Bidirectional Encoder Representations from Transformers for the COVID-19 vaccine stance classification

Abhinav Kumar[1], Pradeep Kumar Roy[2] and Jyoti Prakash Singh[3]

[1]*Department of Computer Science & Engineering, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India*
[2]*Indian Institute of Information Technology Surat, Gujarat, India*
[3]*National Institute of Technology Patna, India*

Abstract

Vaccine-related information is awash on social media platforms like Twitter and Facebook. One party supports vaccination, while the other opposes vaccination and promotes misconceptions and misleading information about the risks of vaccination. The analysis of social media posts can give significant information into public opinion on vaccines, which can help government authorities in decision-making. In this work, an ensemble-based BERT model has been proposed for the classification of COVID-19 vaccine-related tweets into AntiVax, ProVax, and neural sentiment classes. The proposed model performed significantly well with a micro F_1-score of 0.532 and an accuracy of 0.532 and achieved the second rank in the shared competition.

Keywords

COVID-19, Vaccination, Deep learning, Machine learning, Social media, BERT

1. Introduction

The COVID-19 outbreak shows no signs of slowing down, and vaccination looks to be the only long-term cure. People all around the globe began to share their thoughts about the vaccination when the first vaccine with a 90 percent efficacy rate was revealed on November 9, 2020 [1]. Many people, however, are sceptical of vaccinations for a variety of reasons. Social media sites like Twitter and Facebook are inundated with vaccine-related information [2]. One group of individuals is in favour of vaccination, while another opposes vaccination and spreads myths and false information about the dangers of vaccination. The study of social media posts can provide valuable insight of public opinion on vaccinations, which can aid government agencies in making decisions about their future steps.

User-generated social media data has been used in the past during disasters to inform others about the situation and assist victims [3, 4, 5, 6, 7]. Singh et al. [3] extracted several textual features from the flood related tweets to classify it into informative and not-informative tweets. Kumar et al. [4] proposed deep multi-modal neural network by combining imagery and textual contents of the disaster-related Twitter posts to classify it into informative and not-informative

Forum for Information Retrieval Evaluation, December 13-17, 2021, India
abhinavanand05@gmail.com (A. Kumar); pradeep.roy@iiitsurat.ac.in (P. K. Roy); jps@nitp.ac.in (J. P. Singh)

contents. Kumar el al. [5] utilized earthquake related tweets to identify vocation names mentioned in the tweets. Baweja et al. [7] proposed a machine learning-based model to first identify the need and availability of various resources during the disaster and then extract tweets of individuals expressing the same need and availability.

A few work [8, 9, 10, 11, 12] have been reported by the researchers where they tried to identify COVID-19 fake news from the social media. Glazkova et al. [8] developed a fine-tuned ensemble-based model composed of three parallel BERT models pre-trained on COVID-19 social media postings in order to detect COVID-19 fake news. Wani et al. [9] experimented with a number of deep learning models, including CNN, LSTM, and several BERT versions whereas [10, 11, 12] used various conventional machine learning classifiers such as Decision Tree, Gradient Boosting, Support Vector Machine, Random Forest, Logistic Regression, and Naive Bayes to identify COVID-19 fake news. Recently, Cotfas et al. [1] proposed several conventional machine learning classifiers such as Naive Bayes, Random Forest, Support Vector Machine and deep learning models such as Bi-LSTM, CNN, and BERT to classify tweets into *neural*, *against*, and *in favour* of vaccination classes.

This work proposes an ensemble-based BERT (Bidirectional Encoder Representations from Transformers) model to classify COVID-19 vaccination related tweets into AntiVax ("the tweet is against the use of vaccines"), ProVax ("the tweet supports / promotes the use of vaccines"), and Neural ("the tweet does not have any discernible sentiment expressed towards vaccines or is not related to vaccines") classes. The proposed model is validated with the shared task *IRMiDis FIRE-2021*.[1]

The rest of the sections are organized as follows: section 2 discusses the proposed ensemble-based BERT model in detail. Section 3 list the finding of the proposed system and finally the paper is concluded in section 4.

2. Methodology

The systematic diagram for the proposed ensemble-based BERT model can be seen in Figure 1. The proposed model is validated by the dataset published in the *IRMiDis FIRE-2021* shared task.[2]. The dataset contains 1,010 tweets of neural , 991 tweets of ProVax, and 791 tweets of AntiVax classes.

The proposed model uses CT-BERT (COVID-Twitter-BERT) [13] to fine-tuned with the of COVID-19 vaccination stance classification tasks. The CT-BERT model was trained using 160 million tweets from January 12 to April 16, 2020, all of which included at least one of the keywords "wuhan," "ncov," "coronavirus," "covid," or "sars-cov-2." These tweets were then filtered and preprocessed, yielding a total training sample of 22.5 million tweets (containing 40.7 million sentences and 633 million tokens). The detail description of the CT-BERT model can be seen in Müller et al. [13].

