Advances in Informatics, Management and Technology in Healthcare
J. Mantas et al. (Eds.)

doi:10.3233/SHTI220699

Artificial Intelligence-Based Models for Predicting Vaccines Critical Tweets: An Experimental Study

Uzair SHAH[a], Hazrat ALI[a], Tanvir ALAM[a], Mowafa HOUSEH[a] and Zubair SHAH[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*

Abstract. We studied the suitability of Artificial Intelligence (AI)-based models to predict vaccine-critical tweets on the social media platform Twitter. We manually labeled a sample of 800 tweets as either “vaccine-critical” (i.e, anti-vaccine tweets, mentioned concerns related to vaccine safety and efficacy, and are against vaccine mandates or vaccine passports) or “other” (i.e., tweets that are neutral, report news, or are ambiguous) and used them to train and test AI-based models for automatically predicting vaccine-critical tweets. We fine-tuned two pre-trained deep learning-based language models, BERT and BERTweet, and implemented four classical AI-based models, Random Forest, Logistics Regression, Linear Support Vector Machines, and Multinomial Naïve Bayes. We evaluated these AI-based models using f1 score, accuracy, precision, and recall in three-fold cross-validation. We found that BERTweet outperformed all other models using these measures.

Keywords. vaccines, tweets, machine learning, deep learning

1. Introduction

With the rapid spread of the COVID-19 pandemic, vaccination against the disease was a strong hope to prevent the infection and curtail the spread. However, the misinformation on social media can enormously change the behavior of the public towards the vaccine. For example, vaccine critical posts on Twitter may affect public opinion towards vaccination. Research studies have found that anti-vaxxers have an increased following on social media platforms [1]. Individuals who receive more vaccine critical information are more likely to be reluctant to get vaccinated [1].

The analysis of information on Twitter can provide valuable insights into public opinion [2]. Many studies have demonstrated the extraction of valuable information from Twitter data analysis [3], [4]. However, manual evaluation of the information on Twitter is a resource-intensive task as identified by previous studies too [5]. Monitoring the misinformation on social media is challenging due to the amount of data being generated every day (for example, 500+ million Tweets are generated per day [6]). Artificial Intelligence (AI) can identify important information on Twitter by automatically processing the text [7]. Furthermore, AI techniques enable the rapid analysis of vaccine-related Tweets to identify valuable information such as the barriers in vaccine uptake [8].

[1] Corresponding Author, Zubair Shah, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar. Email: zshah@hbku.edu.qa.

However, these studies rely mostly on a large amount of manually labeled Twitter data, or their scope is limited to traditional machine learning classifiers only. For example, the authors in [9] compared SVM and Naïve Bayes for vaccine stance classification; however, the performance was limited as the reported F1 score was 0.25 only. The authors in [10] used SVM to perform classification of tweets that were initially labeled manually; however, the performance of the SVM model lacked consistency as the F1 score had a high deviation and ranged from 0.22 to 0.92.

In this work, we demonstrate the use of a pre-trained deep learning model that can automatically identify the tweets that are vaccine critical. In addition, the method requires only a small amount of labeled data, thus, effectively reducing the required human effort. Our thorough comparison of two deep learning models with four traditional machine learning models demonstrates the effectiveness of the deep learning models. Furthermore, improved performance has been achieved by using only a small manually labeled dataset.

2. Methods

Data Collection and Labeling: We have collected a large sample of vaccine-related tweets from the Twitter social media platform using its API (application programming interface). These tweets were collected using keywords, such as “vax”, “vaccin”, “immunis”, “immuniz” that are often used by Twitter users when they talk about vaccine-related issues. These keywords match various forms of vaccine-related content such as the keyword vax can match vax, vaxxers, anti-vax, anti-vaxxers, and many such variants. From this large sample of tweets, we randomly selected around 800 tweets and labeled them manually by two experts who have experience in the fields of social media data analytics and public health. Tweets were labeled as either “vaccine-critical” (i.e, anti-vaccine tweets, mentioned concerns related to vaccine safety and efficacy, and are against vaccine mandates or vaccine passports) or “other” (i.e., tweets that are neutral, report news, or are ambiguous) [9]. There were 394 tweets labeled as vaccine-critical and 406 tweets were labeled as others.

AI Models Implementation: We implemented several AI-based models using the Python programming language and its widely available libraries such as NumPy, pandas, transformers, scikit-learn, scipy, and nltk. We fine-tuned two pre-trained deep learning-based language models, BERT (Bidirectional Encoder Representations from Transformers) and BERTweet (a variant of BERT specifically designed for English tweets) using the Python transformers library from Hugging Face2 and implemented four classical AI-based models, Random Forest, Logistics Regression, Linear Support Vector Machines (Linear SVM) and Multinomial Naïve Bayes (Multinomial NB) using the Python scikit-learn3 library. For BERT and BERTweet, we did not perform any preprocessing of the tweets as these models have tokenizers that take care of preprocessing, but for classical AI-based models, we preprocessed the manually labeled tweets by removing emojis, none printable characters, and stopwords such as the, a, in, on, of, etc. We also used the nltk library to perform stemming and lemmatization of text of tweets. Finally, we used the TfidfTransformer module of the scikit-learn library to convert all the tweets' texts to TF-IDF (term frequency-inverse document frequency),

[2] https://huggingface.co/docs/transformers/index

[3] https://scikit-learn.org/stable/

which features vectors for each tweet that represents the importance of words in tweets to the whole corpus of the tweets. We set the parameters of the TfidfTransformer module as max_features=4000, min_df=5, max_df=0.7.

