Analysis of misinformation concerning covid vaccine

(1305 words)

MARCH 27, 2022.

INDIANA UNIVERSITY (BLOOMINGTON)

Authored by: VARSHA RAVI VARMA
IU ID: 2000751388

Sem: Spring 2022

1. Introduction

> *"More than 500 websites have promoted misinformation about the coronavirus – including debunked claims about the vaccine , according to a firm that rates the credibility of websites" – USA Today News [1]*

Social media enables misinformation and disinformation to be spread rapidly to millions of people in the current world. According to studies, false information circulates more widely than the truth online [2]. Due to social media's global reach with its rapid amplification capability, information can quickly permeate the Internet and become reinforced, potentially resulting in an "infodemic". An excess of information can have negative consequences. In the case of the COVID-19 epidemic, for example, people have performed seemingly harmless acts such as shaving their heads and gargling with salt water, but they have also committed illegal and damaging acts such as arson. Misinformation in health care has identified that false and misleading claims adversely affect people's attitude toward vaccines. This infodemic may significantly alter the course of the pandemic by interfering with public health interventions such as mask-wearing, social distancing, and vaccination. Vaccination is particularly important in light of the infodemic, as it holds the key to returning to pre-pandemic conditions.

A majority of the world's population has received at least one dose of the COVID vaccine. In certain populations, however, vaccine hesitancy is leading to the occurrence of new outbreaks of illnesses at alarming rates. Social media platforms are increasingly popular among supporters of the anti-vaccination movement, who disseminate misleading information about the safety and effectiveness of vaccines [3]. The intent of this study was to explore and analyze misinformation relating to COVID vaccine on the Twitter social media platform using sentiment analysis. Presented in this paper are the findings of this analysis of the extent of worldwide exposure to the COVID infodemic through Twitter platform, i.e., how much misleading information is being spread in the form of tweets among all the covid vaccine related tweets.

2. Research question

The research question for this study can be phrased as follows:
How many of the Tweets related to the covid-19 vaccine contain misinformation?

3. Method

3.1. Data

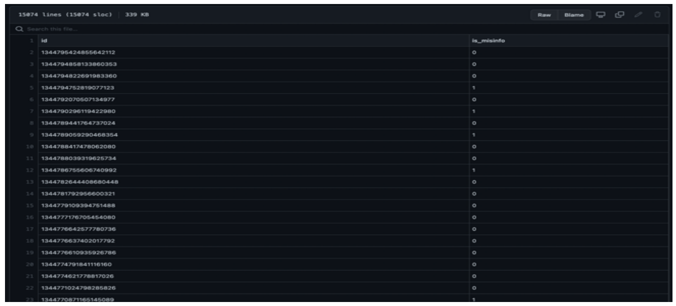
Data source: Twitter [4]

Data: Tweets related to covid vaccine

Data timeline: November 2020 – July 2021

3.1.1. Training phase data:

A pre-existing novel dataset that contains 15 million mutually verified vaccine related tweets and 15 thousand labeled tweets for the detection of vaccine misinformation is available in the GitHub repository [5]. The VaccineTweets folder includes tweets related to the vaccination process from November 2020 through July 2021, arranged by week and the labeled folder consists of over 15000 tweets identified as covid vaccine misinformation.

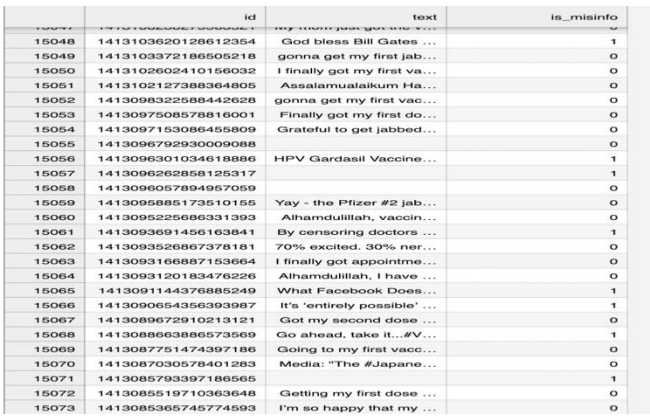


15074 lines (15074 sloc) | 339 KB

	id	is_misinfo
2	1344795424855642112	0
3	1344794858133860353	0
4	1344794822691983360	0
5	1344794752819077123	1
6	1344792070507134977	0
7	1344790296119422980	1
8	1344789441764737024	0
9	1344789059290468354	1
10	1344788417478062080	0
11	1344788039319625734	0
12	1344786755606740992	1
13	1344782644408680448	0
14	1344781792956600321	0
15	1344779109394751488	0
16	1344777176705454080	0
17	1344776642577780736	0
18	1344776637402017792	0
19	1344776610935926786	0
20	1344774791841116160	0
21	1344774621778817026	0
22	1344771024798285826	0
23	1344770871165145089	1

