1. How do social media posts correlate with vaccination rates?

2. How different is the sentiment about vaccination by region?

3. How accurately can the vaccination rate be predicted from social media data of a particular region?

Sources:

1. <https://www.proquest.com/docview/2590054784?https://www.proquest.com/coronavirus/advanced&pq-origsite=summon> [2]
2. <https://f1000research.com/articles/10-236/v3> [3]
3. Chang C, Monselise M, Yang CC. What Are People Concerned About During the Pandemic? Detecting Evolving Topics about COVID-19 from Twitter. J Healthc Inform Res 2021 Jan 17:1-28 [1]
4. Lamsal R. Coronavirus (COVID-19) tweets data set. IEEE Data Port. URL: https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset [accessed 2021-08-31] [4]

Introduction

For our background study we drew on research conducted in a similar vein as our project. Because our project seeks to answer the following research questions: 1) How do social media posts correlate with vaccination rates? 2) How different is the sentiment about vaccination by region? 3) How accurately can the vaccination rate be predicted from social media data of a particular region? it became clear that sentiment analysis, location plotting, and vaccination rate projection are our project’s core concepts. Thus, we employed these terms to guide our background study. What will follow is a summary of our background study in regards to the aforementioned core terms.

Other Approaches and Similar Research Contexts

We examined two major research projects that were performed in similar contexts to ours. The first of these sought to “learn about public sentiment for a period of 60 days when the vaccines were started in the United States” [2]. This study based its analysis on the Lamsal R. Coronavirus tweets dataset [4] that “uses over 90 keywords and hashtags to monitor the real-time coronavirus-related tweets from February 05, 2020 till present” [1]. Further, the researchers in this study cleaned the data set so that they only analysed tweets from the 16th of December to the 13th of February because this was when the US Food and Drug Administration had just authorized the Pfizer-BioNTech and Moderna vaccines for emergency use [1]. This study relied on Twitter’s tweet lookup API to retrieve the content and metadata of each tweet for sentiment analysis because the aforementioned dataset merely shared the ID for each tweet [1]. This will be referred to as the Monselise et al. study.

The second study we examined was performed by Sayeed Al-Zaman M.D. It was interested in “understanding the characteristics of coronavirus disease 2019 (COVID-19) vaccine-related social media content and how users engage with them” [3]. Al-Zaman’s study based its research on previous research that categorized COVID-19 online information into 11 different categories that were then simplified into the two broader categories of useful information and misinformation or faulty information [3]. This study was based in Bangladesh so, it was determined that Facebook would be the best data source as it is the most popular form of social media in the region [3]. Instead of sentiment analysis however, Al-Zaman’s study employed content and interaction analyses to evaluate “vaccine-related social media content and how users engage with them” [3].

Sentiment Analysis

Sentiment Analysis is defined as “the classification of text, images, or audio into a set of one or more sentiments” [1]. This is the primary method of analysis we decided to employ in our project. In the Monselise et al. study, where researchers sought to analyse Twitter data with respect to vaccine sentiments, they conducted a “two-step approach where the first step consists of computing the polarity score of tweets that classifies the tweet as positive, neutral, or negative, followed by classifying the tweet into 1 of 5 emotions (anger, fear, joy, hopefulness, sadness)” [2].

This study employed the Valence Aware Dictionary and sEntiment Reasoner (VADER) Python library to perform the first step of sentiment analysis. VADER was chosen by these researchers because it “is more sensitive to expressions of sentiment in the social media context, it performs better than other rule-based classification algorithms in this context, and it outperforms individual human raters in the F1 score” [2].

