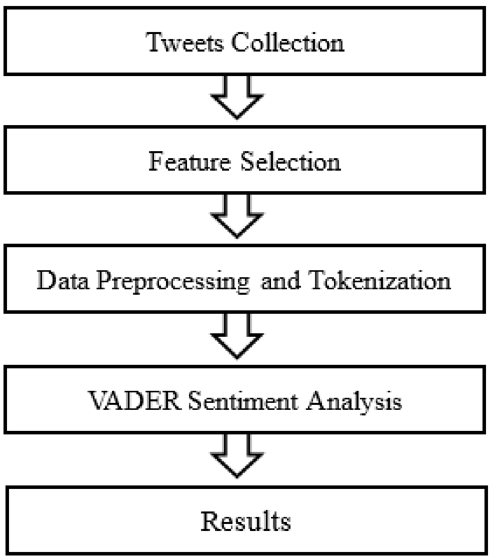

Figure 1: Flowchart of analyzing the research paper

10.5824/ajite.2022.02.001.x

After the preprocessing section, as shown in Figure 1, all tweets were primarily divided into tokens for the sentiment analysis. In this study, we use the Valence Aware Dictionary for Sentiment Reasoner (VADER) to classify the sentiments expressed in Twitter data related to Covid19. VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to the sentiments expressed in social media. It is an open-source tool. VADER also takes into consideration word order and degree modifiers (Chauhan et al., 2018: p. 487).

Developed by Hutto and Gilbert (2014), VADER (F1 = 0.96) actually outperforms even individual human raters (F1 = 0.84) at correctly classifying the sentiment of tweets. We preferred VADER in our study as it has a high classification success especially in the analysis of tweets. In addition, when we examined other studies conducted with VADER, we realized that it was very successful in analyzing emotions of Social Media texts (Cavnar and Trenkle, 1994; Ramteke et al., 2016; Elbagir and Yang, 2019). The VADER Sentiment Analyzer has been used to classify pre-processed tweets as positive, negative, neutral or compound. The compound value is a useful metric for measuring the sentiment in a given tweet. We use the following threshold values and classes in our study, unlike the generally used threshold values to classify tweets as positive, negative and neutral:

Highly Positive sentiment:(compound score > 0.501), assign score = 2
Positive sentiment:(compound score > 0.001) and (compound score < 0.501), assign score = 1
Neutral sentiment:(compound value > -0.001) and (compound value < 0.001), assign score = 0
Negative sentiment:(compound score <-0.001) and (compound score >-0.501), assign score = -1
Highly Negative sentiment: (compound score < -0.501), assign score = -2

Thus, we obtained a more sensitive measurement for the sentiment analysis of tweets by obtaining more than three classes (Highly Positive, Positive, Neutral, Negative, Highly Negative).

Word clouds are useful tools to visually summarize large amounts of text data. In this study, Python's Wordcloud library was used to visualize the most frequently collected text data on a monthly basis. In addition, N-grams were applied on the tweets in order to better understand the contents of tweets.

3. FINDINGS AND DISCUSSION

In order to compare the accuracy of the classification results obtained, a total of 900 tweets from 300 each for three different classes (positive, negative, neutral) were manually evaluated and compared with the classification results made by VADER. According to the confusion matrix given in Table 1, the classification accuracy rate obtained with VADER is 85%. In addition, three classes were used instead of five classes due to the difficulties that may be experienced in perceiving the differences between certain classes in determining the classification performance.

Table 1. Confusion Matrix

		Predicted		
		Neutral	Neutral	Positive
Actual	Neutral	**252**	24	24
	Neutral	20	**275**	5
	Positive	18	48	**234**

Before evaluating the classification results, the Word Cloud given in Figure 2 was used to examine the general structure of the tweets. Word Cloud or Tag Cloud is a data visualization technique that helps to display words in a text or tweet on a chart, where important words are displayed larger and less important words are displayed smaller or not displayed at all. In Natural Language Processing, useful information about large amounts of data to be processed through Word Cloud is obtained and visualized. The word clouds given in Figure 2 were handled separately for each month within the six months examined in the study. Figure 2a was obtained using 211,771 tweets shared with the relevant hashtag in January. As can be seen, the words *"wuhan"*, *"coronavirus"*, *"outbreak"*, *"china"* are frequently used in these tweets. Figure 2b was obtained using 996,748 tweets shared with the relevant hashtag in February. As can be seen, the words *"coronavirus"*, *"covid19"* are frequently used in these tweets. Figure 2c was obtained using 18,155,235 tweets shared with the relevant hashtag in March. As can be seen, the words *"social distance","coronavirus"*, *"covid19"*, are frequently used in these tweets. Figure 2d was obtained using 15,435,085 tweets shared with the relevant hashtag in April. As can be seen, the words *"social distance"," coronavirus"*, *"covid19"*, *"stay home"* are frequently used in these tweets. Figure 2e was obtained using 11,059,289 tweets shared with the relevant hashtag in May. As can be seen, the words *"social distance"," coronavirus"*, *"covid19"*, *"practice social"* are frequently used in these tweets. Figure 2f was obtained using 6,813,248 tweets shared with the relevant hashtag in June. As can be seen, the words *"social distance"," coronavirus"*, *"covid19"*, *"wear mask"* are frequently used in these tweets.