Fig 1: View of a the initial pre-existing labelled covid vaccine data.

https://github.com/SakibShahriar95/ANTiVax/blob/main/Labeled/VaxMisinfoData.csv



	id	text	is_misinfo
15048	1413103620128612354	God bless Bill Gates ...	1
15049	1413103372186505218	gonna get my first jab...	0
15050	1413102602410156032	I finally got my first va...	0
15051	1413102127388364805	Assalamualaikum Ha...	0
15052	1413098322588442628	gonna get my first vac...	0
15053	1413097508578816001	Finally got my first do...	0
15054	1413097153086455809	Grateful to get jabbed...	0
15055	1413096792930009088		0
15056	1413096301034618886	HPV Gardasil Vaccine...	1
15057	1413096262858125317		1
15058	1413096057894957059		0
15059	1413095885173510155	Yay - the Pfizer #2 jab...	0
15060	1413095225686331393	Alhamdulillah, vaccin...	0
15061	1413093691456163841	By censoring doctors ...	1
15062	1413093526867378181	70% excited. 30% ner...	0
15063	1413093166887153664	I finally got appointme...	0
15064	1413093120183476226	Alhamdulillah, I have ...	0
15065	1413091144376885249	What Facebook Does...	1
15066	1413090654356393987	It's 'entirely possible' ...	1
15067	1413089672910213121	Got my second dose ...	0
15068	1413088663886573569	Go ahead, take it...#V...	1
15069	1413087751474397186	Going to my first vacc...	0
15070	1413087030578401283	Media: "The #Japane...	0
15071	1413085793397186565		1
15072	1413085519710363648	Getting my first dose ...	0
15073	1413085365745774593	I'm so happy that my ...	0

Fig 2: View of a the labelled covid vaccine dataset after the retrieval of texts using the tweet id from the pre-existing dataset shown in Fig 1.

To train the classification model, we will use the labeled dataset containing around 15000 tweets from the above-mentioned GitHub repository as the training dataset. Tweepy is used to retrieve tweet text based on the tweet id (shown in Fig 2), which is not present in the initial pre-existing data set (shown in Fig 1). As a result, the csv file will contain three columns, namely tweet ID, tweet text linked to that id, and label indicating misinformation (shown in Fig 2). In order to analyze misinformation contained within the tweets, we use this data to train the model.

3.1.2. Analysis phase data:

To analyze and get the results, we collect our own data from twitter using tweepy, since pre-existing data cannot be used during the testing phase. The dataset contains approximately 1000 tweets. Using tweepy, the dataset is retrieved with two columns, upon which the trained model analyzes and labels each tweet as either 0 or 1, where 1 indicates that the tweet is inaccurate. Based on the outputted dataset, one can calculate the magnitude of misinformation.

	id	text	is_misinfo
1007	1412560685956046856	Cramer reviews Nvidi...	
1008	1412560684160798725	goodnight besties - if i...	
1009	1412560676342673410	To Try to Spur Vaccin...	
1010	1412560668562059267	Today in a conference...	
1011	1412560654867730434	Idk if it's from the lack...	
1012	1412560599658078214	Pin Code:[410301] R...	
1013	1412560571128483843	Dear @POTUS give ...	
1014	1412560565680103426	If massive adverse eff...	
1015	1412560548076605449	Message to the vacci...	
1016	1412560529906831363	Important reminder of...	
1017	1412560526635241473	The world is worried a...	
1018	1412560495224164354	Do you think the vacci...	
1019	1412560489251540992	If I paint my door post...	
1020	1412560486202150918	Our hard work is not li...	
1021	1412560459161427968	Work in the constructi...	
1022	1412560455747313665	Good news - got an a...	
1023	1412560437908951042	' ATAGI took the narro...	
1024	1412560416664961024	I heard the Pfizer vax ...	
1025	1412560380115787777	Anyone else in the sa...	
1026	1412560378882523140	Accidentally administ...	
1027	1412560375887896581	Organizers planning f...	
1028	1412560365737627652	July 2021 COVID-19 ...	
1029	1412560354064875521	That's what I'm talkin...	
1030	1412560350319321090	Why the US has not a...	
1031	1412560334607450114	How do we go from if ...	
1032	1412560331151392771		

