

Analyzing Twitter, Reddit and 4chan Trends on Covid-19

Umesh Kumar Gandla
ugandla1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Harshavardhan Maragoni
hmarago1@binghamton.edu
Binghamton university
Binghamton, NY, USA

Goutham Metta
gmetta1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Manikya Pandey
mpandey2@binghamton.edu
Binghamton University
Binghamton, NY, USA

Ashika Yellamelli
ayellam1@binghamton.edu
Binghamton University
Binghamton, NY, USA

ABSTRACT

The key to controlling the COVID-19 pandemic is having an understanding of how people react to the COVID-19 immunizations. Examining public opinion is necessary in order to comprehend the reaction. Traditional surveys collect data on a small scale, are expensive and time-consuming, and only cover a restricted number of health-related topics. We are gathering textual information from social networking sites like Twitter and Reddit and 4chan to get a sense of how people feel about the Covid-19 immunization. These websites serve as forums for exchanging knowledge and viewpoints on numerous international issues including COVID-19. The Aim of this study was to compare public opinions and suggestions on covid19 pandemic using the data derived from the social media platforms like Twitter and Reddit and also to study on politics which is derived from Reddit data set.

KEYWORDS

Twitter:
SARS-CoV-2, COVID-19, Vaccines, Coronavirus, Pandemic

Reddit:
Coronavirus, COVID19 Positive, COVID19_support, china_Flu, COVID19, CovidVaccinated, vaccine, pandemic, coronavirusUS, Covldiots.

1. INTRODUCTION

A particular strain of flu first appeared in Wuhan, China, in January 2020, and it quickly spread throughout the world via tourists. In March 2020, the World Health Organization classified it as a pandemic.

Up until mid-May 2021, there have been more over 158 million confirmed instances of Covid-19 worldwide, and the number of fatalities has surpassed 3.2 million. More than 32 million cases and 580,000 fatalities were recorded in the

United States. Because of the lockdowns brought on by this pandemic, more people are now using social media sites like Twitter and Reddit to share news and express their thoughts. These forums have been used by both individuals and government representatives to frequently discuss COVID-19-related policies and news.

Social media, according to research trends, will be crucial for promoting public health. Social media can assist in public health surveillance by facilitating real-time communication at a low cost and monitoring public reaction to health crises. In this case, Twitter and Reddit emerge as crucial information sources that mostly contain personal opinions and responses to those viewpoints rather than posts from carefully constructed content developers.

As more people are getting vaccinated, the COVID-19 dataset can be used to assess how the general population feels about vaccination, how that feeling has changed over time, and what effects vaccination has had. The herd immunity is negatively impacted by low vaccination adoption. Therefore, a significant success element for creating non-pharmaceutical therapies and managing the COVID-19 is the public's response to COVID-19 vaccinations. In order to study the trend in user sentiments linked to the Covid-19 Vaccine, this project will extract, examine, and explore postings from Twitter and Reddit that are related to the topic.

2. Background

The disease has had an unheard-of global impact when the World Health Organization proclaimed COVID-19 a pandemic on March 11, 2020. Twitter, 4chan, Reddit and other social media platforms can be used as a resource to monitor public sentiments and behavior during a crisis and to improve situational awareness. The information gathered can subsequently be used to assist communication and health promotion campaigns, as well as to better understand public attitudes and behaviors during the COVID-19 issue. The aim of this study was to compare public attitude towards the pandemic.

3. DATA SETS

3.1 . Twitter

Data	Data type
_id	object
data.edit_history_tweet_ids.0	int64
data.id	int64
post	object

We collected the above data set of the twitter like _id, data.edit history, data id and the posts which were done by the specific id's. we collected data of 1 million tweet posts from Nov 1 2022 to Nov 14 2022 i.e 14 days of the data is extracted for twitter data set. The extracted data is stored in MongoDB in the form of a csv file.

3.2 . Reddit

Data	Data type
_id	object
index	int64
subreddit	object
post	object
selftext	object
ups	int64

downs	int64
score	int64
created_utc(date)	object
kind	object

We collected the above data set of reddit and we extracted the 3 million posts from Nov 1 2022 to Nov 14, 2022 i.e 14 days of data was extracted from reddit. The extracted data is stored in MongoDB in the form of csv file.