Figure 2. Word Cloud Representations

An N-gram is an N-character segment of a longer text. Generally divides the string into a series of overlapping N-grams (Cavnar & Trenkle, 1994: p. 3). In this study, 2-grams, 3-grams, and 4-grams were used on a word basis for each tweet obtained. N-grams given in Table 2 were handled separately for each month within the six months examined in the study. From 2-grams, the words *"coronavirus, outbreak"*, *"coronavirus, coronavirus"*, *"wuhan, coronavirus"* were frequently used in tweets in January, while the words *"coronavirus, covid19"*, *"covid19, coronavirus"* began to be used more frequently in February, especially with the definition of the virus and its spread in the world. After March, the words *"social, distance"*, *"stay, home"* have started to be used more frequently, especially with restriction applications. In May and June, the words *"social, distance" and "wear, mask"* were used more together with reopening. Similarly, from 3-grams, the words *"public, health, emergency"*, *"coronavirus, sars, flu"*, *"world, health, organization"* attracted attention at the beginning of the outbreak, and in the following time *"practice, social, distance"*, *"mask, social, distance"*, *"maintain, social, distance"* were used more together. From 4-grams, the results of *"public, health, emergency, international" in January*, *"virus, coronavirus, sars, flu"* in February, *"copy, official, last, deliver"*, *"boris, johnson, test, positive"* in March, *"frontline, keep, safe, coronavirus"*, *"stay, home, stay, safe"* in April and "*social, distance, wear, mask"* in May and June are quite remarkable in terms of the course of the outbreak.

Table 2: Top ten N-grams by months

2-gram	3-gram	4-gram
January		
coronavirus, outbreak	public, health, emergency	public, health, emergency, international
coronavirus, coronavirus	global, health, emergency	health, emergency, international, concern
china, coronavirus	confirm, case, coronavirus	declare, global, health, emergency
wuhan, coronavirus	world, health, organization	coronavirus, wuhanflu, 2019ncov, ncov2019
coronavirus, china	coronavirus, death, toll	coronavirus, death, toll, rise
case, coronavirus	first, case, coronavirus	world, health, organization, declare
novel, coronavirus	death, toll, rise	declare, public, health, emergency
confirm, case	emergency, international, concern	coronavirus, global, health, emergency
new, coronavirus	coronavirus, outbreak, china	coronavirus, public, health, emergency
coronavirus, case	health, emergency, international	wuhan, citizen, plainly, tell
February		
coronavirus, covid19	coronavirus, sars, flu	virus, coronavirus, sars, flu
covid19, coronavirus	virus, coronavirus, sars	see, detail, virus, coronavirus
coronavirus, outbreak	detail, virus, coronavirus	detail, virus, coronavirus, sars
covid, 19	see, detail, virus	coronavirus, sars, flu, china
novel, coronavirus	test, positive, coronavirus	coronavirus, sars, flu, deathtoll
coronavirus, case	world, health, organization	sars, flu, deathtoll, china
coronavirus, coronavirus	new, coronavirus, case	diamond, princess, cruise, ship
confirm, case	sars, flu, china	coronavirus, disease, 2019, covid19
case, coronavirus	new, case, coronavirus	sars, flu, china, trump
china, coronavirus	sars, flu, deathtoll	survive, see, detail, virus
March		
social, distance	practice, social, distance	copy, official, last, deliver
covid, 19	test, positive, covid19	deliver, copy, official, last
coronavirus, covid19	test, positive, coronavirus	stay, home, stay, safe
stay, home	maintain, social, distance	boris, johnson, test, positive
covid19, coronavirus	call, chinese, virus	social, distance, stay, home
covid19, pandemic	covid19, social, distance	act, deliver, copy, official
chinese, virus	confirm, case, covid19	must, act, deliver, copy
test, positive	stay, home, stay	sign, must, act, deliver
practice, social	social, distance, stay	prince, charles, test, positive
due, covid19	copy, official, last	prime, minister, boris, johnson