First, the CT-BERT model is fine-tuned with the three different validation split. To split the provided training set into train and validation, 42, 10, and 20 are used as the random state (RS). To fined-tuned the CT-BERT on our dataset, a maximum length of the tweets is used as 30,

[1] https://sites.google.com/view/irmidisfire2021/home?authuser=0
[2] https://sites.google.com/view/irmidisfire2021/home?authuser=0

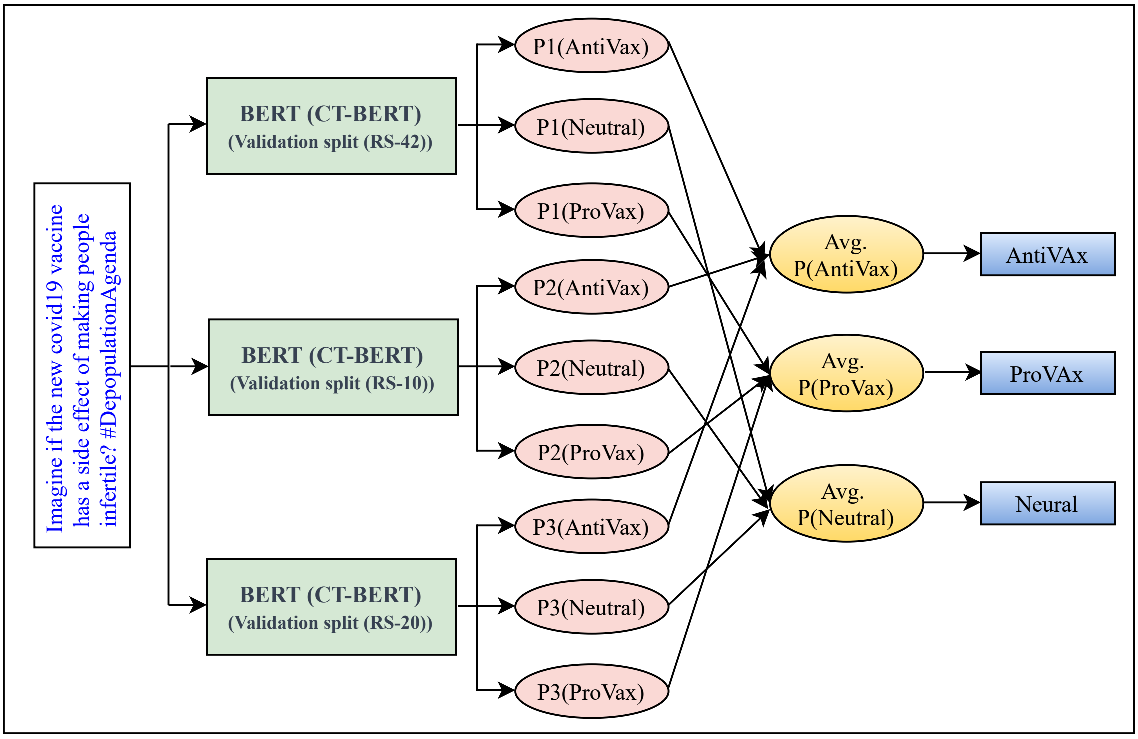

Figure 1: Proposed ensemble-based model for the COVID-19 vaccine stance classification

batch size of 32, a learning rate of $2e^{-5}$ and we trained the model for 20 epochs. After training three different models with different validation split, the class probability is then predicted for provided testing samples. The class-wise probability form all the three trained models is then averaged to get the final probability for each class *AntiVax*, *ProVax*, and *Neutral* (as can be seen in Figure 1. Finally, the test sample belonging to that class which has the highest average probability.

3. Results

The result of the proposed model is measured in terms of macro F_1-score and accuracy. Along with this, class-wise precision, recall, and F_1-score for the validation data are also shown to understand the class-wise performance of the model on the validation set. The results for the three different CT-BERT models fine-tuned on the different validation sets are listed in Table 1.

The CT-BERT (RS-42) achieved both macro F_1-score and an accuracy of 0.85. Similarly, CT-BERT (RS-20) achieved macro F_1-score and accuracy of 0.86 whereas CT-BERT (RS-10) model achieved a macro F_1-score of 0.86 and an accuracy of 0.85.