Fig 3: View of the initially retrieved unlabeled covid vaccine testing dataset form twitter using

3.2. Analysis

> "In Statistics, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naïve) independence assumptions between the features. They are among the simplest Bayesian network models, but coupled with kernel density estimation, they can achieve high accuracy levels." – Wikipedia [6]

As a part of this study, we will use Naive Bayes classifier to train and test our model to analyze and predict tweets. Text classification, for example, can use the Naive Bayes classifier with discrete features. With a pre-existing dataset of 15,000 tweets as a training set, we train a model to classify tweets as 0 or 1 (where 1 indicates that the tweet contains misinformation) using the evaluate and predict functions. Our model is validated using this dataset and cross validation (n folds). Moreover, we use this trained model to label the self-collected dataset for analysis. Using this labeled dataset, we analyze the magnitude of misinformation being spread on Twitter regarding the covid vaccine.

Sklearn is a simple and efficient tool used for predictive data analysis that is built on NumPy, SciPy and matplotlib. Matplotlib is used to visualize the analyzed data. "sklearn.naive_bayes.MultinomialNB" is used to train the model to classify the tweets into the two categories, i.e., misinformation (1) and valid information (0). After training the model and labeling the tweets, matplotlib is used to plot a bar graph that represents the percentage of each type of information (whether misinformation or valid information). Note that there will be a few rows with null values due to reasons such as data retrieval failure, etc. and such rows will be removed from the dataset while training to avoid affecting the model accuracy. We use the evaluate and predict functions to train the model using various libraries and a pipeline. We retrieve the accuracy of the model and the best parameters that acquires the best accuracy as well.

```
import pandas as pd
import nltk
import numpy as np
from pprint import pprint
from time import time

from sklearn.model_selection import GridSearchCV
from sklearn.pipeline import Pipeline
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_classification
from sklearn.naive_bayes import MultinomialNB
from sklearn.multioutput import MultiOutputClassifier
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.metrics import accuracy_score
from sklearn.metrics import f1_score, precision_score, recall_score
from IPython.display import Image
from sklearn.model_selection import KFold

import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
from nltk.corpus import stopwords
nltk.download('stopwords')
data = pd.read_csv('Training_dataset_15k.csv')
testing = pd.read_csv('Testing_dataset_1k.csv')
```

Fig 4: Glimpse of the python libraries used for the analysis.

4. Results

According to the analysis conducted, approximately 30% of tweets about the COVID vaccine on Twitter contain misinformation. A large number of covid vaccine-related tweets on Twitter are misinformed. Fig. 6 below presents the results of the trained prediction model, such as model accuracy, best parameters, number of valid and misleading information obtained from the testing section of the training dataset, and pipeline parameters. Figure 7 displays the main results of the analysis, which depicts the percentage of valid information related to Covid vaccine, as well as misinformation.

These are the clear results of the analysis of misinformation concerning covid vaccines:

Magnitude of misinformation related to covid vaccine on twitter platform: 28.28%
Magnitude of valid information related to covid vaccine on twitter platform: 71.72%
Model prediction accuracy: 97.20%

```
Performing grid search...
pipeline: ['vect', 'tfidf', 'clf']
parameters:
{'clf__alpha': (0.001, 0.01, 0.1),
 'tfidf__norm': ('l1', 'l2'),
 'tfidf__use_idf': (True, False),
 'vect__max_df': (0.5, 0.75, 1.0),
 'vect__max_features': (None, 200, 500),
 'vect__ngram_range': ((1, 1), (1, 2))}
Fitting 3 folds for each of 216 candidates, totalling 648 fits
done in 56.638s

Best score: 0.972
Best parameters set:
        clf__alpha: 0.1
        tfidf__norm: 'l2'
        tfidf__use_idf: False
        vect__max_df: 1.0
        vect__max_features: None
        vect__ngram_range: (1, 2)
(array([0, 1]), array([2508, 1314]))
----------------------------------------------------------------------
Micro-average quality numbers
Precision: 0.9751, Recall: 0.9751, F1-measure: 0.9751
Macro-average quality numbers
Precision: 0.9715, Recall: 0.9734, F1-measure: 0.9724
```

Fig 6: Results of the training model showing best parameters, accuracy and other model information.

