Gurneet Chhabra; Gaurav Chattree

CSC 440.2: Data Mining

October 28, 2021

Course Project Proposal

**USING REDDIT DATA TO UNDERSTAND VACCINCE HESITANCY CONCERNS AMONG HEALTH CARE WORKERS**

**Problem Statement**

As of October 25, 2021, more than 244 million, SARS-CoV-2 infections (Severe Acute Respiratory Syndrome Corona Virus) had been diagnosed, with approximately 4.9 million deaths confirmed from COVID-19 [1]. The 73rd World Health Assembly passed a resolution in May 2020 acknowledging the need of widespread immunization as a public-health objective and a priority for preventing, controlling and terminating SARS-CoV-2 transmission [2]. There are presently more than 150 vaccine candidates’ worldwide, 485 vaccine trials ongoing, and 23 COVID-19 vaccines approved [3].

Before the COVID-19 pandemic, people's unwillingness to acquire safe and recommended accessible vaccinations, known as 'vaccine hesitancy,' was already an increasing problem. Based on the Strategic Advisory Group of Experts on Immunization (SAGE), vaccine hesitancy is the term used to describe: “delay in acceptance or refusal of vaccination despite availability of vaccination services” [4].

Vaccine hesitancy has been acknowledged as a major global health threat by the World Health Organization (WHO) [5]. Denial of the COVID-vaccine reduces the likelihood of herd immunity and may prolong the pandemic. Vaccine apprehension among health care workers, such as, doctors and nurses, as well as those who provide indirect health care and services, such as, helpers, laboratory technicians and even medical waste handlers, is particularly concerning. Health care workers (HCWs) are not only at a higher risk of getting and transmitting the disease, but they can also have a significant impact on patient vaccination decisions. Vaccinated doctors and nurses are more likely to encourage and advice others to get vaccinated [6].

According to a poll in Israel, 78 percent of physicians, 61 percent nurses and 74 percent of general public planned to get a COVID-19 vaccine. In two other studies, low acceptance rates were also seen among Hong Kong nurses, and just 36 percent of US health care workers stated they were likely to take the vaccination as soon as it became available in late 2020s, out of this 56% said that they were not sure and would wait to review more data [7] [8] [9]. According to surveys, the number of Americans who do not intend to get vaccinated fell from September to December 2020 among the general population (it declines from 38% to 32% in one survey and 34% to 27% in another one) [10] [11]. Although this is encouraging, establishing herd immunity may necessitate vaccinating a larger proportion of population10.

Importance of understanding and tackling health care workers vaccination perceptions is critical for promoting COVID-19 vaccine uptake and it can also educate us a lot about how to deal with future infectious disease crises.

To overcome this problem, Reddit provides a great platform being the 7th most visited website in the United States by November 2020. As compared to other social media platforms which requires users to fill in their profile information to sign-up, reddit only asks for username. This helps is achieving an anonymous environment which is conducive to users sharing their true opinion on topics.

This systematic study aim to assess the hesitancy for COVID-19 vaccine(s) among Reddit users who are health care practitioners, to determine the scope of the problem, which can serve as a starting point for investigating the factors influencing the COVID-19 vaccine hesitancy among heath care workers.

**Data Acquisition**

We will be using reddit data for our analysis. We will scrap the Reddit data using Python Reddit API Wrapper (PRAW) library in Python. Out data of interests is just the submissions and the top comments. Also, we are only interested in the opinions of health care worker. In order to identify the reddit users who are health care workers, we will be using subreddits which are made specifically for these health care workers. Some of the subreddits that we have filtered so far are:

