Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Analyzing Iranian Opinions toward COVID-19 Vaccination

Zahra Bokaee Nezhad, Mohammad Ali Deihimi

PII: S2772-7076(22)00003-0
DOI: https://doi.org/10.1016/j.ijregi.2021.12.011
Reference: IJREGI 50

To appear in: IJID Regions

Received date: 9 November 2021
Revised date: 6 December 2021
Accepted date: 26 December 2021

Please cite this article as: Zahra Bokaee Nezhad, Mohammad Ali Deihimi, Analyzing Iranian Opinions toward COVID-19 Vaccination, IJID Regions (2022), doi: https://doi.org/10.1016/j.ijregi.2021.12.011

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Highlights

- Twitter is a good source for investigating public opinions.
- Positive polarity toward imported vaccines is slightly dominant in Iran.
- Health agencies can work on Twitter to raise proponents of COVID-19 vaccination.
- Health agencies can work on Twitter to reduce opponents of COVID-19 vaccination.
Title Page

Analyzing Iranian Opinions toward COVID-19 Vaccination

Zahra Bokaee Nezhad *1 and Mohammad Ali Deihimi 2

1,2 Ph.D. Candidate at Department of Computer Science, Shiraz University, Shiraz, Iran

*1Corresponding author: sattar khan blvd., no.12, 302, 718495336, shiraz, fars province, iran.
Email: zboae@gmail.com
Mobile: +989171885220 - Office-Phone: +987138330187 – Fax: +987138428418

2. Pasdaran street, no.66, 193, 718395436, shiraz, fars province, iran. Email:
m.a.deihimi@gmail.com
Mobile: +9891711153529

Abstract:

Objectives: The present study aimed to assess Iranian tweets to (1) Analyze Iranian views toward COVID-19-vaccination. (2) Compare Iranian views toward homegrown and imported COVID-19-vaccines. (3) Present an effective model for Sentiment Analysis tasks on critical issues like COVID-19-vaccination.

Design or methods: Persian tweets were retrieved mentioning homegrown and imported vaccines between 1-April-2021 and 30-September-2021. We identified sentiments of retrieved tweets using a deep learning sentiment analysis model. We used a sarcasm detection model based on Random Forest classifier to discover sarcastic tweets and minimize misclassification. Finally, we investigated Iranian views toward COVID-19-vaccination.

Results: (1) We found subtle differences in number of positive sentiments toward homegrown and imported vaccines, the latter had dominant positive polarity. (2) Negative sentiments
regarding homegrown and imported vaccines increased in some months. (3) We observed no significant differences between percentage of overall positive and negative opinions toward vaccination.

Conclusion: It is worrisome that negative sentiments toward homegrown and imported vaccines in some months increases in Iran. Health organizations can focus on Twitter to promote positive messaging toward COVID-19-vaccination. Sarcasm detection enabled us detect tweets that ironically stated positive sentiments toward vaccination, it improved accuracy of the sentiment analysis results. Our Sentiment Analysis-Sarcasm Detection model is a reliable tool for classification problems.

Keywords: SARS-CoV-2, COVID-19, Public Health, Vaccination, Sentiment Analysis
Introduction:
COVID-19 is an infectious disease caused by the SARS-CoV-2 virus. (World Health Organization, 2020). The development of vaccines against COVID-19 has been a global purpose since the World Health Organization declared the pandemic. (Marcce & Likic, 2021). Attaining a level of herd immunity by vaccination could be complicated as public opinions toward vaccines can change based on different events and even vary between different COVID-19 vaccines. (Chen Lyu, et al., 2021)

Since Iran’s government has aimed to end the pandemic by effective vaccination, they supported Iranian scientists to develop COVIran Barekat as a homegrown vaccine. (Abdoli, et al., 2021). Besides that, several imported vaccines are currently used there, including Sputnik Light, Janssen, Pfizer/BioNTech, AstraZeneca/Oxford, Moderna, Sinopharm, etc. (McGillUniversity, 2021). Comparing homegrown and imported vaccines is a hot topic in Iran. Iranian have engaged in comparing vaccines to decide which one they should get. They also use social media such as Twitter to express their views online, so COVID-19 has been a popular subject on Twitter since January-2020 (Sattar & Arifuzzaman, 2021). Hence, Twitter lets health organizations track public perception of COVID-19 vaccination and helps them with better plans to increase the uptake of COVID-19 vaccines and end the pandemic. As a result, the present study assessed Iranian tweets to (1) Analyze Iranian views toward COVID-19 vaccination. (2) Compare Iranian views toward homegrown and imported COVID-19 vaccines. (3) Present an effective model for Sentiment Analysis (SA) task on critical issues such as vaccination.