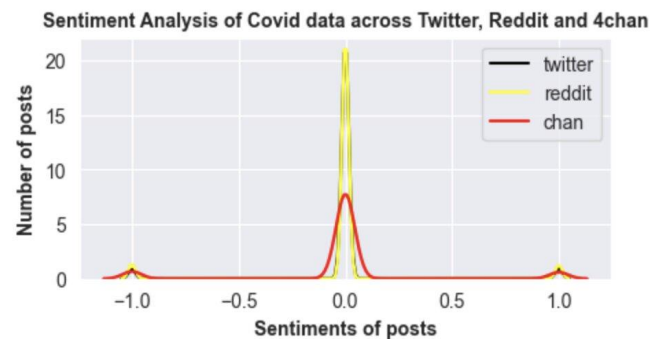
3.3. 4chan

Data	Datatype
_id	object
post	object
now(date)	object
fname	object

We collected the above data set of 4chan and we extracted 225K posts from Nov 1 2022 to Nov 14, 2022 i.e 14 days of data was extracted from 4chan. The extracted data is stored in MongoDB in the form of csv file.

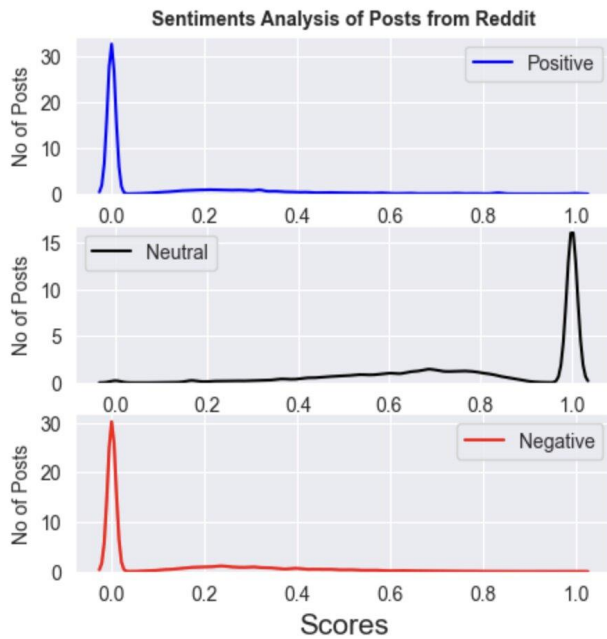
4. Discussions

4.1. Sentimental Analysis on Twitter, reddit, 4chan



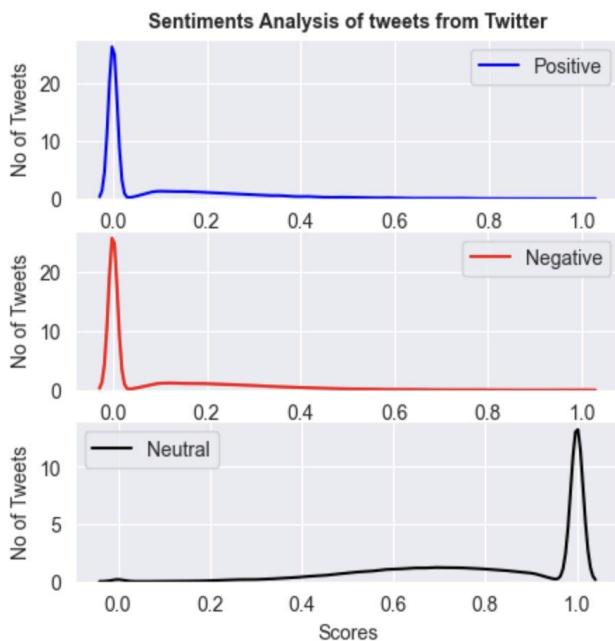
Above linear graph shows the sentimental analysis of posts or tweets from Twitter, Reddit and 4chan. In the above graph 0 means a post is neutral, 1 means post is positive and -1 means a post is negative. The graph plots all posts collected as positive, neutral and negative.

