

Overview of the FIRE 2022 track: Information Retrieval from Microblogs during Disasters (IRMiDis)*

Soham Poddar¹, Moumita Basu^{2,*}, Saptarshi Ghosh³ and Kripabandhu Ghosh⁴

¹Indian Institute of Technology, Kharagpur

²Amity University, Kolkata

¹Indian Institute of Technology, Kharagpur

⁴Indian Institute of Science Education and Research, Kolkata

Abstract

Microblogging sites such as Twitter play an important role in dealing with various mass emergencies including natural disasters and pandemics. The FIRE 2022 track on Information Retrieval from Microblogs during Disasters (IRMiDis) focused on two important tasks – (i) to detect the vaccine-related stance of tweets related to COVID-19 vaccines, and (ii) to detect reporting of COVID-19 symptom in tweets.

Keywords

Twitter, microblogs, COVID-19, vaccine stance, tweet classification

1. Introduction

Microblogging social media sites like Twitter are often used to gather real-time information about events happening in the real-world, such as during disasters and pandemics. The ‘Information Retrieval from Microblogs during Disasters’ (IRMiDis) series of shared tasks aims to provide datasets and shared tasks for development of IR/NLP/ML techniques that can be utilized for better management of such crises situations.

In particular, during pandemics such as COVID-19 where complete vaccination is the long-term solution to fight against the disease, social media can be utilized to understand public sentiments towards vaccines [1, 2]. During such pandemics, social media can also be utilized to gain real time insights about symptoms being reported by people, which can be used to track the spread of the disease [3]. To address the two aforementioned important problems, we offered two shared tasks in the FIRE 2022 IRMiDis track.

Task 1: COVID vaccine-stance detection: The only long-term remedy for the COVID-19 pandemic seems to be through society-scale vaccination. However, quite a few people are skeptical about the use of vaccines owing to various reasons, including the politics involved and

FIRE 2022: Forum for Information Retrieval Evaluation, December 9-13, 2022, India

* You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

*Moumita Basu

† These authors contributed equally.

✉ sohampoddar26@gmail.com (S. Poddar); mbasu@kol.amity.edu (M. Basu); saptarshi.ghosh@gmail.com (S. Ghosh); kripa.ghosh@gmail.com (K. Ghosh)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Category of tweets	Count	Tweet excerpt
Anti-Vax	245	I'm not getting any vaccine for Covid. 99% survival rate. Just give me HCQ
Pro-Vax	229	FDA approves Pfizer vaccine just now. It's been a good day for decisions.
Neutral	244	Yup talked about Covid the vaccine, didn't mention the election really much at all.

Table 1
Summary of the test set used in Task 1

the fact that vaccines have been rushed into production. It is important to understand public sentiments towards vaccines, and social media can be used to gain a lot of data quickly about people talking about vaccines.

Building an effective classifier to predict the user-stance (towards vaccines) from social media posts (e.g., microblogs) becomes a crucial first step in any kind of analysis towards vaccine stance. In this task, the participants need to develop automated methods to identify the stance of a tweet (actually, of the user posting the tweet) towards COVID-19 vaccines. Here the tweets are to be classified into three classes – Anti-Vax (against vaccines), Pro-Vax (supports vaccines) and Neutral.

Task 2: Detection of COVID-19 symptom-reporting in tweets: Quickly identifying people who are experiencing COVID-19 symptoms is important for authorities to arrest the spread of the disease. In this task, we specifically explore if tweets that report about someone experiencing COVID-19 symptoms (e.g., ‘fever’, ‘cough’) can be automatically identified. We call such tweets “*symptom-reporting tweets*”.

Note that, simply identifying tweets that contain mentions of COVID-19 symptoms is *not* helpful, since these tweets can contain lots of irrelevant information. For instance, a tweet mentioning “weekend football fever” contains the symptom-word “fever” but is clearly *not* a symptom-reporting tweet. Again, a tweet giving just general information about potential symptoms of COVID-19 is *not* a symptom-reporting tweet. In fact, our analyses showed that a very large majority of tweets that include COVID-symptom words are *not* symptom-reporting tweets, i.e., these tweets do not inform about some person experiencing COVID-19 symptoms. Thus it is important to build an effective classifier to understand which tweets actually inform about someone experiencing COVID-19 symptoms.

In this task, participants need to develop a 4-class classifier on tweets that can detect tweets that report someone experiencing COVID-19 symptoms. The 4 classes are based on who’s reporting the symptoms – (i) Primary Reporting (self-reporting), (ii) Secondary Reporting (reporting for friends or family members), (iii) Third-party Reporting (reporting for celebrity), (iv) Non-Reporting (tweets which do not report about anyone experiencing COVID-19 symptoms).

Team Id	Method	Accuracy	Macro-F1
Data@IITD	Fine tuned CT-BERT	0.770	0.773
IREL	CT-BERT	0.582	0.528
Team-NISER	Pre-trained BERT with data augmentation techniques	0.517	0.514
Thapar Programmers	Fine Tuned BERT	0.529	0.503
DataWiz	XGBoost with Count Vectorizer	0.506	0.490
GAFA	Fine tuned BERT	0.490	0.483
AmiTechies	TF-IDF Vectorizer and Bernoulli Naïve Bayes	0.490	0.473
NLP Learners	TF-IDF Vectorizer and Random Forest	0.391	0.389
UBCS	TF-IDF Vectorizer and Multinomial Naïve Bayes	0.327	0.319
Vishal Nair	TF-IDF vectorization and MLP Classifier	0.353	0.304
Infrared IR	Doc2Vec and SVM	0.396	0.295
Subinay-IISERK	TF-IDF Vectorizer and SVM	0.275	0.281
SSN_NLP	ensemble of xlmroberta, roberta and albert	0.272	0.270

Table 2

Comparison among some of the submitted runs in Task 1 (COVID vaccine stance detection). Runs are ranked in decreasing order of F1-score. We are only reporting the best performing run of each team

2. Task1: COVID-19 Vaccine Stance Classification from Tweets

The datasets: The training dataset comprised of a set of 2,792 tweets from the dataset provided by Cotfas et al. [2] that contains COVID vaccine-related tweets posted during Nov-Dec 2020. The tweets are labeled as Anti-Vax, Pro-Vax or Neutral.