To identify public opinions toward COVID-19 vaccination, we first need to assign a polarity of 'positive', 'negative', and 'neutral' to each retrieved tweet. To do so, we utilized a pre-trained Persian SA model. We deployed deep learning classifiers based on CNN-LSTM Hybrid Model
(Convolutional Neural Network – Long Short Term Memory) for sentiment classification, which proved high accuracy in previous works (BokaeeNezhad & Deihimi, 2019). Moreover, sarcasm in a tweet may cause unreliable determination of sentiments. Hence, detecting it in a tweet can improve SA results (Schifanella, et al., 2016). Since Persian speakers often use sarcasm due to their language's nature, Twitter involves many sarcastic tweets in Persian (Golazizian, et al., 2020). In this study, sarcasm detection enabled us to detect several tweets that ironically stated a positive sentiment toward vaccination. So for increasing the accuracy of the SA model and minimizing potential misclassifications, we used a pre-trained sarcasm detection model to modify each tweet’s sentiment label. Random Forest classifier was used for sarcasm detection. Subsequently, we created the first vaccine-related dataset in Persian for analyzing Iranian opinion toward COVID-19 vaccination.

To the best of our knowledge, the present study is the first attempt to analyze public concerns regarding COVID-19 vaccines in Iran. Figure 1 shows the workflow of the suggested methodology.

Materials and Methods:

Data acquisition:

Figure 1 depicts an overview of the proposed model architecture. Python programming language library called "Tweepy" was connected to the Twitter academic API to collect related tweets. We conducted separate searches on Twitter regarding imported and homegrown vaccines. The search keywords for the homegrown vaccine were "بکیت، واکسن داخلی" (Barekat vaccine, homegrown vaccine). Accordingly, the keywords of retrieving tweets for foreign vaccines were "اکسین آسترازئناکا، واکسن فايزر، واکسن یونیورسال، واکسن سینوفارم، واکسن خارجی" (AstraZeneca vaccine, Oxford vaccine, Pfizer vaccine, Moderna vaccine, Sinopharm, foreign vaccine).
Therefore, all Persian tweets posted in the time frame from 1-April-2021 to 30-September-2021 that related to our keywords were retrieved.

Data preprocessing:

In this step, we cleaned our tweets from non-Persian tweets, URLs, retweets, mentions, and some special characters such as ‘^ % # - +’. To clean all this trash, we used “re” Python module. This step also consisted of two sub-processes as (1) Conversion of emoticons to words. To do so, we used the Emoji Dictionary proposed by (BokaeeNezhad & Deihimi, 2020). That is, changing all emoticons by their related words. (2) Conversion of Persian slang and proverb with direct meaning using Proverb Dictionary proposed by (BokaeeNezhad & Deihimi, 2020). It can help the classifiers to detect sarcasm with more accuracy (Bouazizi & Ohtsuki, 2016). This step was done manually by five Persian linguists. Linguistics checked all retrieved tweets and converted each slang or proverb to its related direct meaning. Then we applied a toolkit named Hazm, for Persian text preprocessing tasks. This toolkit performs various activities, including normalization, space correction, tokenization, stemming, parts of speech tagging, etc.

Persian Datasets

After preprocessing step, we combined the results of imported vaccine searches to develop our first dataset: Foreign-Vaccine dataset. Accordingly, we built our second dataset: Homegrown-Vaccine dataset.

