Sentiment Analysis and Machine Learning Classification of COVID-19 Vaccine Tweets: Vaccination in the Shadow of Fear-trust Dilemma

Samet Tüzemen[1], Özge Barış-Tüzemen[2] and Ali Kemal Çelik*[1]
1 Department of Business Administration, Ardahan University Çamlıçatak Ardahan, Türkiye
2 Department of Econometrics, Karadeniz Technical University Kalkınma Trabzon, Türkiye
E-mail: samettuzemen@ardahan.edu.tr, ozgebariss@gmail.com, alikemalcelik@ardahan.edu.tr
* Corresponding author

Keywords: COVID-19 vaccine, sentiment analysis, machine learning, text mining, twitter

Received: March 4, 2022

In addition to infecting millions of people and causing hundreds of thousands of deaths, COVID-19 has also caused psychological and economic devastation. Studies on the vaccine, which is considered to be the only way to eliminate this pandemic, have been rapidly completed and more than 10 vaccines have begun to be applied worldwide by 2021. One of the biggest obstacles to the fight against COVID-19 is the hesitation against the vaccine. The fear factor, fed by incomplete and false information spreading rapidly through social media applications such as Twitter, is thought to be the main reason for this hesitation. In this study, the general sentiment against the COVID-19 vaccine is analyzed. For this, in the first week of January 2021, more than 8000 tweets are extracted with R statistical software and Twitter API, and appropriate sentiment analysis methods are applied. On the other hand, accuracy values are obtained by applying Logistic Regression and Naïve Bayes methods, which are effective and widely used supervised machine learning methods, for sentiment classification. Although the results indicate that there is a positive attitude about the vaccine, it is remarkable that the rate of negative sentiments is relatively high (30%). Trust is the dominant sentiment on the positive side, while fear is the dominant sentiment on the negative side. According to the results of the classification methods, accuracy values are close to 90%.

Povzetek: Študija obravnava splošno razpoloženje glede cepiva za COVID-19 na Twitterju.

1 Introduction

COVID-19 is a disease caused by the virus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is transmitted from person to person, affecting the respiratory tract. This disease, which emerged with the detection of the first case in Wuhan province of China in December 2019 and spread all over the world in a few months, was declared a pandemic by the World Health Organization (WHO) on March 11th, 2020 and has been the most important agenda item in the world until today. COVID-19 is transmitted by inhaling droplets emitted by sick individuals during a speech or sneezing/coughing. For this reason, it is recommended to pay attention to social distance, use of masks, and cleanliness as a method of protection from disease. Since these recommendations were found to be insufficient to contain the spread of the disease, governments have implemented various advanced measures such as restrictions and closures.

The measures taken by many countries around the world, in the form of travel and gathering bans and lockdowns, have contributed to overcoming the periods in which the spread of the epidemic accelerated, called waves, and ensured the control of the epidemic to a certain extent, as seen in Figure 1. However, the restriction of economic activity and social life in this dimension has created great pressure on individuals and the economies. Especially individuals who are currently trying to cope with the shock effect of a worldwide epidemic have also faced the loss of social interaction. As a result, individuals have started to experience disorders such as stress, anxiety, and depression [1].

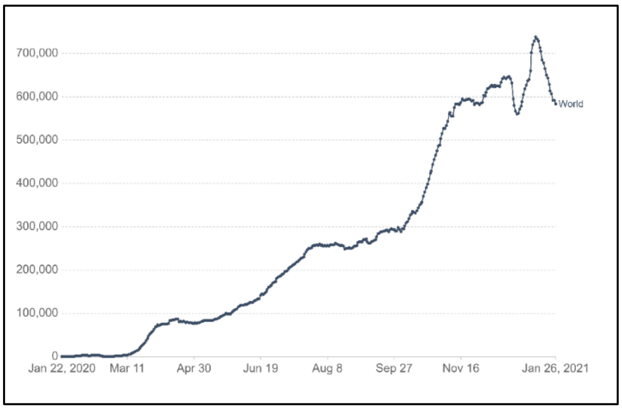

Figure 1: Daily new confirmed COVID-19 cases (world).

On the other hand, these measures had a great impact on macroeconomic indicators such as unemployment and budget balance. This situation caused a sudden decline in the Gross Domestic Product of the countries, causing almost all of them to experience a large decrease. As seen in Figure 2, the shrinkage experienced in GDP has found an average of 4.4%. In addition, there also have been dramatic increases in unemployment rates. All these negativities experienced in economic terms ultimately caused the psychological conditions of individuals to get even worsen [3].