April		
social, distance	practice, social, distance	frontline, keep, safe, coronavirus
covid, 19	test, positive, covid19	provide, frontline, keep, safe
covid19, pandemic	maintain, social, distance	stay, home, stay, safe
coronavirus, covid19	social, distance, rule	uk, govt, provide, frontline
covid19, coronavirus	follow, social, distance	copy, official, last, deliver
covid19, case	social, distance, measure	deliver, copy, official, last
stay, home	new, covid19, case	safe, coronavirus, sign, petition
fight, covid19	social, distance, guideline	keep, safe, coronavirus, sign
due, covid19	fight, covid, 19	safe, enforce, mask, usage
covid19, crisis	covid, 19, pandemic	keep, safe, enforce, mask
May		
social, distance	practice, social, distance	social, distance, wear, mask
covid, 19	test, positive, covid19	wear, mask, social, distance
covid19, pandemic	mask, social, distance	pandemic, news, coronavirus, covid19
covid19, case	social, distance, rule	global, pandemic, news, coronavirus
coronavirus, covid19	maintain, social, distance	bring, total, confirm, case
covid19, coronavirus	new, covid19, case	follow, social, distance, guideline
due, covid19	social, distance, measure	mask, practice, social, distance
wear, mask	follow, social, distance	stay, home, stay, safe
covid19, crisis	social, distance, guideline	new, death, bring, total
covid19, test	new, case, covid19	social, distance, stay, home
June		
social, distance	test, positive, covid19	social, distance, wear, mask
covid, 19	mask, social, distance	wear, mask, social, distance
covid19, pandemic	new, covid19, case	break, social, distance, rule
covid19, case	practice, social, distance	mask, practice, social, distance
due, covid19	social, distance, rule	pandemic, news, coronavirus, covid19
wear, mask	maintain, social, distance	global, pandemic, news, coronavirus
coronavirus, covid19	social, distance, measure	report, today, utc, time
new, case	social, distance, mask	bring, total, confirm, case
covid19, coronavirus	new, case, covid19	wear, mask, practice, social
test, positive	follow, social, distance	social, distance, measure, place

In this study, we use the VADER to classify the sentiments expressed in Twitter data related to Covid-19. The compound scores of tweets were grouped into five categories: Highly Positive, Positive, Neutral, Negative, Highly Negative.