With 803278 tweets collected from Twitter, the Homegrown-Vaccine dataset contained 400839 (49.9%) tweets, and the Foreign-Vaccine dataset included 402439 (50.09%) tweets accordingly.
**Vaccine Sentiment Analysis**

We used the pre-trained hybrid deep learning model proposed by (BokaeeNezhad & Deihimi, 2019), to assign three polarity scores (‘positive,’ ‘negative,’ and ‘neutral’) to each tweet. The model trained on a Persian database consisted of 11616 tweets. The proposed model was based on CNN-LSTM architecture and proved the effectiveness of using deep learning classifiers on Persian datasets. In this architecture, CNN was used as a feature extractor for LSTM on textual input data. The proposed model also used Word2vec as word embedding. Using the model, we labeled both datasets with positive(+1), negative(-1), and neutral(0) labels. Figure 2 illustrates the percentage of tweets in each sentiment class on both datasets. As shown in the figure, before using the sarcasm detection model, the positive sentiment towards foreign vaccine accounted for 46% of tweets(n=185,121), followed by the negative sentiment for 42% (n=169,024) and neutral sentiment for 12%(n=48,292) respectively. In this stage, we observed a slight difference between positive and negative sentimental scores toward foreign vaccines. On the other hand, as can be seen in figure 2, the positive sentiment for the homegrown vaccine accounted for 44% of tweets(n=176,369), followed by the negative sentiment for 36%(n=144,302) and neutral sentiment for 20%(n=80,167), respectively. As it can be seen, the number of neutral views on the homegrown vaccine was 8% higher than the same sentiment in foreign vaccines.

**Vaccine Sarcasm Detection**

As sarcasm makes it challenging to recognize the sentiment of a tweet correctly, sarcasm detection is a significant step to SA (Son, et al., 2019) and it enabled us to detect several tweets that ironically stated a positive sentiment toward vaccination. Hence, this study considered the possibility of sarcasm in tweets. We used a pre-trained Machine Learning model for Persian sarcasm detection proposed by (BokaeeNezhad & Deihimi, 2019) to detect sarcastic tweets. The
model trained on 10023 Persian tweets. Based on the model, three sets of features were extracted, namely (1)Deep-Polarity-Feature, (2)POS-Feature, and (3)Punctuation-Feature. The features were extracted to cover different types of Persian sarcasm. Based on these features, all of the retrieved tweets were represented by feature vectors. The explanation of each set of features is provided in the following sections.

(1) Deep-Polarity-Feature

This feature focused on sentence-level inconsistency. For example, in the retrieved tweet “من واقعاً افتخار می‌کنم که وکسین داخلی داریم حتی مبخور یک منابع کنند از این *** تسین! (I'm so proud to have a homegrown vaccine! They will force us to get this F** damn without any doubt!). There is a sentence-level inconsistency between the first and second parts of the tweet (I'm so proud to have a homegrown vaccine! with a positive sentiment, and they will force us to get this F** damn without any doubt! with a negative sentiment). Hence, the sarcasm detection model considered each tweet with more than 12 tokens as a Multiple-Sentence-Tweet (BokaeeNezhad & Deihimi, 2020). Each Multiple-Sentence-Tweet was divided into two parts. Thus, for the mentioned tweet, we have:

Part 1: I'm so proud to have a homegrown vaccine!

Part 2: They will force us to get this F** damn without any doubt!

As mentioned previously, if there was any sentiment inconsistency between the first and second parts of the tweet, the tweet might be sarcastic. Hence, in this stage, the Sarcasm Detection model used the CNN-LSTM model again to detect the sentiment of each part of Multiple-Sentence-Tweets separately. Consequently, two new binary features, dpf1(deep-polarity-feature 1) and dpf2(deep-polarity-feature 2) were introduced. Dpf1 was activated if the first and second parts of the Multiple-Sentence-Tweet did not have the same sentiment. Dpf2 was activated if the
first or second parts of the *Multiple-Sentence-Tweet* did not have the same sentiment as the whole *Multiple-Sentence-Tweet*'s sentiments.

(2) POS-Feature

This step was conducted to prepare another binary feature named POS-feature. Based on the proposed approach, two POS patterns appeared in most Persian sarcastic tweets:

1-pronoun+adverb+adjective+verb like: خداوند من چقدر خوشبختم که قراره واکسن ایرانی بزنم!! (OMG! How lucky I am to receive the homegrown vaccine!!)

2-noun+adjective+adjective+adjective like: وای یه واکسن خطرناک مشکوک خارجی دیگه! چه باحال! (How Cool! Another dangerous, suspected imported vaccine!)

In this part, the model checked all tweets' POS tags. If any tweets contained one of these patterns, the POS-feature was activated.