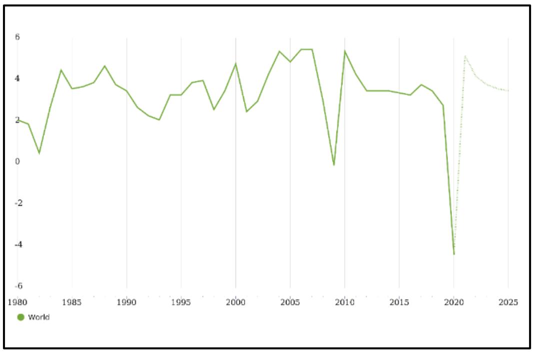

Figure 2: Real GDP growth (Annual %).

Simultaneously with the emergence of the disease, scientists in many parts of the world began working to develop a vaccine that would be effective against the virus. It is known that years, not months, are needed for an effective and safe vaccine to emerge after all procedures are completed. Despite this, a great effort has been made for an effective vaccine that will end the COVID-19 pandemic, and the development process of at least 3 vaccines has been completed before the end of 2020. In some countries such as China, United Kingdom, and Russia, it has even been granted permission to use these vaccines in emergencies. As of January 2021, 10 vaccines have been used by various countries and over 70 million people have been vaccinated. Figure 3 shows the course of vaccination in 10 countries where the most doses are applied.

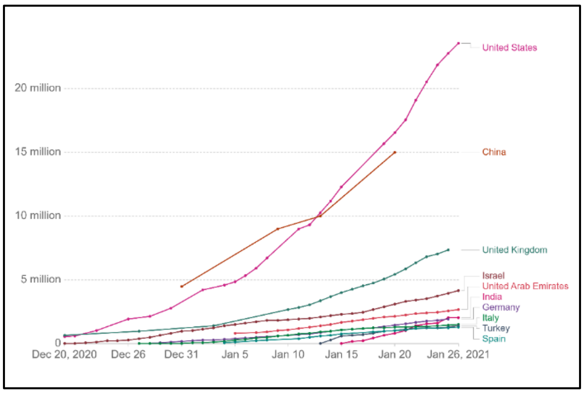

Figure 3: Cumulative COVID-19 vaccines doses administered (highest 10 countries).

Considering the psychological and economic destruction of this pandemic, it is expected that the beginning of the vaccination process, which is likely to end the epidemic, would have a positive effect on people. This virus has infected 100 million people worldwide and caused the death of 2 million people as of January 2021. Despite this phenomenon, the positive attitude towards the vaccine is not high as is expected. Although it varies based on countries, it is observed that there is a remarkable rate of skepticism in the society against the vaccine [5]. The most important factor that triggers this attitude, which is an important obstacle in the effective fight against a pandemic, is the rapid spread of misleading information based on conspiracy theories. Social media has become the primary communication tool that enables information to spread rapidly around the world. However, the accuracy of the aforementioned information cannot always be guaranteed, and this causes information corruption. This situation makes it difficult to manage the perception of society in such an important period. As a result, even the vaccine, which is the world's only hope to end this global crisis, was faced with a significant negative response.

This study aims to reveal the public sentiment against the COVID-19 vaccine as of the first week of 2021 by examining the posts (tweets) from Twitter, which is an important social media tool with a large user base, while the ongoing vaccination activities are given. For this purpose, using the R statistical software, Twitter posts are compiled, and sentiment analysis is performed with the data cleaned with appropriate methods. Then, the efficiency of the established models is examined by applying Logistic Regression and Naïve Bayes Classification methods, which are the most frequently used machine learning methods.

2 Background and literature review

Studies on COVID-19 disease have increased dramatically after spread around the world and declared as a global pandemic by WHO. Not only the medical effects of COVID-19 on sick people, but also the psychological and behavioral effects on the whole society, and even the socio-economic effects on the countries are examined in these studies. Considering the scope and impact of the disease, evaluating these studies independently from each other will prevent one to understand the real dimension of each effect. For this reason, some of the wide-ranging researches are referred and their contribution to this study is examined.

As in every large socio-economic incident, the primary impact of the COVID-19 outbreak has been on the psychology of individuals. Especially, the increasing number of cases and deaths and the restrictions imposed by the governments started to create increasing pressure on the individuals in the society. It is thought that determinants such as education, age, gender, and social status have an important contribution to the extent of this pressure. Accordingly, certain groups are experiencing unemployment and cost of living pressure, while others must cope with concerns such as education and socialization. Psychological problems such as stress, anxiety disorder, and depression accompany this pressure. With the rapid spread of the disease itself, the spread of true and false information about it on social media has increased the extension of individual traumas. Although efforts are made to alleviate social trauma through various methods such as free online group therapies, some researchers argue that the effects of this trauma will extend into the post-pandemic period and only then will its profound effects be understood [1, 6, 7, 8, 9, 10, 11].