Three different models (i) Ensemble-based CT-BERT model, (ii) CT-BERT (RS-42), and (iii) CT-BERT (RS-10) were submitted for the final evaluation on the private testing dataset provided by the organizer. The result of the proposed ensemble-based CT-BERT model is listed in Table

Table 1
Results for the CT-BERT models for different validation split

Model	Class	Precision	Recall	F_1-score	Accuracy
CT-BERT (RS-42)	AntiVax	0.84	0.83	0.83	0.85
	Neutral	0.89	0.82	0.85	
	ProVax	0.82	0.89	0.85	
	Macro Avg.	0.85	0.85	0.85	
CT-BERT (RS-20)	AntiVax	0.86	0.84	0.85	0.86
	Neutral	0.87	0.87	0.87	
	ProVax	0.85	0.87	0.86	
	macro avg	0.86	0.86	0.86	
CT-BERT (RS-10)	AntiVax	0.92	0.86	0.88	0.85
	Neutral	0.86	0.85	0.85	
	ProVax	0.80	0.86	0.83	
	Macro Avg.	0.86	0.85	0.86	

Table 2
Results for the proposed ensemble-based BERT model for COVID-19 stance classification on testing dataset

Models	Ranks	Runs	Accuracy	macro-F_1-score
Ensemble-based CT-BERT	2	Run-1	0.555	0.556
CT-BERT (RS-42)	3	Run-2	0.549	0.548
CT-BERT (RS-10)	4	Run-3	0.532	0.532

2. The proposed ensemble-based CT-BERT model achieved a micro F_1-score of 0.556 and an accuracy of 0.555. The CT-BERT (RS-42) achieved a macro F_1-score of 0.548 and an accuracy of 0.549. The CT-BERT achieved a macro F_1-score of 0.532 and an accuracy of 0.532.

4. Conclusion

During the COVID-19 pandemic, social media such as Twitter and Facebook are flooded with COVID-related information. A significant amount of fake information and myths are also posted by the people on the vaccination. In this work, we have proposed an ensemble-based model that classified COVID-19 vaccination-related tweets into three categories such as AntiVax, ProVax, and Neutral. The proposed model is performed significantly well in the shared task and achieved a macro F_1-score of 0.532 and an accuracy of 0.532. In the future, a more robust system can be made by integrating linguistic, character-level, and word-level features together with the ensemble-based model.

References

[1] L.-A. Cotfas, C. Delcea, I. Roxin, C. Ioanăş, D. S. Gherai, F. Tajariol, The longest month: Analyzing covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement, IEEE Access 9 (2021) 33203–33223.

[2] M. M. Müller, M. Salathé, Crowdbreaks: tracking health trends using public social media data and crowdsourcing, Frontiers in public health 7 (2019) 81.

[3] J. P. Singh, Y. K. Dwivedi, N. P. Rana, A. Kumar, K. K. Kapoor, Event classification and location prediction from tweets during disasters, Annals of Operations Research 283 (2019) 737–757.

[4] A. Kumar, J. P. Singh, Y. K. Dwivedi, N. P. Rana, A deep multi-modal neural network for informative twitter content classification during emergencies, Annals of Operations Research (2020) 1–32.

[5] A. Kumar, J. P. Singh, Location reference identification from tweets during emergencies: A deep learning approach, International journal of disaster risk reduction 33 (2019) 365–375.

[6] A. Kumar, J. P. Singh, S. Saumya, A comparative analysis of machine learning techniques for disaster-related tweet classification, in: 2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)(47129), IEEE, 2019, pp. 222–227.

[7] S. Baweja, A. Aggarwal, V. Goyal, S. Mehta, Automatic retrieval of actionable information from disaster-related microblogs., in: FIRE (Working Notes), 2017, pp. 43–45.

[8] A. Glazkova, M. Glazkov, T. Trifonov, g2tmn at constraint@ aaai2021: exploiting ct-bert and ensembling learning for covid-19 fake news detection, arXiv preprint arXiv:2012.11967 (2020).

[9] A. Wani, I. Joshi, S. Khandve, V. Wagh, R. Joshi, Evaluating deep learning approaches for covid19 fake news detection, in: Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers, Springer Nature, 2021, p. 153.

[10] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, T. Chakraborty, Fighting an infodemic: Covid-19 fake news dataset, Communications in Computer and Information Science 1402 (2021).

[11] A. R. Mahlous, A. Al-Laith, Fake news detection in arabic tweets during the covid-19 pandemic, International Journal of Advanced Computer Science and Applications 12 (2021). URL: http://dx.doi.org/10.14569/IJACSA.2021.0120691. doi:10.14569/IJACSA.2021.0120691.

[12] T. Felber, Constraint 2021: Machine learning models for covid-19 fake news detection shared task, arXiv preprint arXiv:2101.03717 (2021).

[13] M. Müller, M. Salathé, P. E. Kummervold, Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter, arXiv preprint arXiv:2005.07503 (2020).