(3) Punctuation-Feature

Many research has shown that punctuation has a remarkable impact on text classification. In other words, sarcasm affects words and meaning and also translated into a particular use of punctuation or repetition of words to show some special feelings such as hatred, wonder, exaggeration, etc. (Tungthamthiti, et al., 2014)

Owing to this fact, the Sarcasm Detection model counted the number of repeated sequences of exclamation (!) and question marks (?) separately. In addition, it counted the number of repetitive characters such as ِهنگبِه،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ،خخشخ，etc.,(sarcastic signs appeared in several Persian sarcastic tweets). Then, the model considered two new binary features, *Low-Punc-feature* and *High-Punc feature*. For this stage, five Persian linguistics investigated several sarcastic tweets to find the optimum number for the Punctuation-feature activation. Hence these features were activated as follow:
Low-Punc-feature activated if the number of ? or ! or repetitive characters < 3.

High-Punc-feature activated if the number of ? or ! or repetitive characters ≥ 3.

Consider the following tweet “"أره خوشحالکم حق انتخاب ندارم و بايد واکسن جيني بزنم!!!!!!!" (Yah! I am happy to have no right and should get the Chinese vaccine!!!!!). The model detected the sarcasm in this tweet by discovering a repeated exclamation mark (High-Punc-feature was activated).

After all sets of features were extracted, the model used the Random Forest classifier to detect sarcastic tweets. Since we used the pre-train Sarcasm Detection model, the process of Random Forest classifier started with no need to receive any training data as input. Hence, it separately got both homegrown-vaccine and foreign-vaccine datasets to label each tweet with either ‘sarcastic’(0) or ‘not sarcastic’(1).

After using the sarcasm detection model on both datasets, we applied our methodology based on (Yunitasari, et al., 2019) works. Suppose a tweet was classified as "positive"(+1) during the SA stage, yet it was classified as "sarcastic"(0) during the sarcasm detection stage. In that case, the tweet's sentiment was reversed to "negative." In other words, the sentiment of a tweet was changed when the tweet was classified as "positive" and "sarcastic" simultaneously. Let's consider the following tweet, "من چقدر خوشحالم خدايا مرسی بالاخره بعد از مدت ها انتظار آمرز فايرزوم رو زدم !! و احتمالا با عارضه های شدیدکا نا تب بیمارم " (Thank God I am blissfully happy since I finally gave my Phizer vaccine today, and I am likely to die tonight for its interesting side effect!!). The SA model classified the tweet as "positive," while the sarcasm detection model labeled it "sarcastic."

As a result, the tweet's sentiment was reversed, classified as "negative." Table 1 showed some cases of sarcasm in both datasets, which misled the SA results.

We assumed our SA + sarcasm detection model was our final model, which can be used in critical classification tasks like COVID-19 vaccination. Therefore, we tested its performance first
and compared it with the SA model itself to ensure our datasets were well-performed with high-accuracy labeling using the sarcasm detection model. That is, one time, we tested our SA model on datasets. Then we integrated the SA and sarcasm detection models and tested the integration model on our datasets.

Since we had two datasets in this study as Home-grown vaccine dataset and the foreign-vaccine dataset, we tested our model using cross-validation on each dataset separately and finally considered the average performance metrics achieved from both datasets. To do so, for each dataset, we randomly selected 90% of all labeled data as train set and the remaining of them as the test set. This study used the scikit-learn library, which implements cross-validation by using KFold() scikit-learn class. The results are presented in Table 2.

Figure 3 illustrates the percentage of tweets in each sentiment class after using the sarcasm detection model. As shown in the figure, the positive sentiments towards foreign vaccines accounted for 43% of tweets (n=173,048), followed by the negative sentiments for 45% and the neutral sentiments for 12%, respectively. The results revealed that positive views towards foreign vaccines decreased by about 3% after deploying the sarcasm detection model. On the other hand, figure 3 indicates that the positive sentiments towards the homegrown vaccine accounted for 40% of the tweets (n=160,335), followed by the negative sentiments for 40% and the neutral sentiments for 20%. After deploying the sarcasm detection model, we found that the number of positive and negative opinions towards homegrown vaccines became the same. Moreover, the positive sentiments towards the homegrown vaccine decreased by about 4% in this stage.
Result and Discussion:

In this section, we first tested the effectiveness of sarcasm detection in our study to see whether it can be a reliable method for improving SA results. Afterward, we investigated the Iranian opinions toward COVID-19 vaccines.