Today, people prefer to share their feelings on social media. For this reason, social media applications such as Twitter have become a very large and important data source to measure the feelings of individuals and societies in the face of certain events. In this process, many studies have been conducted using tweets to investigate how people feel about the COVID-19 outbreak. In these studies, which are called sentiment analysis, researchers have applied various machine learning classification methods such as Logistic Regression, Naïve Bayes, Vector Support Machine, and Recurrent Neural Network, which are widely used today. The results vary according to the demoFigureic structure and the measures taken by the governments. However, it is seen that the presence of high polarity in relatively homogeneous groups, the reaction of individuals to an event is generally directly related to their characteristics [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22].

The impact of COVID-19 on the world economy has been devastating. In addition to the restriction of international transportation and trade, because of the lockdown implementations and the dramatic decline in commercial activity, the economies around the world rapidly entered a severe recession. As a natural consequence of this, unemployment rose to record levels and financial difficulties increased significantly. Although developed and rich countries have taken some measures to mitigate the impact of the pandemic on the general economy and individuals, the crisis has deepened in countries that are already experiencing difficulties. Another reason for the psychological disorders that coped with the COVID-19 pandemic is the mentioned unemployment. While most research show that unemployment is an important trigger of stress and depression, it also emphasizes that this situation causes an increase in suicide cases. Again, the researchers concluded that the groups considered as minorities are more vulnerable when considering the factors of race, gender, and age in the face of the above factors [23, 3, 24, 25, 26, 27, 28, 29].

Experts point out the need for herd immunity to end this pandemic and therefore stop the material and moral losses. Accordingly, in order for the epidemic to slow down and disappear, at least 60% of the population must be immunized [30]. This can be done in two ways. First, this proportion of people should get sick. Second, people should be vaccinated at this rate. There have been countries that have tried the first method at the early stage of the pandemic. However, with the realization that the cost of this is too high to be incurred, the second method has become the only hope for the whole world. Especially the shock wave experienced at the beginning of the pandemic provided support for vaccination studies to a large extent.

With the availability of at least a few different vaccines in the last months of 2020, conspiracy theories that spread rapidly on social media emerged. These conspiracy theories have triggered an unsafe environment for the vaccine. In this context, the hesitation against the COVID-19 vaccine is being studied extensively. [31] argue that the current hesitation against vaccination will not disappear in a short time, even with a devastating pandemic such as COVID-19, and this should be tackled at the local level. [32] found that 71.5% of the participants were willing to get the COVID-19 vaccine in their survey in June 2020, with 13,426 people from 19 countries. [33] measured vaccine literacy and attitude against possible COVID-19 vaccine in their survey study for Italy. Again, the results of this study carried out in June 2020, show that there is an 80-90% positive attitude towards the vaccine.

Examining the size of the hesitation against vaccines in China in May 2020, [34] found that 95% of the participants trust the vaccine to be developed in the country and 83% want to get the COVID-19 vaccine when it is ready. [35] conducted a similar study for the USA in May 2020 and as a result, 69% positive response was obtained for the COVID-19 vaccine. On the other hand, [36] revealed in the survey they conducted for the UK in September 2020 that 54% of the participants had a positive approach to the COVID-19 vaccine. [37], who examined the rate of refusal of the COVID-19 vaccine by 5 consecutive survey studies in France between May and October 2020, found that this rate gradually increased. According to the findings of [38], who conducted similar research for Italy, the situation was in line with the results of the previous study. Researchers have stated that there is a decrease in the intention of vaccination between the two stages of the epidemic in Italy and that the proportion of people who intend to be vaccinated is not enough to end the epidemic (pp. 786-787).

Many similar studies have been conducted for many countries such as Finland, Israel, Pakistan, and

Indonesia. According to the findings and the joint opinion of the researchers, the skeptical attitude towards the COVID-19 vaccine is alarming and urgent action should be taken against it [39, 40, 41, 5, 42, 43]. On the other hand, in the survey study conducted for nursing students, who are healthcare professionals of the future, it was revealed that 63% of the participating students intend to get the COVID-19 vaccine [44]. This rate clearly shows that even young people with health education have a skeptical attitude towards vaccination. When the studies are evaluated as a whole, it is seen that the size of the hesitation against the vaccine is at a worrying level. Almost all researchers agree that a proactive method should be followed in solving this problem.

3 Data set and methodology

In this study, public sentiment analysis about the COVID-19 vaccine is examined. The data set used for this purpose is extracted from Twitter with the keywords "coronavirus vaccine" on January 9th, 2021, using the R statistical software and the “rtweet” library. More than 8000 tweets are converted to lowercase letters, freed from repetitions, punctuations, numbers, stop words, URLs, and non-ASCII words, and finally lemmatized in order to make them ready for analysis. Finally, the remaining 7935 tweets have been converted into a term document matrix.