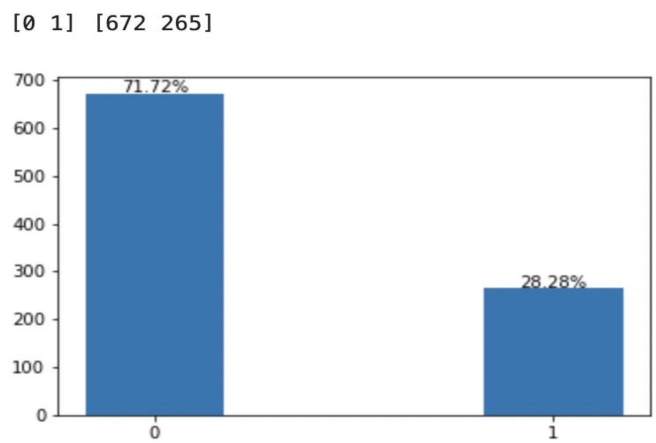


Fig 7: Bar graph showing the percentages of both valid information (0) and misinformation (1)

5. Conclusions and Limitations

Models based on machine learning are effective at detecting misinformation about COVID vaccines on social media platforms. The vaccine-related misinformation on social media may exacerbate vaccine hesitancy, preventing the development of vaccine-induced herd immunity and possibly causing an increase in the number of infections caused by the new strain of COVID. It is important to understand vaccine hesitancy caused due to misinformation spread through the lens of social media.

Based on indirect measures and human-reported values from the dataset available on GitHub, the results of this work may contain confounding factors or biases in response. As an example, an individual who labels the data manually while creating the pre-existing dataset will have different perspectives than others and this will likely lead to inaccurate information. Additionally, the amount of information we can retrieve regarding covid vaccine is also limited. It is possible that there are more misleading posts on Twitter that were taken down in response to the uproar. Thus, our main findings which are associated with rumor exposure should be consistent with the views of only Twitter users.

Therefore, we can infer that approximately 1/3rd of the information posted on Twitter regarding covid vaccine is inaccurate. This study suggests that accurate information promulgating public awareness about the disease's risks and side effects is crucial for the widespread acceptance of vaccines. Given the public's susceptibility to and wide distribution of vaccination-related rumors, we stressed the importance of avoiding falsehoods and spreading the right information.

6. References

Cover page image credits:

- **https://www.edweek.org/leadership/covid-19-vaccines-and-schools-your-questions-answered/2021/01**
- **https://wsma.org/Shared_Content/News/Membership_Memo/2021/January_21/state_redoubles_covid_19_vaccination_efforts_how_wsma_members_can_help**

Introduction:

[1] "COVID, vaccine misinformation spread by hundreds of websites, analysis finds", *Reporter Daniel Funke at dfunke@usatoday.com*, USA Today.

[2] : Singh K, Lima G, Cha M, Cha C, Kulshrestha J, Ahn Y-Y, et al. (2022) Misinformation, believability, and vaccine acceptance over 40 countries: Takeaways from the initial phase of the COVID-19 infodemic. PLoS ONE 17(2): e0263381. https://doi.org/10.1371/journal. pone.0263381

[3] Muric G, Wu Y, Ferrara E, COVID-19 Vaccine Hesitancy on Social Media: Building a Public Twitter, Data Set of Antivaccine Content, Vaccine Misinformation, and Conspiracies, JMIR Public Health, Surveill 2021;7(11):e30642, URL: https://publichealth.jmir.org/2021/11/e30642, DOI: 10.2196/30642

Data:

[4] Twitter website, https://twitter.com/

[5] Hayawi, Kadhim, Sakib Shahriar, Mohamed Adel Serhani, Ikbal Taleb, and Sujith Samuel Mathew. “ANTi-Vax: a novel twitter dataset for COVID-19 vaccine misinformation detection.” Public Health 203 (2022): 23-30

Analysis:

Tadesse, Michael M., et al. "Detection of depression-related posts in reddit social media forum." *IEEE Access* 7 (2019): 44883-44893.

Schrading, Nicolas, et al. "An analysis of domestic abuse discourse on reddit." *Proceedings of the 2015 conference on empirical methods in natural language processing*. 2015.

Mohan, Shruthi, et al. "The impact of toxic language on the health of reddit communities." *Canadian Conference on Artificial Intelligence*. Springer, Cham, 2017.

Copland, Simon. "Reddit quarantined: Can changing platform affordances reduce hateful material online?." *Internet Policy Review* 9.4 (2020): 1-26.

De Choudhury, Munmun, and Sushovan De. "Mental health discourse on reddit: Self-disclosure, social support, and anonymity." *Eighth international AAAI conference on weblogs and social media*. 2014.

Transparency report 2020: https://www.redditinc.com/policies/transparency-report-2020-1

Transparency report 2021: https://www.redditinc.com/policies/transparency-report-2021-2