The effectiveness of Sarcasm Detection

In this stage, we evaluated our result with and without sarcasm detection model to demonstrate that the SA model in sensitive topics such as Vaccine Opinion needs to be improved by integrating some methods such as sarcasm detection model. On the other hand, vaccination in Iran is faced with some public obligations, and in such situations, native people usually tend to use sarcastic assertions to convey their opposition. (Karim, et al., 2021) Therefore, to prove that sarcasm detection on this occasion can be a reliable method to improve the accuracy of SA, we provided several tests on our approach in this section first. To do so, we used a k-fold cross-validation method to estimate the accuracy of the model. As presented in Table 2, sarcasm detection improved the accuracy and precision of SA. The accuracy of the SA model when using the sarcasm detection model was consistently above 79% in every fold, which confirmed that the sarcasm detection model could be stable to improve the performance of SA. Yet, the recall values declined in each fold, which meant that the correct sentiment prediction decreased. One reason was that we considered sarcasm only with negative meaning, while in reality, it is reasonable to have sarcasm with a positive connotation (Yunitasari, et al., 2019). Nevertheless, by improving the SA accuracy using the sarcasm detection model, we had two separated datasets with positive, negative, and neutral labels.
Overall sentiment towards COVID-19 vaccines

The frequency of the collected tweets regarding COVID-19 vaccines over six months is illustrated in figure 4 for each vaccine group separately. As shown in figure 4, there was a spike in the number of foreign vaccines tweets during the second week of April, with approximately 17000 tweets. A possible interpretation of this could be the declaration banning the import of UK and US COVID-19 vaccines in Iran during those times. Another notable rise in the weekly number of tweets can be observed in August's second and third weeks. That was the time when Iran's government officially announced that permission for AstraZeneca/Oxford, Pfizer/BioNTech, and Moderna was issued to import to Iran.

On the other hand, figure 4 reveals a rise in the number of homegrown vaccine tweets in August's first and second weeks, with about 17100 tweets. During those times, a public awareness asked the government to submit Barekat vaccine’s application to WHO for evaluation. The public awareness raised many people’s hesitation towards the homegrown vaccine. Another significant increase regarding Barekat vaccine was observed in May's second and third weeks when Iran's government officially declared successfully developing Barekat vaccine.

Negative Sentiment towards COVID-19 Vaccines

The distribution of negative sentiments towards COVID-19 vaccines is illustrated in Figure 5. As shown in Figure 5, there was no statically significant difference in the negative sentiment of tweets towards foreign vaccines from April to late July. However, Pfizer's side effects in Iran showed a notable rise in negative tweets between late August and September. This upward trend could also be related to some specific tweeter's accounts that tried to make negative opinions towards particular groups of vaccines(Yousefinaghani, et al., 2021). During other months of study, the average negative sentiment towards foreign vaccines had no significant change.
On the other hand, as shown in figure 5, the negative sentiment towards Barekat vaccine dramatically increased at the beginning of April. As mentioned before, some news claimed about banning the import of the UK and US COVID-19 vaccines in that period. Such reports could correlate with spikes in the negative sentiment toward Barekat vaccine. However, the negative sentiment toward Barekat vaccine remained steady until late August. In fact, between late August and September, an increase in negative sentiments towards the homegrown and foreign vaccines came about during the same period. A possible explanation could relate to the reports claiming Iran’s government mandated Barekat vaccine due to Pfizer reported side effects. Hundreds of people opposed Barekat vaccine mandates. However, since no mandate has happened, the negative sentiment toward Barekat decreased in late September.

Figure 6 depicts the number of tweets with positive, negative, and neutral sentiments per month for each dataset. The figure shows that positive tweets towards foreign vaccines were the dominant sentiments for approximately four months of study. In comparison, the amount of positive and negative sentiments towards the homegrown vaccine were almost the same. By analyzing sentiments between both groups of COVID-19 vaccines, our results indicated that while the negative sentiment related to foreign vaccines increased, the positive sentiment regarding them decreased no more than 10% for the first four months of the study. However, the negative sentiments towards foreign vaccines saw their first peak in late August and September (for approximately 15%). At the same time, the positive sentiment regarding them decreased dramatically by about 25%. A possible explanation is that these vaccines initially were perceived as "safe" amongst Iranian people until August. In late August, the announcement of some foreign vaccines side effects significantly decreased their positive sentiment. At the same time, a 12% decrease in the positive sentiment towards the homegrown vaccine was observed.
Interestingly, in early April, the announcement of banning the UK and Us vaccines led to a rise in the positive sentiments toward foreign vaccines. In contrast, it caused a decrease in positive sentiments towards the homegrown vaccine as well. In addition, there was no significant difference in the number of neutral tweets for each vaccine group during the study.