To analyze the general attitude towards COVID-19 vaccines, the sentiment analysis method, which is a frequently used and effective method in big data analytics, is used. Sentiment analysis is defined as the classification of the main idea in a text with the applications of natural language processing and text analytics. Sentiment analysis aims to understand the attitude of the author by detecting the emotion polarization in a text and classifying it as positive, negative, and neutral ([45] pp. 53-54). For this, a dictionary-based emotion score is determined for each word in the text using flexible and open-source programming languages such as R and Python and related packages. Later, this score, determined on the basis of words, is calculated for the whole text. As a result, the text is classified as positive, negative, or neutral [45].

Despite the important advantage of being less complex, the score calculation method is not efficient enough in some cases. A positive sentence with negative score words will be evaluated as negative by this method. On the other hand, the machine learning approach, which automates processes more, is widely used in sentiment classification. In particular, a sentiment classification model is created by training the available data with supervised machine learning methods, and the obtained accuracy values are compared. This comparison is used to determine if the model has been set correctly, or if there is an overfit or underfit issue. Logistic Regression and Naïve Bayes, which are among these supervised machine learning methods, are used in this study.

Logistic regression is a common type of generalized linear models and models the probability of some events occurring as a linear function of a set of predicted values. In other words, the Logistic Regression method tries to estimate the probability of the dependent variable having a certain value instead of estimating the value of the dependent variable. For example, instead of guessing whether a soccer team will beat the round it played, it tries to predict the probability of passing the round. The actual state of the dependent variable is determined by looking at the estimated probability. If the predicted probability is greater than 0.50, the estimate is closer to YES (i.e., to pass the round), otherwise, the failure to pass the round is more probable. Logistic Regression is used only when the dependent variable is a categorical binary (0 or 1, YES or NO, etc.). In this case, these two possibilities are calculated as $P(y_j = 0) = 1 - p_j$ and $P(y_j = 1) = p_j$with the available data. In this case, the linear logistics (logit) model is established as follows ([46] pp. 157-158).

$$\log\left(\frac{p_j}{[1-p_j]}\right) = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \cdots + \beta_n X_{nj}$$

Naïve Bayes is a simple but effective machine learning classification method that uses the Bayes rule based on the assumption of conditional independence of variables. Bayes theory is a method of calculating the probability of event A to occur depending on event B. It is basically formulated as follows:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

where, $P(A)$ and $P(B)$ are the probability of occurrence of events A and B, respectively, $P(B \mid A)$ is the conditional probability of event B to event A, and lastly, $P(A \mid B)$ is the conditional probability of event A to event B. Based on this, the Naïve Bayes classification equation is simply shown as:

$$P(y \mid x_1, \dots, x_j) = \frac{P(x_1, \dots, x_j \mid y)P(y)}{P(x_1, \dots, x_j)}$$

where, $P(y \mid x_1, \dots, x_j)$ is the posterior conditional probability of class (y) to observation values (x_n), $P(x_1, \dots, x_j \mid y)$ is the conditional probability of observation values to class. Finally, while $P(y)$ is the prior probability of the class, $P(x_1, \dots, x_j)$ is called the marginal probability ([47] pp. 279-280).

4 Findings

In this part of the study in which the sentiment analysis about the COVID-19 vaccine is examined, the obtained findings are presented. Accordingly, the frequency distribution of 200 or more words in extracted and cleaned tweets with appropriate methods is presented in Figure 4.

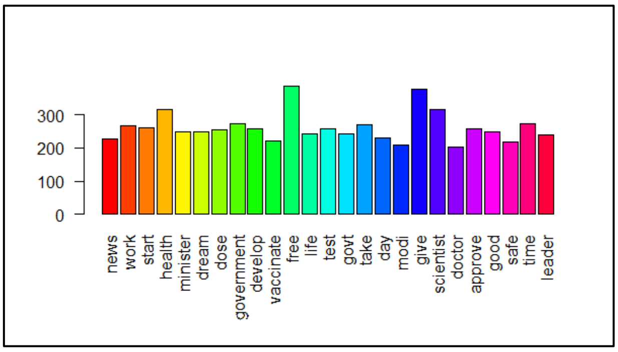

Figure 4: Word frequency of COVID-19 vaccine tweets.

As seen in Figure 4, 4 words namely, free, give, scientist, and health are used more than 300 times. Another important result that is seen from the Figure is that there are words with positive meanings such as good, safe, and approve among the words used more than 200 times, and the absence of words with negative meanings. On the other hand, all the words in the tweets used in the study are presented in Figure 5 in the form of a word cloud.