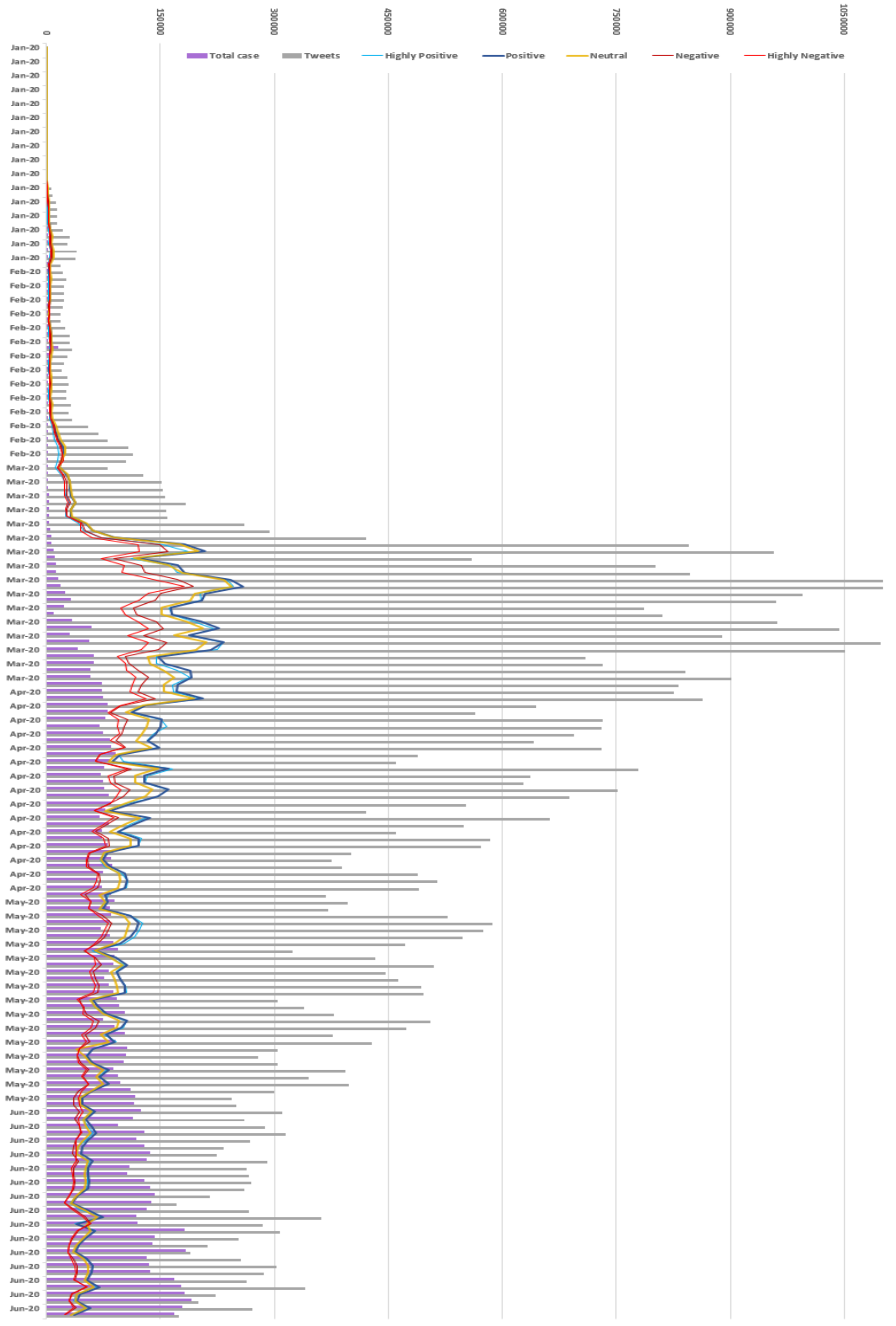

Figure 3: Word Cloud Representations

Figure 3 shows the daily classification results of tweets obtained for 6 months within the scope of the study. It also shows the daily number of confirmed cases and the total number of tweets as well as the classification results. As a result of the classification carried out on approximately 52 million tweets, 22% of the total tweets are high positive, 23% of the total tweets are positive, 22% of the total tweets are neutral, 17% of the total tweets are negative and 15% of the total tweets are high negative. The values shown in Figure 3 can be accessed in detail in Appendix 1. As a result of the findings of this research, contrary to expectations, although the number of positive tweets is high in both total and daily tweets, this situation is not related to the performance of VADER, which is a lexicon-based method. Besides that, the 85% accuracy rate obtained with the manual evaluation made within the scope of the research supports this situation. Although tweets with negative emotional polarity were more on a daily basis in the initial phase of the period examined in the study, this situation reversed in the later stages of the pandemic and it was understood as a result of the classification that tweets with more positive emotional polarity on a daily basis. In particular, with a general review of positive tweets, as the main reasons for the increase in these positive tweets; possible vaccine studies, better understanding of the uncertainties at the beginning of the pandemic, the development of effective solutions against the virus, the decrease in the number of cases, leaving the first pandemic wave behind and a positive change in the perspective on the future of the pandemic can be listed.

As seen in Figure 3 and Appendix 1, while the negative tweets of the first stages of the outbreak were more intense, positive tweets became more intense, especially after the month of March, when the Covid-19, virus spread all over the world. The reason for the increase in these positive tweets is that people get used to the fight against the coronavirus and the tweets about how to overcome this process with the least possible loss, rather than the negative effects of it. In addition, with the increase in the measures taken against the virus, the exponential increase in the spread of the virus has decreased, and people have more positive feelings towards the passing of the pandemic. Besides that, as experts have stated, Covid-19, a SARS-based virus, has positively affected the perspectives of individuals in the society towards the course of the pandemic with the arrival of summer, thanks to more ventilation of indoor environments, the decrease in the rate of Covid-19 transmission and the flexibility in restrictions. Also, this situation presents interesting findings for the examination of epidemiology, sociology and psychology.

Especially in many countries, in March and April, when various restrictions and lockdown were experienced, the number of tweets posted with the hashtag determined within the scope of the study is quite high. It is also seen that the tweets posted in this period contain more positive emotions. This result obtained in this context is quite striking. In the last period of the 6-month period examined, we see that although the increase in the number of confirmed cases continues, the number of tweets about Covid-19 has decreased and the tweet emotions are about the same daily.

As can be seen from Table 2 and Figure 3, it is observed that the perception towards the Covid-19 virus in the society changes as the pandemic period progresses. In particular, the positive developments in vaccine studies and the reduction in the number of periodic cases and the relaxation in restrictions have caused a change in the perspective of the society towards Covid-19.