Conclusion

In this study, we analyzed the sentiments of 803278 Persian tweets concerning COVID-19 vaccines retrieved between 1-April-2021 and 30-September-2021. We used a deep learning model for SA and a machine learning model for sarcasm detection to classify vaccine-related tweets more accurately. We concluded that (1)sarcasm detection enabled us to detect several tweets that ironically stated a positive sentiment toward vaccination. So, it improved the accuracy of the SA results, and our SA-Sarcasm Detection model can be a reliable tool for further classification problems.

Our results also indicated (2)a subtle difference in the number of positive sentiments toward the homegrown and foreign vaccines, and the latter had the dominant positive polarity. In fact, sentiments regarding vaccination remained positively stable throughout the first four months of study. However, we observed a slight decrease in the users’ desire to take the vaccine when reports on vaccines’ side effects increased in early August and September. (3)Moreover, the negative sentiment regarding homegrown and imported vaccines seems to be increasing in some months which is worrisome. We also observed no significant differences between the percentage of overall positive and negative opinions toward vaccination amongst Iranian people.

Additionally, we concluded that the issue of the mandatory homegrown vaccine directly led to negative opinions toward it and banning the import of foreign vaccines caused positive sentiments toward them and negative views toward the homegrown vaccines accordingly.
Since public healthcare agencies aim to increase the uptake of COVID-19 vaccines to end the pandemic, they can focus on social media such as Twitter to promote positive messaging toward vaccination.

One of our study's limitations is that collected tweets included just a short time of vaccine availability. Further work can focus on vaccine-related tweets after September, when most people were actively receiving vaccines. Furthermore, the present study did not explore the attitude of Twitter users towards each vaccine separately. We aim to identify more vaccine sentiments and compare their progression by time, post engagement metrics such as retweets, favorites, replies, and account characteristics to enhance our work.

Acknowledgment

Declaration of interests

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

Conflict of Interest and Authorship

- All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

- This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

- The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Funding Source
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Ethical Approval statement**

Hereby, I, Zahra Bokaee Nezhad, consciously assure that for the manuscript *Analyzing Iranian Opinions toward COVID-19 Vaccination* the following is fulfilled:

1) This material is the authors' own original work, which has not been previously published elsewhere.

2) The paper is not currently being considered for publication elsewhere.

3) The paper reflects the authors' own research and analysis in a truthful and complete manner.

4) The paper properly credits the meaningful contributions of co-authors and co-researchers.

5) The results are appropriately placed in the context of prior and existing research.

6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.

7) All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

I agree with the above statements.

Date: 11/9/2021

Corresponding author’s signature: Zahra Bokaee Nezhad
References
Abdoli, A. et al., 2021. Safety and potency of BIV1-CovIran inactivated vaccine candidate for SARS-CoV-2: A preclinical study. *Reviews in Medical Virology*, pp. 1-20.

BokaeeNezhad, Z. & Deihimi, M., 2020. Sarcasm Detection in Persian. *Journal of Information and Communication Technology*.

BokaeeNezhad, Z. & Deihimi, M. A., 2019. A COMBINED DEEP LEARNING MODEL FOR PERSIAN SENTIMENT ANALYSIS. *IJUM Engineering Journal*, pp. 129-139.

Bouazizi, M. & Ohtsuki, T., 2016. A Pattern-Based Approach for Sarcasm Detection on Twitter. *IEEE Access*.

Chen, T. & Dredze, M., 2018. Vaccine Images on Twitter: Analysis of What Images are Shared. *Journal of medical Internet research*.

Chen Lyu, J., Le Han, E. & Luli, G. K., 2021. COVID-19 Vaccine–Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. *JOURNAL OF MEDICAL INTERNET RESEARCH*, 23(16), pp. 1-12.

Golazizian, P. et al., 2020. *Irony Detection in Persian Language: A Transfer Learning Approach Using Emoji Prediction*. s.l., European Language Resources Association, p. 2839–2845.