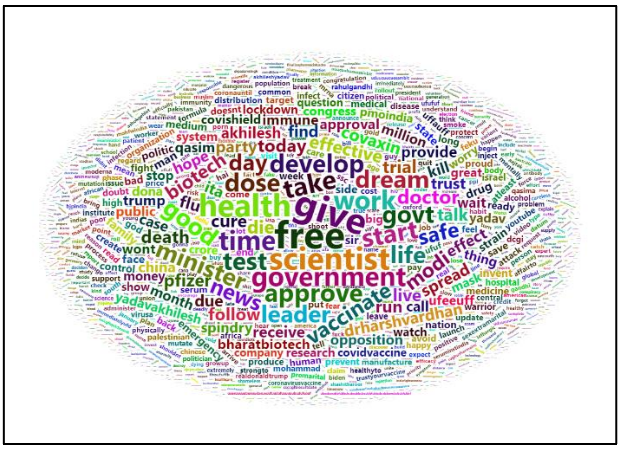

Figure 5: Word cloud of COVID-19 vaccine tweets.

Although it has a lower frequency, there are also negative words such as stop, kill, opposition, and death in tweets, as seen in Figure 5. When evaluated as a whole, the distribution of sentiments in the tweets of this study is shown in Figure 6.

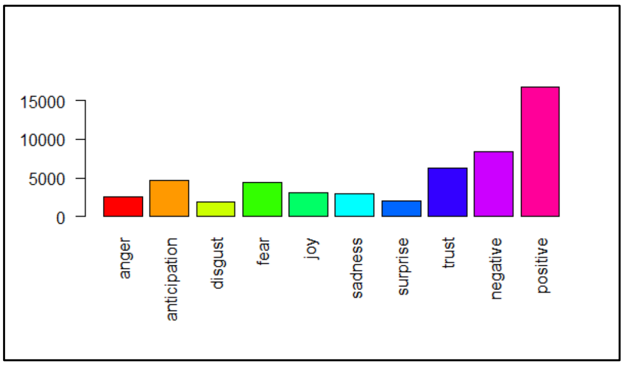

Figure 6: Sentiment frequency distribution of COVID-19 vaccine tweets.

As seen in Figure 6, the strongest sentiment against the COVID-19 vaccine as of the first week of January 2021 is positive. Accordingly, positive sentiments are about twice as much as negative sentiments. With a simple approach, it can be said that approximately 66% of the tweets about the COVID-19 vaccine are positive and approximately 33% are negative. These results coincide with the findings of [35], [36], and [44] examined in the literature section of the study. Although these rates give the impression that there is a positive approach to the vaccine at first glance, it is critically close to 60%, which is required for the immunity rate also known as herd immunity, to end the pandemic. Therefore, it is not wrong to comment that the rate of negative sentiments towards the vaccine is high. On the other hand, it is seen that feelings of trust and hope are dominant on the positive side, and fear is dominant on the negative side.

In order to classify tweets with supervised machine learning methods, the tweets with positive emotion score are marked as 1, and the ones with negative scores are marked as 0. With this marking, the Logistic Regression model is established and applied to the train and test data set separated as 80%-20%. The obtained confusion matrix and accuracy values are presented in Table 1.

Table 1: Confusion matrix and accuracy values for logistic regression.

Train data			**Test data**		
	Actual			*Actual*	
Prediction	**Negative**	**Positive**	**Prediction**	**Negative**	**Positive**
Negative	509	161	**Negative**	129	38
Positive	388	4109	**Positive**	83	1009
Accuracy: ***0,8937***			*Accuracy:* ***0,9039***		

As seen in Table 1, the Logistic Regression model has made the negative and positive classification of the sentiment of tweets with very high accuracy. In addition, the sensitivity value giving the true positive rate is calculated as 0.9637 for the train data set, while the specificity value giving the true negative rate is calculated as 0.6085. On the other hand, the results of the Naïve Bayes model created for emotion classification of tweets about the COVID-19 vaccine are presented in Table 2.

Table 2: Confusion matrix and accuracy values for naïve bayes.

Train data			**Test data**		
	Actual			*Actual*	
Prediction	**Negative**	**Positive**	**Prediction**	**Negative**	**Positive**
Negative	544	254	**Negative**	132	65
Positive	353	4016	**Positive**	80	982
Accuracy: ***0,8825***			*Accuracy:* ***0,8848***		

As seen in Table 2, where the results of the Naïve Bayes classification model are presented, the accuracy values are close to the results of Logistic Regression. The sensitivity value for the train data set is 0.9405 and the specificity value is 0.6065. When the findings are compared with the results obtained from the Logistic Regression model, it is seen that the Logistic Regression classification model is slightly more effective.