We also collected tweets using various COVID-related keywords, including generic keywords (e.g. ‘vaccine’, ‘vaxxer’), names of COVID-19 vaccines or their manufacturers (e.g. ‘pfizer’, ‘covaxin’), and so on. We randomly selected 2,400 distinct tweets from all the tweets posted during March–December 2020. We got each of these tweets labelled into the three classes (Anti-vax, Pro-vax, Neutral) by three annotators on the Prolific crowdsourcing platform (<https://prolific.co/>). There was at least majority agreement for 2,321 tweets (out of the 2,400), i.e., at least 2 out of the 3 annotators provided the same label. 718 tweets from this were provided as the test dataset and the rest were added to the training dataset.

Table 1 states the distribution of classes in the test dataset along with an example tweet in each class. More details about the data collection and annotation process can be found in the prior work [1].

Methods: In Task 1 we received participation from 13 teams, and 32 runs were submitted. All the teams used common NLP pre-processing techniques and a variety of classification strategies that includes traditional classifiers (such as Multinomial Naïve Bayes and Support Vector Machines), Neural network-based classifiers (MLP), ensemble methods (such as Random Forest) and transformer based pre-trained models such as BERT and Covid-Twitter-BERT [4] (abbreviated as CT-BERT). The summary of the methods is depicted in Table 2.

Category of tweets	Count	Tweet excerpt
Primary Reporting	105	I recently travelled to jaipur and have cough and slight pressure in upper middle chest
Secondary Reporting	29	today, my friend's wife reported high fever so he took her to the belliaghatta hospital
Third-party Reporting	48	Bangladeshi national with high body temperature sent back
Non-Reporting	218	Wash hands, maintain hygiene and cough etiquettes

Table 3
Summary of the test set used in Task 2

Team Id	Method	Accuracy	Macro-F1
Data@IITD	Fine-tuned CT-BERT	0.820	0.746
Thapar Programmers	Fine-tuned BERT	0.790	0.740
DataWiz	XGBoost with Count vectorizer	0.507	0.428
Subinay-IISERK	TF-IDF Vectorizer and SVM	0.295	0.324
Vishal Nair	TF-IDF vectorization and MLP Classifier	0.435	0.239
UBCS	TF-IDF Vectorizer and Multinomial Naïve Bayes	0.351	0.212

Table 4
Comparison among all the submitted runs in Task 2 (detection of COVID-19 symptom reporting tweets). Runs are ranked in decreasing order of F1-score.

Results: We considered two standard classification metrics over the test set – Accuracy and macro-F1 Score (primary metric). Table 2 ranks the submitted runs in decreasing order of the primary metric. We observed recent transformer based models like CT-BERT have outperformed all the other models.

3. Task 2: Detection of COVID-19 Symptom-Reporting in Tweets

The datasets: We crawled English tweets from February 2020 - June 2021 using keywords related to COVID-19 symptoms (e.g., ‘fever’, ‘cough’). This list of symptoms were compiled from the list of symptoms of COVID-19 given by WHO ¹ and by Sarker et al. [3]. We took a random sample from our collected set of tweets and got about 2K tweets annotated into the four classes by human workers. The four classes were Primary, Secondary, Third-party and Non-Reporting (as explained in the Introduction). We split this annotated set of tweets in the ratio 80%-20% and released them as train and test sets. An examples tweet from each of these classes is given in Table 1, along with the distribution of tweets in the test dataset.

Methods: In Task 2, 6 teams participated and 6 runs were submitted. Participating teams employed similar kind of methods that they used in Task 1. The summary of methods is illustrated in Table 4.

¹<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>

Results: We considered two standard classification metrics over the test set – Accuracy and macro-F1 Score (primary metric). Table 4 ranks the submitted runs in decreasing order of the primary metric.

4. Conclusion and Future Directions

The FIRE 2022 IRMiDis track compared the performance of various methods for (i) detecting the stance of the community regarding COVID vaccines, (ii) detecting symptom-reporting COVID-19 tweets. We hope that the test collections developed in this track will be utilized by the research community in the development of better models for both the tasks in future.

Acknowledgments

The track organizers thank all the participants for their interest in this track, and the FIRE authorities for their support in running the track.

References

- [1] S. Poddar, M. Mondal, J. Misra, N. Ganguly, S. Ghosh, Winds of change: Impact of covid-19 on vaccine-related opinions of twitter users, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 16, AAAI Press, 2022, pp. 782–793.
- [2] L.-A. Cotfas, C. Delcea, I. Roxin, C. Ioanăș, 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.
- [3] A. Sarker, S. Lakamana, W. Hogg-Bremer, A. Xie, M. A. Al-Garadi, Y.-C. Yang, Self-reported covid-19 symptoms on twitter: an analysis and a research resource, *Journal of the American Medical Informatics Association* 27 (2020) 1310–1315.
- [4] 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).