Karim, M. G., A. J. & Bagheri, M. S., 2021. Sarcasm Is the Key: A Gender-Based Study of Impoliteness Strategies in Persian and American Comedy Series. *Journal of Research in Applied Linguistics*, 12(1), pp. 115-126.

Lyu, H. et al., 2021. Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination. *Intelligent Medicine*, pp. 1-34.

Marcec, R. & Likic, R., 2021. Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgraduate Medical Journal*, p. 1.

McGillUniversity, 2021. *COVID-19 Vaccine Tracker*. [Online] Available at: https://covid19.trackvaccines.org/country/iran-islamic-republic-of/

Sattar, N. S. & Arifuzzaman, S., 2021. COVID-19 Vaccination Awareness and Aftermath: Public Sentiment Analysis on Twitter Data and Vaccinated Population Prediction in the USA. *Applied Sciences*, 11(13).

Schifanella, R., Juan, P. d., Tetreault, J. & Cao, L., 2016. *Detecting Sarcasm in Multimodal Social Platforms*. Amsterdam, Association for Computing MachineryNew YorkNYUnited States.

Son, L. H., Kumar, A., Sangwan, S. R. & Arora, A., 2019. Sarcasm Detection Using Soft Attention-Based Bidirectional Long Short-Term Memory Model With Convolution Network. *IEEE Access*, 7(1), pp. 23319 - 23328.

Tunghamthiti, P., Shirai, K. & Mohd, M., 2014. *Recognition of Sarcasms in Tweets Based on Concept Level Sentiment Analysis and Supervised Learning Approaches*. Phuket, Thailand, s.n., p. 404–413.

World Health Organization, 2020. *https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---20-april-2021*. [Online].

Yousefinaghani, S. et al., 2021. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*, Volume 108, pp. 256-262.
Yunitasari, Y., Musdholifah, A. & Sari, A. K., 2019. SARCASM DETECTION FOR SENTIMENT ANALYSIS IN INDONESIAN LANGUAGE TWEETS. *Indonesian Journal of Computing and Cybernetics Systems, 13*(1), pp. 53-62.

**Figure 1.** Flowchart of the proposed model.
Figure 2. Percentage of sentiment polarities towards imported and home-grown vaccines before using the sarcasm detection model
Figure 4. The frequency of the collected tweets regarding COVID-19 imported and Home-grown vaccines over six months.

Figure 5. The distribution of negative sentiments towards COVID-19 imported and home-grown vaccines.
Figure 6. The distribution of vaccine opinions towards COVID-19 imported and Home-grown vaccines over six months.
Figure 3. Percentage of sentiment polarities towards imported and home-grown vaccines after using the sarcasm detection model.
Table 1

Retrieved tweets containing sarcasm with their sentiment polarities.

| Label | Twit |
|-------|------|
| Positive | خیلی خوبه بعد از عمری واکسن به ما هم رسید تازه اگر این همون واکسن خارجی خوبه یانه که نمیکشتنمون! 😏 |
| Positive | (=That's nice, after ages we finally received the vaccine, I wish it were that excellent foreign vaccine which could not harm us! 😏) |
| Positive | خدا رو شکر تزریق واکسن به همسایه من شروع شده چقدر خوبه که قراره موس خدمشگاهی دانشدمون باشیم |
| Positive | (Thank God, its turn to my peers to get the vaccine, that's interesting to become our scientist Laboratory mice!!!) |
| Positive | باشه، اصلا برکت عالیه ولی مبشه توروخد!!! ................................. اگه برهکت برکت نکفید؟؟؟!!! |
| Positive | Ok, Barkat is wonderful! But would you please not say Barekat!!?? |

Table 2

Testing the results of the sentiment analysis model with and without sarcasm detection.
| Test                                      | Accuracy Average | Precision Average | Recall Average |
|-------------------------------------------|------------------|-------------------|----------------|
| k-5 Sentiment Analysis model              | 0.733            | 0.730             | 0.984          |
| k-5 Sentiment Analysis model + Sarcasm Model | 0.812            | 0.826             | 0.911          |
| k-10 Sentiment Analysis model             | 0.760            | 0.762             | 0.981          |
| k-10 Sentiment Analysis model + Sarcasm Model | 0.793            | 0.821             | 0.909          |
| k-15 Sentiment Analysis model             | 0.720            | 0.732             | 0.972          |
| k-15 Sentiment Analysis model + Sarcasm Model | 0.791            | 0.823             | 0.904          |