AI Models Training and Testing: We trained and tested all the above-mentioned AI models in a stratified three-fold cross-validation scheme using the manually labeled tweets. Three-fold cross-validation is a standard mechanism in the field of machine learning where data is divided into three equal portions in a stratified manner where class labels are distributed in equal proportion in each portion. AI models are then trained and tested in three rounds, where in each round two portions of the data are used for training, and the third unseen portion by AI-based model is used for testing. We evaluated all the six AI models using f1 score, accuracy, precision, and recall measures and reported the results average over the three-fold cross-validation scheme for these measures.

3. Results and Discussion

The results of AI-based models are provided in Table 1. The best results are highlighted in bold. It can be seen that deep learning-based fine-tuned pre-trained language models performed better than classical AI-based models in all measures. Generally, deep learning-based models require a large amount of data for training to learn various hidden features in the data and generalize well to perform better on unseen test data, but with the latest transfer learning technology pre-trained models can be fine-tuned using a small sample as pre-trained models have already learned much of the features in its previous training on similar tasks. Therefore, even with a small sample of manually labeled tweets both BERT and BERTweet have performed better than the classical models Random Forest, Logistic Regression, Linear SVM, and Multinomial NB. The BERTweet model is slightly better than its parent model BERT using all measures as BERTweet was specifically enhanced for the classification of English tweets. Among the classical models, Multinomial NB is better than Random Forest, Logistic Regression, and Linear SVM in all measures except recall. Random Forest is better than Logistic Regression, Linear SVM, and Multinomial NB in terms of recall.

Table 1. The results of AI-based models averaged over three-fold cross-validation.

AI-based Models	F1-Score	Accuracy	Precision	Recall
BERT	0.8508	0.8575	0.8784	0.8249
BERTweet	**0.8594**	**0.8650**	**0.8824**	**0.8376**
Random Forest	0.7025	0.7162	0.7263	0.6802
Logistic Regression	0.6991	0.7138	0.7248	0.6751
Linear SVM	0.6926	0.7038	0.7082	0.6777
Multinomial NB	0.7032	0.7225	0.7429	0.6675

These results are useful for future research studies to show which models should be utilized when vaccine-related tweets are required to be labeled as vaccine-critical. We also make our code, manually labeled tweets, and model labeled tweets available for other researchers on GitHub to be used for future research. Here is the GitHub link: https://github.com/Uzshah/vaccine-tweets-classification.

4. Conclusions and Future Direction

Vaccine-related tweets are often posted on social media. Many social media users are exposed to these tweets as it appears in their tweet feeds and they might use the information shared through tweets for their health-related decisions. Vaccine-critical tweets might be harmful to social media users who are active consumers of these tweets and this could lead to public health problems. Therefore, it is very important to develop methods to automatically predict if a tweet is vaccine-critical or not. This work studied the suitability of various AI-based models for automatically predicting vaccine-critical tweets and found that fine-tuning BERTweet, an existing pre-trained deep learning-based language model is better than using all other models. Future studies can investigate including additional data such as metadata from tweets to fine-tune BERTweet and see if that improves the model performance.

References

[1] Burki T. The online anti-vaccine movement in the age of COVID-19. The Lancet Digital Health. 2020 Oct 1;2(10):e504-5.

[2] Biswas R, Alam T, Househ M, Shah Z. Public Sentiment Towards Vaccination After COVID-19 Outbreak in the Arab World. In Informatics and Technology in Clinical Care and Public Health 2022;57-60. IOS Press.

[3] ElFadl A, Shah U, Rehman SU, Ali R, Shah Z. News on Twitter: Engagement, Exposure and Estimating Credibility using Machine Learning. In 2021 8th International Conference on Behavioral and Social Computing (BESC) 2021 Oct 29;pp1-5. IEEE.

[4] Dyda A, Shah Z, Surian D, Martin P, Coiera E, Dey A, Leask J, Dunn AG. HPV vaccine coverage in Australia and associations with HPV vaccine information exposure among Australian Twitter users. Human vaccines & immunotherapeutics. 2019 Aug 3;15(7-8):1488-95.

[5] Shah Z, Surian D, Dyda A, Coiera E, Mandl KD, Dunn AG. Automatically appraising the credibility of vaccine-related web pages shared on social media: a Twitter surveillance study. Journal of medical Internet research. 2019 Nov 4;21(11):e14007.

[6] Online twitter statistics, Available at: https://www.internetlivestats.com/twitter-statistics/, Accessed: 09 Mar 2022.

[7] Chen Q, Leaman R, Allot A, Luo L, Wei CH, Yan S, Lu Z. Artificial intelligence in action: addressing the COVID-19 pandemic with natural language processing. Annual review of biomedical data science. 2021 Jul 20;4:313-39.

[8] Lanyi K, Green R, Craig D, Marshall C. COVID-19 Vaccine Hesitancy: Analysing Twitter to Identify Barriers to Vaccination in a Low Uptake Region of the UK. Frontiers in Digital Health. 2021;3.

[9] Kunneman F, Lambooij M, Wong A, Bosch AV, Mollema L. Monitoring stance towards vaccination in twitter messages. BMC medical informatics and decision making. 2020 Dec;20(1):1-4.

[10] Shapiro GK, Surian D, Dunn AG, Perry R, Kelaher M. Comparing human papillomavirus vaccine concerns on Twitter: a cross-sectional study of users in Australia, Canada and the UK. BMJ open. 2017 Oct 1;7(10):e016869.