The results obtained as a consequence of this study will help policy makers to control and manage social perception during the Covid-19 pandemic process. At the same time, it can be used as a tool to take the pulse of the society, as well as to determine the action plans and restrictions that can be taken regarding the pandemic. In addition, the findings obtained within the scope of this study constitute a secondary data source for the use of a wide variety of disciplines, examining many social and individual factors throughout the pandemic process.

CONCLUSION

In this study, it was aimed to investigate the attitudes of people on the outbreak by analyzing the sentimental of the English tweets obtained from all over the world in the first 6 months of the Covid-19 outbreak. Especially considering that the emotional state of people during the outbreak varies from day to day, the analysis of all the tweets obtained may not reflect the correct results. It is aimed to determine the changes in the relevant period by classifying the tweets obtained from Twitter on a daily basis. There is no such comprehensive sentimental and emotion analysis study in the literature on the Covid-19 outbreak, including other outbreaks in the past. In this study, a better understanding of people emotions was provided during the outbreak by examining especially for a wide period.

When the results obtained in the study were examined, while there were more negative tweets about Covid-19 in the early stages of the outbreak, more positive tweets were shared by people with the later stages of the outbreak. This result obtained in this context is quite striking. In addition, it is seen that in the period under review, people shared less tweets about the Covid-19 outbreak and their interest in this issue decreased. In March and April, when the outbreak intensified, the number of tweets posted with the hashtag determined within the scope of the study is quite high, and the tweets posted during the period contain more positive emotions and this reflects a striking situation.

10.5824/ajite.2022.02.001.x

REFERENCES

Ahmed, M. E., Rabin, M. R. I., & Chowdhury, F. N. (2020). COVID-19: Social Media Sentiment Analysis on Reopening. *arXiv preprint arXiv:2006.00804*.

Andrade, Francisca Marli Rodrigues de and Barreto, Tarssio Brito and Herrera-Feligreras, Andrés and Ugolini, Andrea and Lu, Yu-Ting, (2020). Twitter in Brazil: Discourses on China in Times of Coronavirus. *Available at SSRN: https://ssrn.com/abstract=3608566 or http://dx.doi.org/10.2139/ssrn.3608566*

C. Kaur and A. Sharma, (2020) "EasyChair Preprint Twitter Sentiment Analysis on Coronavirus using Textblob".

Cavnar, W. B., & Trenkle, J. M. (1994, April). *N-gram-based text categorization*. In Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval (Vol. 161175).

Chauhan, V. K., Bansal, A., & Goel, A. (2018). Twitter sentiment analysis using vader. International *Journal of Advance Research, Ideas and Innovations in Technology (IJARIIT*), 4(1), 485-489.

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS one*, 5(11), e14118.

Dubey, A. D. (2020). Decoding the Twitter Sentiments towards the Leadership in the times of COVID-19: A Case of USA and India. *Available at SSRN 3588623*.

Elbagir, S., & Yang, J. (2019). *Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment*. In Proceedings of the International MultiConference of Engineers and Computer Scientists (pp. 122-16).

Fung, I. C. H., Fu, K. W., Ying, Y., Schaible, B., Hao, Y., Chan, C. H., & Tse, Z. T. H. (2013). Chinese social media reaction to the MERS-CoV and avian influenza A (H7N9) outbreaks. *Infectious diseases of poverty*, 2(1), 31.

Hutto, C. J., & Gilbert, E. (2014). *Vader: A parsimonious rule-based model for sentiment analysis of social media text*. In Eighth international AAAI conference on weblogs and social media.

Kim, E. H. J., Jeong, Y. K., Kim, Y., Kang, K. Y., & Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science*, 42(6), 763-781.

Kruspe, A., Haeberle, M., & Zhu, X. X. (2020). Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic

McNeill, A., Harris, P. R., & Briggs, P. (2016). Twitter influence on UK vaccination and antiviral uptake during the 2009 H1N1 pandemic. *Frontiers in public health*, 4, 26.

Pokharel, B. P. (2020). Twitter Sentiment analysis during COVID-19 Outbreak in Nepal. *Available at SSRN: https://ssrn.com/abstract=3624719 or http://dx.doi.org/10.2139/ssrn.3624719*

Prabhakar Kaila, D., & Prasad, D. A. (2020). Informational flow on Twitter–Corona virus outbreak–topic modelling approach. *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 11(3).

Ramteke, J., Shah, S., Godhia, D., & Shaikh, A. (2016, August). *Election result prediction using Twitter sentiment analysis.* In 2016 international conference on inventive computation technologies (ICICT) (1), 1-5. IEEE.