4.2 Sentimental analysis on reddit



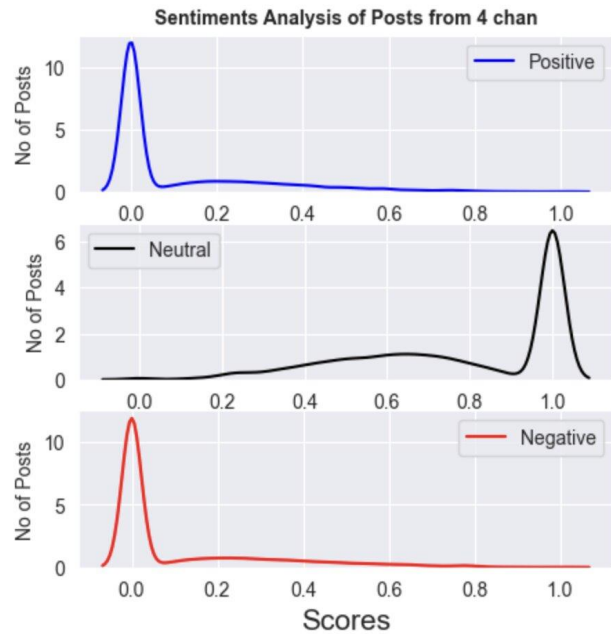
The above graph shows sentiment scores of posts and gives scores as positive, negative and neutral scores. The graph plots positive, neutral, negative scores of all posts in the Reddit dataset as subplots.

4.3 Sentimental analysis on Twitter



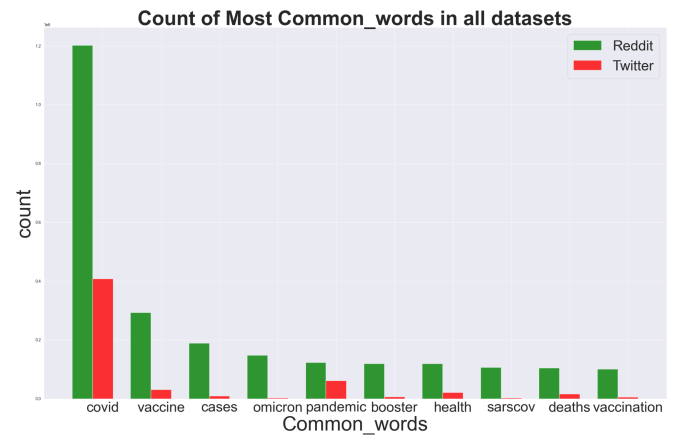
The above graph shows sentiment scores of posts or tweets and gives scores as positive, negative and neutral scores. The graph plots positive, neutral, negative scores of all posts in Twitter dataset as subplots.

4.4 Sentimental Analysis on 4chan



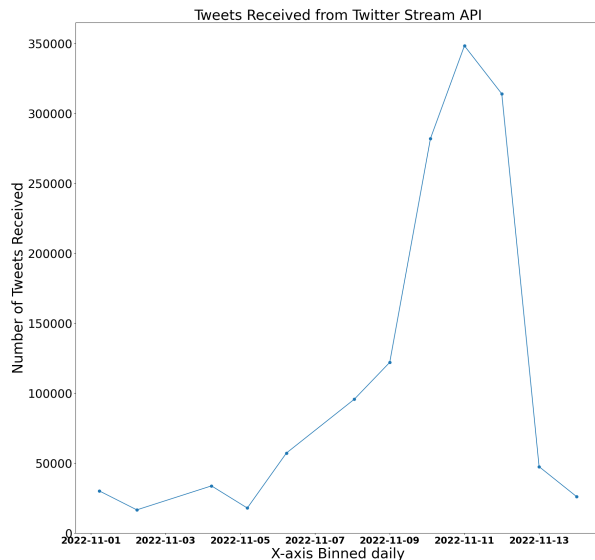
The above graph shows sentiment scores of posts or tweets and gives scores as positive, negative and neutral scores. The graph plots positive, neutral, negative scores of all posts in 4chan dataset as subplots.

4.5 Analysis on common words in twitter, reddit datasets



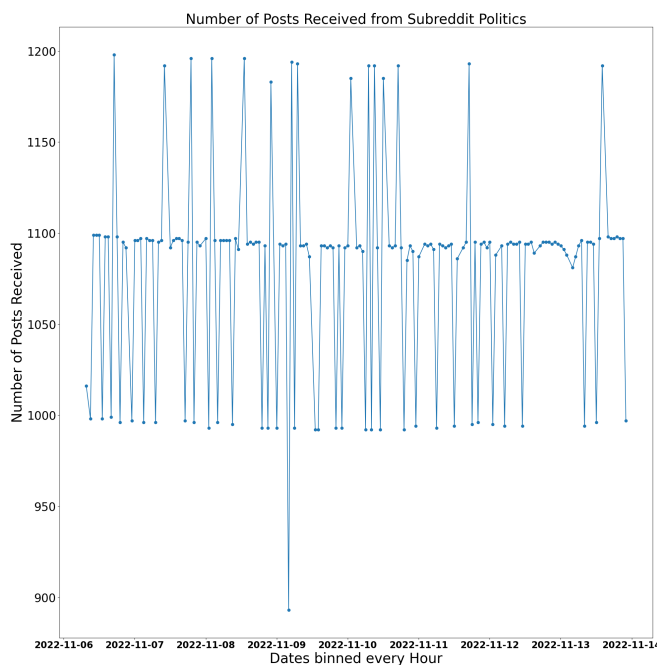
The above graph shows the count of posts for Twitter and Reddit datasets which have common words in all the two datasets.