* **r/medicine (368K members)** – It is a virtual lounge for physicians and other medical professionals from around the world to talk about the latest advances, controversies, ask questions of each other, have a laugh, or share a difficult moment.
* **r/nurses (16.9K members)** - A discussion page for Nurses!
* **r/nursepractitioner (16.0K members)** - This is a platform designed to inform and unite the NP community. Asking for advice, practice information, the job market, and general banter is encouraged!
* **r/doctors (8.0K members)** - Redditors who are also medical doctors
* **r/studentnurse (105K members)** - A community for nursing students and nurses who love students.
* **r/medicalschool (439K members)** - An international community for medical students.
* **r/physicianassistant (47.3K members)** - This is a subreddit to share information about Physician Assistants (PAs). This sub is open to PAs, MD/DOs, NPs, Nurses, any other medical professional.
* **r/residency (117K members)** – It is a community of interns and residents who are just trying to make it through training! This is a subreddit specifically for interns and residents to get together and discuss issues concerning their training and medicine/surgery.
* **r/AskDocs (339K)** - All flaired medical professionals on this subreddit are verified by the mods

The submissions and top comments will be scrapped only for the users present in these group.

**Algorithmic Choices/ Ideas**

We plan on using the following topic modelling and emotion tagging algorithms for our research. We are currently unsure of the in-depth implementation of these algorithms in our work.

* Latent Dirichlet Allocation (LDA)
* Non-Negative Matrix Factorization (NMF)
* Latent Semantic Analysis (LSA)
* Parallel Latent Dirichlet Allocation (PLDA)
* Pachinko Allocation Model (PAM)
* Linguistic Inquiry and Word Count (LIWC)

**Performance Evaluation**

Coherence Score will be our primary performance metric. Several attempts will be made by iterating over different number of topics until a number is found that produces both coherent topics and has a relatively high coherence score.

**References:**

[1]. World Health Organization. [https://covid19.who.int/](https://covid19.who.int/%20) (accessed 27 October 2021).

[2]. World Health Organization. [https://apps.who.int/gb/ebwha/pdf\_files/WHA73/A73\_R1-en.pdf](https://apps.who.int/gb/ebwha/pdf_files/WHA73/A73_R1-en.pdf%20) (19 May 2020).

[3]. McGill COVID19 Vaccine Tracker Team <https://covid19.trackvaccines.org/vaccines/> (accessed 27 October 2021)

[4]. MacDonald, N.E.; Sage Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. Vaccine 2015, 33, 4161–4164.

[5]. World Health Organization. Ten threats to global health in 2019. Accessed October 23, 2021. <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019>

[6].Paterson P, Meurice F, Stanberry LR, Glismann S, Rosenthal SL, Larson HJ. Vaccine hesitancy and healthcare providers. Vaccine. 2016;34(52):6700-6706.

[7]. Wang K, Wong ELY, Ho KF, et al. Intention of nurses to accept coronavirus disease 2019 vaccination and change of intention to accept seasonal influenza vaccination during the coronavirus disease 2019 pandemic: a cross-sectional survey. Vaccine. 2020;38(45):7049-7056.

[8]. Kwok KO, Li KK, Wei WI, Tang A, Wong SYS, Lee SS. Influenza vaccine uptake, COVID-19 vaccination intention and vaccine hesitancy among nurses: a survey. *Int J Nurs Stud*. 2021;114:103854.

[9]. Shekhar R, Sheikh AB, Upadhyay S, et al. COVID-19 vaccine acceptance among health care workers in the United States. *Vaccines (Basel)*. 2021;9(2):119.

[10]. Nguyen KH, Srivastav A, Razzaghi H, et al. COVID-19 vaccination intent, perceptions, and reasons for not vaccinating among groups prioritized for early vaccination — United States, September and December 2020. MMWR Morb Mortal Wkly Rep. 2021;70:217-222.

[11]. Hamel L, Kirzinger A, Muñana C, Brodie M. KFF COVID-19 vaccine monitor: December 2020. Kaiser Family Foundation. Dec. 15, 2020. Accessed April 23, 2021. <https://www.kff.org/coronavirus-covid-19/report/kff-covid-19-vaccine-monitor-december-2020/>