5 Conclusion

The aim of this study is to measure and evaluate the attitude towards the COVID-19 vaccine with social media, which has become the most important communication tool today. For this purpose, the tweets about the vaccine are extracted and sentiment analysis about the vaccine is made with various classification methods. For this, on January 9th, 2021, more than 8000 tweets belonging to the previous week are extracted via R statistical software and Twitter API, and the obtained data set is cleaned through appropriate libraries and made ready for analysis. The results of the sentiment analysis and machine learning classification are shared in the findings section of this study.

When the results of the study are evaluated, it is seen that positive sentiments about the COVID-19 vaccine are more than negative ones. Therefore, the vaccine, which is seen as the best possible solution to the major problems caused by the pandemic, is generally accepted. On the other hand, the high rate of negative sentiments is worrisome. Similar to the study conducted by [48], this rate (more than 30%) is an indication that hesitation against vaccination should be evaluated carefully. The results obtained with the classification of sentiments reveal that the most dominant sentiment among negative sentiments is 'fear'. Thus, in order to ensure that the fight against the COVID-19 pandemic is not interrupted and the desired level of immunity is achieved, those in the public decision-making position must take strategic steps to combat fear and the underlying uncertainty. The most basic way of this is to fight against misinformation that spread rapidly, especially in social media, by sharing effective and accurate information.

Logistic Regression and Naïve Bayes supervised machine learning methods are applied to classify tweets about the COVID-19 vaccine and their effectiveness is determined and compared. According to the findings, both methods have very high classification efficiency. However, positive sentiment classification is more successful than negative sentiment classification in both methods. It can be thought that the reason for this is the way the negative feelings are expressed (using "not good" instead of "bad"). In this study, which tries to take a snapshot of the attitude towards the vaccine in terms of its results, it is recommended to examine the sentiment locally in future studies on this subject and to investigate how certain practices or developments affect the attitude towards the vaccine instantly.

References

[1] Pillay, A. L., & Barnes, B. R. Psychology and COVID-19: impacts, themes and way forward. *South African Journal of Psychology*, 50(2), 148–153, 2020. https://doi.org/10.1177/0081246320937684

[2] Ourworldindata.org. Retrieved from https://ourworldindata.org/covid-cases?country=~OWID_WRL (Accessed: 27.01.2021)

[4] IMF World Economic Outlook (October 2020). Retrieved from https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/WEOWORLD (Accessed: 21.01.2021)

[3] Achdut, N., & Refaeli, T. Unemployment and Psychological Distress among Young People during the COVID-19 Pandemic: Psychological Resources and Risk Factors. *International Journal of Environmental Research and Public Health*, 17(19), 7163, 2020. http://dx.doi.org/10.3390/ijerph17197163

[5] Sallam, M. COVID-19 vaccine hesitancy worldwide: a systematic review of vaccine acceptance rates. *medRxiv*, 2021. https://doi.org/10.1101/2020.12.28.20248950

[6] Marmarosh, C. L., Forsyth, D. R., Strauss, B., & Burlingame, G. M. The psychology of the COVID-19 pandemic: A group-level perspective. *Group Dynamics: Theory, Research, and Practice*, 24(3), 122-138, 2020. http://dx.doi.org/10.1037/gdn0000142

[7] Atalan, A. Is the lockdown important to prevent the COVID-19 pandemic? Effects on psychology,

environment and economy-perspective. *Annals of Medicine and Surgery*, 56, 38-42, 2020. https://doi.org/10.1016/j.amsu.2020.06.010

[8] Akat, M., & Karatas, K. Psychological Effects of COVID-19 Pandemic on Society and Its Reflections on Education. *Turkish Studies*, 15, 1-13, 2020. https://doi.org/10.7827/TurkishStudies.44336

[9] Li, S., Wang, Y., Xue, J., Zhao, N., & Zhu, T. The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. *International journal of environmental research and public health*, 17(6), 2032, 2020. https://doi.org/10.3390/ijerph17062032

[10] Aiello, L. M., Quercia, D., Zhou, K., Constantinides, M., Scepanovic, S., & Joglekar, S. How Epidemic Psychology Works on social media: Evolution of responses to the COVID-19 pandemic. *ArXiv*, abs/2007.13169, 2020.