Shin, S. Y., Seo, D. W., An, J., Kwak, H., Kim, S. H., Gwack, J., & Jo, M. W. (2016). High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. *Scientific reports*, 6, 32920.

Van Lent, L. G., Sungur, H., Kunneman, F. A., Van De Velde, B., & Das, E. (2017). Too far to care? Measuring public attention and fear for Ebola using Twitter. *Journal of medical Internet research*, 19(6), e193.

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., ... & Zhao, Y. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *Jama*, 323(11), 1061-1069.

World Health Organization (WHO). "WHO Coronavirus Disease (COVID-19) Dashboard". Available at https://covid19.who.int/?gclid=Cj0KCQjw0rr4BRCtARIsAB0_48O8MNIGaRGaCcjx RCLkiPW 6kidbaM4Fb_xGsU9E7HOnqrjtx8_bLogaAgKVEALw_wcB

World Health Organization. (2020). Coronavirus disease 2019 (COVID-19) situation report–51. Geneva, Switzerland: World Health Organization. https://www.who.int/docs/default-source/coronaviruse/situationreports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10

Zheng, Y. Y., Ma, Y. T., Zhang, J. Y., & Xie, X. (2020). COVID-19 and the cardiovascular system. *Nature Reviews Cardiology*, 17(5), 259-260.

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Chen, H. D. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. nature, 579(7798), 270-273.

APPENDIX

	Confirmed Case	Tweets	Highly Pos.	Positive	Neutral	Negative	Highly Neg.
01.01.20	0	14	3	5	3	3	0
02.01.20	0	15	4	7	3	1	0
03.01.20	0	28	3	10	8	4	3
04.01.20	0	17	5	5	3	2	2
05.01.20	0	27	4	11	3	3	6
06.01.20	0	25	5	2	8	4	6
07.01.20	0	12	1	1	4	4	2
08.01.20	0	69	20	17	16	13	3
09.01.20	0	135	17	25	49	29	15
10.01.20	0	159	18	23	28	68	22
11.01.20	41	188	18	23	28	68	22
12.01.20	0	128	16	32	29	20	9
13.01.20	1	150	20	43	38	23	12
14.01.20	1	165	19	41	47	39	11
15.01.20	0	148	18	33	45	32	15
16.01.20	0	394	31	78	124	68	25
17.01.20	5	695	57	106	195	94	67
18.01.20	17	1243	57	106	195	94	67
19.01.20	60	500	25	85	150	63	35
20.01.20	78	2020	161	407	589	320	268
21.01.20	93	5865	556	1070	1772	1038	839
22.01.20	154	8570	804	1671	2414	1621	1436
23.01.20	133	12021	1188	2323	3305	2322	1890
24.01.20	271	13968	1338	2721	4022	2452	2219
25.01.20	469	14370	1624	2942	3821	2548	2307
26.01.20	689	14360	1644	2746	3597	2550	2539
27.01.20	785	21244	2448	4228	5384	3884	3562
28.01.20	1789	31280	3910	6337	8029	5641	4871
29.01.20	1481	27935	3356	5573	7159	5322	4260
30.01.20	1760	39164	4969	7564	8833	7665	6351
31.01.20	2010	37733	4873	7357	8852	7244	6015
01.02.20	2115	19045	2300	3570	4462	3550	3312
02.02.20	2598	21714	2645	3946	4706	3775	4112