Source 1

1. sentiment analysis:
   1. Sentiment analysis is a research area that involves the classification of text, images, or audio into a set of one or more sentiments [1]. In a similar study, the researchers classified short snippets of text. Likewise, we will performed a similar design for our sentiment analysis (1).
   2. “To detect the sentiment conveyed in the tweets, we utilized a two-step approach. In the first step, we computed the polarity score of our tweets, and based on this score, we classified the tweets as either positive, neutral, or negative. In the second step, we classified the emotional content of the tweet into 1 of the 5 emotions: anger, fear, joy, hopefulness, and sadness”
   3. “The first classification step was performed using the VADER (Valence Aware Dictionary and sEntiment Reasoner) Python library [14].” (1)
   4. “Since VADER is more sensitive to expressions of sentiment in the social media context, it performs better than other rule-based classification algorithms in this context [45]. It has been found that VADER outperforms individual human raters [14] in the F1 score.”
2. Purpose:
   1. Particularly, we were interested in learning about public sentiment for a period of 60 days when the vaccines were started in the United States. (1)
   2. Social media provides a great data source for listening to the public on what they are thinking and what concerns and questions they have. We used Twitter as a proxy for public sentiment and were able to find the most important discussion topics that pertained to COVID-19 vaccines in the early days of the vaccine rollout. Additionally, we were able to classify public sentiment as it pertained to the vaccines and how this sentiment changed over time overall and in each topic as well. The goal of this research was to examine the discussion topics and public sentiment toward COVID-19 vaccines. By studying the topic and sentiment of the discussion on COVID-19 vaccines on Twitter, we may understand public concerns as they happen and learn more accurately about the source of vaccine hesitancy. By learning what drives vaccine hesitancy, we can better address it and formulate tailored and targeted communication. Conversely, we may also learn about the excitement toward the vaccine and study what is going well and what resonates well with the public on social media. This research will use the results uncovered by the topic and sentiment analysis of the Twitter data and suggest actionable insights for practitioners to address COVID-19 vaccine hesitancy. This research will also address how to utilize positive sentiment toward the vaccine (1).
3. data collection
   1. We adopted the coronavirus tweets data set [37] as our data source, which uses over 90 keywords and hashtags [38] to monitor the real-time coronavirus-related tweets from February 05, 2020 till present
   2. Since the US Food and Drug Administration authorized Pfizer-BioNTech COVID‑19 vaccine and Moderna vaccine for emergency use in mid-December, we only kept tweets that were created during a 60-day period between December 16, 2020 and February 13, 2021 for extracting discussion topics and their sentiment from the general public about COVID-19 vaccines.
   3. we employed Twitter’s tweet lookup application programming interface [39] to retrieve the content and metainformation of each retained tweet because the coronavirus tweets data set only shares the IDs of the collected tweets

Source 2

1. Purpose
   1. This exploratory study seeks to understand the characteristics of coronavirus disease 2019 (COVID-19) vaccine-related social media content and how users engage with them
   2. Researchers categorized COVID-19 online information into 11 types: valid information, comforting information, perplexing information, misinformation, disinformation, shocking information, contradictory information, doubtful information, progressive information, postponed information, and confidential information (Ashrafi-Rizi & Kazempour, 2020). Two broader categories may represent all of them: useful information that helps public health communication, and faulty information that interrupts public health communication. For example, social media was responsible for fake COVID-19 prescriptions and medications that claimed many human lives (Islam et al., 2020).
   3. social media helps to run positive and effective virtual campaigns for the COVID-19 vaccine around the world, and, on the other hand, vaccine opposition and misinformation obstruct the vaccination process (Bonnevie et al., 2021).
   4. A report says 31 million Facebook users follow anti-vaccine groups with 17 million YouTube users subscribing to similar accounts (Burki, 2020). Therefore, it is imperative to understand social media users’ sentiments regarding the COVID-19 vaccine, and the nature of vaccine-related content that dominates social media platforms.
2. Interaction & content analysis
3. Findings
   1. vaccine-related links were the highest in percentage (86.14%) of all Facebook content during the period and the highest shared content (56.21%) as well, which counters the result of Baresch et al. (2011) that showed Facebook users’ infrequently (49%) share news through links. Links shared in social media platforms are mostly news items primarily from different online news portals (Baresch et al., 2011), which infers that Bangladeshi Facebook users mostly consume news items related to the COVID-19 vaccine. However, the decline of links suggests that users are increasingly becoming less interested in vaccine-related news. Kleis Nielsen et al. (2020) also found a similar tendency that using Facebook for COVID-19 news and information is declining: from 10 April to 24 June 2020, the percentage declined from 24 to 12. Unlike the present study, however, the previous study considered all COVID-19 issues instead of only the vaccine issue and produced a similar result.
   2. Like and love are the two leading reactions, meaning the valence of users’ reaction to vaccine-related content is positive and, to some extent, highly intense. On the other hand, the lowest instances of the angry reaction indicate users’ minimal negative emotion toward the issue. Both tendencies indicate a positive valence with from moderate to high arousal, suggesting Facebook users’ optimistic attitude toward the COVID-19 vaccine.