4.6 Twitter Tweets trend



The above plot graph illustrates the number of tweets that come in over the sample stream from Nov 1st, 2022 to Nov 14th, 2022. The x-axis is binned daily from and y-axis is the number of tweets received from Nov 1st, 2022 to Nov 14th, 2022.

4.7 Reddit data analysis on Politics



The above plot graph shows the number of submissions per day on r/politics from subreddit from Nov 4th, 2022 to Nov 14th, 2022 with data on x-axis and count and y-axis and x-axis is binned hourly.

4.8 Comparison of data using common keywords over Twitter,Reddit,4Chan

	common_name	%_Reddit	%_Twitter	%_4chan
0	covid	36.618031	34.241056	9.054396
1	vaccine	8.935250	2.624121	1.160820
2	cases	5.747269	0.807893	6.036264
3	omicron	4.495428	0.199352	0.000000
4	pandemic	3.769319	5.148357	0.696492
5	booster	3.645651	0.559059	1.160820
6	health	3.629572	1.754334	6.964920
7	sarscov	3.245414	0.181321	0.000000
8	deaths	3.185331	1.377351	3.482460
9	vaccination	3.069459	0.501778	0.000000

The above table shows the percentage rate of common words posted in twitter, reddit, 4chan data sets. It shows the trends of how people often use words in social media platforms for the covid19 pandemic.

Conclusions:

The findings of our study are in agreement with those in the literature regarding the detrimental effects of the epidemic on people's emotions and psychological states. These Data sets for the research adds to the body of real-world evidence that shows the pandemic's numerous harmful effects on society. It also shows how to employ ML techniques as well as statistical modeling and inference to make better use of real-world data. This open dataset will make it possible for researchers to carry out a range of studies on the psychological and emotional reactions to social distance measures, the detection of false information sources, and the stratified measuring of attitude towards the epidemic in close to real-time. An unprecedented amount of free data is being created for genetics and epidemiological research as the COVID-19 epidemic spreads over the globe.

Reference

1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9521381/>
2. Beraldo, D. 2017. Contentious Branding: Reassembling Social Movements Through Digital Mediators. PhD Dissertation. Faculty of Social and Behavioural Sciences, Amsterdam Institute for Social Science Research, Amsterdam.
3. De Zeeuw, D.; and Tuters, M. 2020. Teh Internet Is Serious Business: On the Deep Vernacular Web Imaginary. *Cultural Politics* 16(2).
4. De Zeeuw, D.; and Tuters, M. 2020. Teh Internet Is Serious Business: On the Deep Vernacular Web Imaginary. *Cultural Politics* 16(2).
5. <https://www.openicpsr.org/openicpsr/project/120321/version/V12/view.jsessionid=30EC3A1F5CFF2E59B3A1E38BA8B94044>
6. <https://ieee-dataport.org/open-access/coronaviruss-covid-19-tweets-dataset>
7. https://www.who.int/emergencies/diseases/novel-coronavirus-2019?adgroupsurvey={adgroupsurvey}&gclid=Cj0KCQiA-JacBhC0ARIsAIxybyMlzcFehi0vZ5rqhon_vm6x_Z2oUtf_odJyWak0DuKMDS5gCJGSSlaAjzYEALw_wcB
8. <https://arxiv.org/abs/2205.06863>
9. https://www.researchgate.net/publication/360641108_Sentiment_Analysis_of_Covid-related_Reddits
10. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0275862>
11. <https://www.thedataincubator.com/blog/2018/06/20/a-study-of-reddit-politics/>
12. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014
13. <http://cs229.stanford.edu/proj2016/report/AcostallcusWegrzynski-Predicting%20user%27s%20political%20support%20from%20their%20Reddit%20comment%20history-report.pdf>