03.02.20	2832	26332	3344	5052	5950	4752	4380
04.02.20	3258	23588	2758	4380	5104	4312	4315
05.02.20	3914	23439	3073	4482	5319	4193	3709
06.02.20	3721	22833	2878	4199	4764	4143	4229
07.02.20	3202	21637	2104	3428	3943	3253	3880
08.02.20	3413	19186	2104	3428	3943	3523	3880
09.02.20	2669	18347	2220	3449	3718	3239	3403
10.02.20	3055	25122	3130	4756	5395	4442	4313
11.02.20	2486	30072	3298	5826	6909	5135	4792
12.02.20	2065	31341	3782	6008	7040	5487	4737
13.02.20	15213	34129	4173	6794	7982	5856	5144
14.02.20	4068	27037	3522	5111	6231	4885	4196
15.02.20	2732	23761	2792	4286	4824	3696	4007
16.02.20	2090	20786	2606	3777	4456	3548	3510
17.02.20	2161	26928	3448	5185	6126	4699	4238
18.02.20	1993	29720	3699	5515	6364	5421	4895
19.02.20	1856	26207	3293	5055	5773	4679	4105
20.02.20	486	25772	3134	4848	5477	4396	3941
21.02.20	1044	32438	3855	5840	7857	5275	4791
22.02.20	1139	29469	3577	5549	6943	5065	4586
23.02.20	1016	33782	4240	6181	7308	6099	5526
24.02.20	624	54552	6653	10055	12330	9767	8794
25.02.20	826	67769	8735	12874	15381	12200	10841
26.02.20	924	81114	10691	15747	18874	14647	13575
27.02.20	1369	107271	15024	21246	24001	19140	18315
28.02.20	1448	113812	16567	22854	24125	20865	20080
29.02.20	1826	104952	14283	19923	20607	19193	22352
01.03.20	1891	81061	12014	16283	16962	14115	14920
02.03.20	2281	127120	19659	25807	27297	21851	21103
03.03.20	2332	152071	23911	30972	30972	26418	24305
04.03.20	2200	153563	24007	31021	31430	26214	23836
05.03.20	2777	156109	24860	31885	33340	26644	23486
06.03.20	3800	183659	29514	37849	38803	31366	27822
07.03.20	3926	157591	25383	31564	31625	27534	26028
08.03.20	3738	159048	25961	32270	32921	26995	27245
09.03.20	4047	259988	41438	52419	52756	46184	44749
10.03.20	4488	293543	49100	61176	60458	51337	45438
11.03.20	6525	420047	72504	89138	87562	72339	60744
12.03.20	6519	845341	153927	181210	176046	148794	120024
13.03.20	9643	956439	186678	209211	199653	159615	122484
14.03.20	11512	559052	111122	120189	114689	88695	72892
15.03.20	11853	801977	164271	173219	160488	124841	102236
16.03.20	11791	846995	171444	181787	178491	130029	99952
17.03.20	14824	1122432	236575	241872	233123	171973	140623
18.03.20	18429	1240137	246668	258247	244095	193063	181508
19.03.20	23977	995221	202880	208549	195591	151533	134862
20.03.20	32892	959817	204373	204216	189478	142981	122420
21.03.20	23448	787080	160852	163674	151200	114195	97988
22.03.20	9336	810054	165854	165502	151992	119774	103327
23.03.20	33722	962267	194040	202645	183734	145000	120456
24.03.20	59957	1042958	220515	226376	206318	154504	133951
25.03.20	30532	888865	188628	188359	168600	127858	106784