[11] Susič, D., Tomšič, J., & Gams, M. Ranking Effectiveness of Non-Pharmaceutical Interventions Against COVID-19: A Review. *Informatica*, 46(4), 2022. https://doi.org/10.31449/inf.v46i4.4181

[12] Tiwari, P., Pandey, H. M., Khamparia, A., & Kumar, S. Twitter-based Opinion Mining for Flight Service Utilizing Machine Learning. *Informatica*, 43(3), 381-386, 2019. https://doi.org/10.31449/inf.v43i3.2615

[13] Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing*, 97, Part A, 2020. https://doi.org/10.1016/j.asoc.2020.106754

[14] de Las Heras-Pedrosa, C., Sánchez-Núñez, P., & Peláez, J. I. Sentiment Analysis and Emotion Understanding during the COVID-19 Pandemic in Spain and Its Impact on Digital Ecosystems. *International journal of environmental research and public health*, 17(15), 5542, 2020. https://doi.org/10.3390/ijerph17155542

[15] Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets. *IEEE Access*, 8, 181074-181090, 2020. https://doi.org/10.1109/ACCESS.2020.3027350

[16] Barkur, G., Vibha, & Kamath, G. B. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian journal of psychiatry*, 51, 102089, 2020. https://doi.org/10.1016/j.ajp.2020.102089

[17] Samuel, J., Ali, G. G. M. N., Rahman, M. M., Esawi, E., & Samuel, Y. COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. *Information*, 11(6), 314, 2020. http://dx.doi.org/10.3390/info11060314

[18] Nemes L., & Kiss, A. Social media sentiment analysis based on COVID-19, *Journal of Information and Telecommunication*, 2020. http://dx.doi.org/10.1080/24751839.2020.1790793

[19] Zhou, J., Yang, S., Xiao, C., & Chen, F. Examination of community sentiment dynamics due to covid-19 pandemic: a case study from Australia. *ArXiv*, abs/2006.12185, 2020.

[20] Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *Journal of medical Internet research*, 22(4), e19016, 2020. https://doi.org/10.2196/19016

[21] Samant, S. S., Murthy, N. B., & Malapati, A. Categorization of event clusters from twitter using term weighting schemes. *Informatica*, 45(3), 2021. https://doi.org/10.31449/inf.v45i3.3063

[22] Chawla, S., & Mehrotra, M. Impact of emotions in social media content diffusion. *Informatica*, 45(6), 2021. https://doi.org/10.31449/inf.v45i6.3575

[23] OECD (2020). Economic Outlook No 108. Retrieved from https://stats.oecd.org/Index.aspx?DataSetCode=EO# (Accessed: 17.01.2021)

[24] Kong, E., & Prinz, D. Disentangling policy effects using proxy data: Which shutdown policies affected unemployment during the COVID-19 pandemic? *Journal of Public Economics*, 189, 104257, 2020. https://doi.org/10.1016/j.jpubeco.2020.104257

[25] Kawohl, W., & Nordt, C. COVID-19, unemployment, and suicide. *The Lancet Psychiatry*, 7(5), 389-390, 2020. https://doi.org/10.1016/S2215-0366(20)30141-3

[26] Bauer, A., & Weber, E. COVID-19: how much unemployment was caused by the shutdown in Germany? *Applied Economics Letters*, 2020. https://doi.org/10.1080/13504851.2020.1789544

[27] Fairlie, R. W., Couch, K., & Xu, H. The Impacts of COVID-19 on Minority Unemployment: First Evidence from April 2020 CPS Microdata. *National Bureau of Economic Research Working*

Paper Series, 27246, 2020. https://doi.org/10.3386/w27246

[28] Raifman, J., Bor, J., & Venkataramani, A. Unemployment insurance and food insecurity among people who lost employment in the wake of COVID-19. *medRxiv:the preprint server for health sciences*, 2020.07.28.20163618, 2020. https://doi.org/10.1101/2020.07.28.20163618

[29] Blustein, D. L., Duffy, R., Ferreira, J. A., Cohen-Scali, V., Cinamon, R. G., & Allan, B. A. Unemployment in the time of COVID-19: A research agenda. *Journal of Vocational Behavior*, 119, 103436, 2020. https://doi.org/10.1016/j.jvb.2020.103436

[30] Randolph, H. E., & Barreiro, L. B. Herd Immunity: Understanding COVID-19, *Immunity*, 52(5), 737-741, 2020. https://doi.org/10.1016/j.immuni.2020.04.012

[31] Dubé, E., & MacDonald, N. E. How can a global pandemic affect vaccine hesitancy? *Expert Review of Vaccines*, 19:10, 899-901, 2020. https://doi.org/10.1080/14760584.2020.1825944

[32] Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., Kimball, S., & El-Mohandes, A. A global survey of potential acceptance of a COVID-19 vaccine. *Nature medicine*, 1–4, 2020. https://doi.org/10.1038/s41591-020-1124-9

[33] Luigi Roberto Biasio, L. R., Bonaccorsi, G., Lorini, C., & Pecorelli, S. Assessing COVID-19 vaccine literacy: a preliminary online survey. *Human Vaccines & Immunotherapeutics*, 2020. https://doi.org/10.1080/21645515.2020.1829315