	Confirmed Case	Tweets	Highly Pos.	Positive	Neutral	Negative	Highly Neg.
01.04.20	73694	830984	166517	173092	154420	124349	112880
02.04.20	73283	825209	167881	171636	154136	120498	109983
03.04.20	75205	863115	204401	206650	193176	143571	130574
04.04.20	80057	643896	124461	131074	126806	97632	96199
05.04.20	79829	564475	111123	112372	104089	83959	81223
06.04.20	76987	731639	150458	151877	135522	106337	95027
07.04.20	69341	729698	158268	149370	131915	102165	93973
08.04.20	74190	693952	141895	143286	125927	98646	96370
09.04.20	83271	641361	132214	133277	118196	91669	84931
10.04.20	85056	730896	148392	147920	138429	104251	101819
11.04.20	91074	488284	94322	97282	90843	70098	70685
12.04.20	84045	460517	101530	86364	82676	64429	65222
13.04.20	76022	689437	165816	159944	146405	110962	106298
14.04.20	71373	547104	132134	128476	116404	88572	81514
15.04.20	74320	550346	129960	129146	116225	90334	84672
16.04.20	75719	66682	159284	160292	139780	109363	98089
17.04.20	82277	614976	145570	146755	129665	100957	92009
18.04.20	84396	485482	105867	110350	101303	83274	84680
19.04.20	78150	366561	79352	83922	77948	62642	62693
20.04.20	69398	576643	132997	136420	122653	95263	89301
21.04.20	82669	481510	110523	113984	101798	80196	75002
22.04.20	72500	398524	96093	94345	84148	64174	59758
23.04.20	72579	512143	124589	120786	111086	80724	74948
24.04.20	80256	511566	119838	120672	110972	82038	78041
25.04.20	88606	344980	77182	80763	75298	56401	55334
26.04.20	84576	323226	71285	74541	71661	53639	52097
27.04.20	86636	346590	76049	82940	79700	55741	52155
28.04.20	73991	433572	99331	102564	95328	69189	67144
29.04.20	66289	447696	106722	106936	97454	70329	66244
30.04.20	72942	432430	105361	103421	93722	67610	62310
01.05.20	84868	323097	80707	76955	70097	50296	45036
02.05.20	90270	354771	78377	81058	78188	57918	59218
03.05.20	83738	325072	71108	73705	69974	54819	55454
04.05.20	85681	463112	109353	110515	101756	74244	67235
05.05.20	83206	524229	126129	121545	109491	85710	81339
06.05.20	70911	503970	122310	117976	106570	80990	76116
07.05.20	84140	478581	115562	111403	102996	76656	71961
08.05.20	88018	416772	101064	96780	89520	65954	63446
09.05.20	94669	287913	60499	64003	63369	48825	51211
10.05.20	90130	386221	83962	90669	83221	65306	63053
11.05.20	88301	445733	102356	106860	99423	71979	65110
12.05.20	81508	386337	92004	92060	84545	60975	56751
13.05.20	75680	407220	96144	97624	88974	64790	59686
14.05.20	82643	433066	104080	103400	92636	68640	64302
15.05.20	88501	432282	104850	103568	94352	68186	61322
16.05.20	92994	260720	58316	60533	58030	42798	41040
17.05.20	95749	293928	62974	67910	64622	48408	50010
18.05.20	102784	331958	77708	79576	73990	52422	48256
19.05.20	74721	436334	104930	105614	95106	68800	61878
20.05.20	90264	414268	100630	99358	92310	64022	57948
21.05.20	102514	324609	76223	78301	71894	51276	46909
22.05.20	91054	373168	91828	89258	82334	57088	52658
23.05.20	106455	267770	61146	60546	40268	43946	43858
24.05.20	104502	239076	52606	53942	51019	40306	41201
25.05.20	102271	264635	58863	60930	57160	44987	42689
26.05.20	88785	342742	79084	81859	74577	55399	51818
27.05.20	94329	299806	69047	70628	65532	47833	46760
28.05.20	96623	348282	81624	81862	74310	55134	55347
29.05.20	110826	268179	58983	60486	59241	42609	46854
30.05.20	116737	211098	43173	46799	44036	35387	41701
31.05.20	116020	214340	42721	48005	43537	36325	43748
01.06.20	124050	275571	61244	63510	58702	43880	48231
02.06.20	113209	228352	47601	52345	48600	37575	42229
03.06.20	93385	251693	54767	58677	51220	42921	44102
04.06.20	129451	277563	62365	65298	57907	45561	46425
05.06.20	118853	233735	51851	54710	48725	38778	39671
06.06.20	128122	204716	41487	46644	41544	36385	38653
07.06.20	136589	197612	39376	45198	40265	34201	38562
08.06.20	131201	255378	57576	61527	55476	41580	39210
09.06.20	108952	229201	52981	55112	51345	36118	33641
10.06.20	105758	231303	53495	55326	51003	36352	35120
11.06.20	128489	234386	53477	56152	51074	37898	35781
12.06.20	136524	227257	51945	55019	49575	36328	34388
13.06.20	142515	181389	40595	41293	39941	29799	29758
14.06.20	137519	144536	30741	33200	32067	24140	24386
15.06.20	132268	230589	43402	54930	49905	37815	34432
16.06.20	118496	313270	73329	75194	66187	50906	47649
17.06.20	120062	245479	37188	39543	53731	58380	56634
18.06.20	181243	267242	63975	63419	57916	41494	40434
19.06.20	142971	220518	51709	52375	48275	35067	33090
20.06.20	138779	181930	41413	42221	39827	29863	29601
21.06.20	183603	164721	35542	37660	34791	27938	28787
22.06.20	132155	222279	51808	53200	49370	35441	32456
23.06.20	134491	255635	61343	60737	55148	40872	37533
24.06.20	136019	250750	58001	59304	54434	40782	38227

26.03.20	55727	1097029	232159	233169	210380	157824	133566	**25.06.20**	168018	230133	52847	53842	49602	37362	36478
27.03.20	41604	1050532	224401	216696	196374	148175	125447	**26.06.20**	177943	305160	66233	69617	61937	53177	54192
28.03.20	62791	709445	144443	147169	134019	104044	92749	**27.06.20**	181502	192334	40526	44221	40356	33099	34129
29.03.20	61740	731933	144123	156437	136176	107925	104092	**28.06.20**	190482	172043	35172	39064	37835	29556	30414
30.03.20	57692	840272	172695	189722	154791	121158	105937	**29.06.20**	178979	236061	53067	57266	51464	38344	35918
31.03.20	58087	901792	188272	190496	168852	134385	116997	**30.06.20**	168859	152412	34783	36097	32283	25070	24176