[34] Lin, Y., Hu, Z., Zhao, Q., Alias, H., Danaee, M., & Wong, L. P. Understanding COVID-19 vaccine demand and hesitancy: A nationwide online survey in China. *PLoS Negl Trop Dis*, 14(12): e0008961, 2020. https://doi.org/10.1371/journal.pntd.0008961

[35] Reiter, P. L., Pennell, M. L., & Katz, M. L. Acceptability of a COVID-19 vaccine among adults in the United States: How many people would get vaccinated? *Vaccine*, 38(42), 6500-6507, 2020. https://doi.org/10.1016/j.vaccine.2020.08.043

[36] Loomba, S., de Figueiredo, A., Piatek, S., de Graaf, K., & Larson, H. J. Measuring the Impact of Exposure to COVID-19 Vaccine Misinformation on Vaccine Intent in the UK and US. *medRxiv*, 2020. https://doi.org/10.1101/2020.10.22.20217513

[37] Hacquin, A. S., Altay, S., de Araujo, E., Chevallier, C., & Mercier, H. Sharp rise in vaccine hesitancy in a large and representative sample of the French population: reasons for vaccine hesitancy. *PsyArXiv*, 2020. https://doi.org/10.31234/osf.io/r8h6z

[38] Palamenghi, L., Barello, S., Boccia, S., & Graffigna, G. Mistrust in biomedical research and vaccine hesitancy: the forefront challenge in the battle against COVID-19 in Italy. *European journal of epidemiology*, 35(8), 785–788, 2020. https://doi.org/10.1007/s10654-020-00675-8

[39] Karlsson, L. C., Soveri, A., Lewandowsky, S., Karlsson, L., Karlsson, H., Nolvi, S., ... & Antfolk, J. Fearing the disease or the vaccine: The case of COVID-19. *Personality and Individual Differences*, 172, 110590, 2021. https://doi.org/10.1016/j.paid.2020.110590

[40] Khan, Y. H., Mallhi, T. H., Alotaibi, N. H., Alzarea, A. I., Alanazi, A. S., Tanveer, N., & Hashmi, F. K. Threat of COVID-19 vaccine hesitancy in Pakistan: the need for measures to neutralize misleading narratives. *The American journal of tropical medicine and hygiene*, 103(2), 603-604, 2020. https://doi.org/10.4269/ajtmh.20-0654

[41] Harapan, H., Wagner, A. L., Yufika, A., Winardi, W., Anwar, S., Gan, A. K., Setiawan, A. M., Rajamoorthy, Y., Sofyan, H., & Mudatsir, M. Acceptance of a COVID-19 Vaccine in Southeast Asia: A Cross-Sectional Study in Indonesia. *Frontiers in public health*, 8, 381, 2020. https://doi.org/10.3389/fpubh.2020.00381

[42] Dror, A. A., Eisenbach, N., Taiber, S., Morozov, N. G., Mizrachi, M., Zigron, A., ... & Sela, E. Vaccine hesitancy: the next challenge in the fight against COVID-19. *European Journal of Epidemiology*, 35(8), 775–779, 2020. https://doi.org/10.1007/s10654-020-00671-y

[43] Al Awaidy, S. T., & Khamis, F. Preparing the Community for a Vaccine Against COVID-19. *Oman medical journal*, 35(6), e193, 2020. https://doi.org/10.5001/omj.2020.130

[44] Kwok, K., Li, K. K., WEI, W., Tang, A., Wong, S., & Lee, S. Influenza vaccine uptake, COVID-19 vaccination intention and vaccine hesitancy among nurses: A survey. *International Journal of Nursing Studies*. 114. 103854, 2021. https://doi.org/10.1016/j.ijnurstu.2020.103854

[45] Luo T., Chen S., Xu G., & Zhou J. Sentiment Analysis. In: Trust-based Collective View Prediction. Springer, New York, NY, 2013. https://doi.org/10.1007/978-1-4614-7202-5_4

[46] Kantardzic, M. *Data mining: Concepts, models, methods, and algorithms (3rd ed.)*. Hoboken, NJ: Wiley-IEEE Press, 2019.

[47] Albon, C. *Machine learning with Python cookbook: Practical solutions from preprocessing to deep learning*. Sebastopol, CA: O'Reilly Media, 2018.

[48] Callaghan, T., Moghtaderi, A., Lueck, J. A., Hotez, P., Strych, U., Dor, A., Fowler, E. F., & Motta, M. Correlates and disparities of intention to vaccinate against COVID-19. *Social science & medicine (1982)*, 113638, 2021. https://doi.org/10.1016/j.socscimed.2020.113638