RSGC
Royal St. George's College

The Young Researcher

2022 Volume 6 | Issue 1

Analyzing Twitter Data to Understand COVID-19 Vaccine Hesitancy Following the Federal Private Sector Mandate

Manisha Palaniappan

Recommended Citation

Palaniappan, M. (2022). Analyzing Twitter data to understand COVID-19 vaccine hesitancy following the federal private sector mandate. *The Young Researcher*, 6(1), 106-119. http://www.theyoungresearcher.com/papers/palaniappan.pdf

ISSN: 2560-9815 (Print) 2560-9823 (Online) Journal homepage: http://www.theyoungresearcher.com

Analyzing Twitter Data to Understand COVID-19 Vaccine Hesitancy Following the Federal Private Sector Mandate

Manisha Palaniappan

Since the rollout of the COVID-19 vaccine, vaccine hesitancy has been prevalent in the US, particularly on social media such as Twitter. Significant efforts to understand the reasons behind anti-vaccine sentiments have been concentrated in the earlier stages of vaccine distribution and research suggests personal freedom to be a primary driver of vaccine hesitancy. This study aims to analyze anti-vaccine sentiment on Twitter following the first federal private sector mandate in November of 2021 to better understand how mandates impact vaccine sentiment in the later stages of the pandemic. Two machine learning models, roBERTa and LDA, were used to analyze a pre-curated Kaggle dataset of COVID-19 vaccine-related tweets. Results suggest that there is an overwhelmingly negative sentiment towards the vaccine. Reasons include efficacy and side-effect concerns, mistrust in institutions, and anti-mandate sentiments. This study identified common reasons behind vaccine hesitancy which future research can hopefully use to create targeted outreach programs.
Key Words: COVID-19, vaccine sentiment, vaccine hesitancy, machine learning

Introduction

COVID-19 has had devastating effects worldwide. According to the World Health Organization, at the time of writing, there have been close to 300 million confirmed cases, including more than 5 million deaths globally (World Health Organization, 2021). In response, there were extensive efforts to develop a vaccine with research beginning in January of 2020. Less than 11 months later, vaccine distribution started, making it the fastest developed vaccine in history (Kuter et al, 2021). Despite such rapid development, vaccination rates have not increased sufficiently mainly due to vaccine hesitancy. According to renowned Canadian physician and professor Noni McDonald, vaccine hesitancy is defined as:

"A delay in acceptance or refusal of vaccination despite availability of vaccination services. Vaccine hesitancy is complex and context-specific, varying across time, place, and vaccines. It is influenced by factors such as complacency, convenience, and confidence" (MacDonald, 2015).

Vaccine hesitancy is not a new phenomenon; in 2017, a measles outbreak occurred in Minnesota due to low vaccination rates (Petraco, 2019). However, vaccine hesitancy has been particularly widespread in the case of COVID-19 vaccines because of the political polarization that has been associated with it (Jiang et al, 2020). To combat low vaccination rates, the federal government implemented vaccine mandates. On November 4, 2021, the white house announced the first federal level vaccination mandate for the private sector (The United States Government, 2021). This

was met with considerable controversy, and on January 13, 2022 the Supreme Court blocked this mandate (Breuninger & Kimball, 2022). Even so, individual states as well as private businesses are still enforcing vaccinations and therefore, it is important to understand how these mandates play into vaccine hesitancy. Proponents of vaccine mandates argue that these mandates will normalize the COVID-19 vaccine and encourage people to get vaccinated

(Ashwell et al, 2021). However, there has still been much opposition from those who are vaccine hesitant. In order to investigate this phenomenon, social media has been used extensively to understand public perception, specifically in the case of the COVID-19 vaccine.

Social Media Analysis

Social media has proven to be a very useful tool in understanding public sentiment towards COVID-19 vaccines because so many Americans use it frequently. According to the Pew Research Center, over 70% of American adults use at least one social media site

(Pew Research Center, 2021). Specifically, Twitter has been used to understand public sentiment in regards to public health and infectious diseases due to low costs associated with this type of research as well as its effectiveness (Sinnenberg et al, 2017). With respect to the pandemic, it has fueled vaccine talk and opinions, which makes it a prominent tool in helping to analyze COVID-19 vaccine sentiment within the US (Featherstone et al, 2020a,b). Particularly, it has aided in understanding vaccine hesitancy. However, because of the extensive amounts of data provided on this platform, it becomes impossible to conduct analyses by hand. Instead, researchers frequently use machine learning models to understand vaccine sentiment and themes behind vaccine hesitancy.

Literature Review

Overview of Literature

Because of the controversial yet essential nature of the COVID-19 vaccinations, it is important to conduct research to understand vaccine sentiments and underlying reasons for vaccine hesitancy. By understanding the causes of vaccine hesitancy, a first step can be taken to finding effective ways to communicate the necessity of getting vaccinated to the unvaccinated population. Therefore, in order to comprehend this topic, researchers have been trying to study vaccine hesitancy since the vaccine was first manufactured (Boucher et al, 2021). However, results have been mixed. Some past research has suggested that negative vaccine sentiment has generally declined as time progressed (Hu et al, 2021). However, others have revealed that there is a fluctuation in negative sentiment, with it declining as positive vaccine efficacy news came out, but then declining afterwards despite no major correlation factors (Fazel et al, 2021). As such, it has been difficult to determine the level of negative public opinion towards COVID-19 vaccines. Furthermore, there has also been research conducted, especially through social media, to discern specific reasons for vaccine hesitancy with respect to COVID-19 vaccines. A review of literature regarding COVID-19 vaccine sentiment conducted at the Sultan Idris Education University revealed three major themes for vaccine hesitancy: concerns regarding the vaccine itself, mistrust in institutions, and individuals' social attributes

(Almoodi et al, 2021). These will be further discussed in the following sections.

Vaccine-related

The first major theme—also the largest one—is specifically vaccine-related. This includes efficacy of COVID-19 vaccines, people's confidence in them, and vaccine safety (Almoodi et al, 2021). One study led by Jean-Christophe Boucher, a reputable researcher at the University of Calgary used sentiment and semantic analysis of Twitter data using machine learning at the time when successful vaccine trials were first announced to understand vaccine hesitancy. They

found that there were many concerns online regarding harmful side effects of vaccines and also a lack of confidence that the vaccine would be effective in ridding the population of COVID-19 (Boucher et al, 2021). These reasons largely stem from the fact that the vaccine was created so quickly, and as such, many people who are vaccine hesitant are afraid that there may not have been enough time to properly create an effective vaccine (Paul et al, 2021). Advocates of the vaccine argue that this is untrue since the fast development was due to previous research accumulated over years, in addition to cooperation of world leaders and plenty of funding (Kuter et al, 2021). Overall, these studies were mostly conducted when the vaccine was first developed, and as such, there has been high uncertainty regarding the long-term impacts of the vaccine, though it is likely to change over time. Therefore, more research needs to be done on vaccine sentiment at the later stages of the pandemic when the vaccine has been more widely distributed.

Mistrust in Institutions

Another major theme present in previous literature regarding reasons for vaccine hesitancy is mistrust in institutions which includes mistrust of the media, international organizations (e.g., pharmaceutical companies), as well as the government—local and federal. In Boucher's social media analysis, he found that there was a large proportion of tweets which expressed criticism of the government and multinational corporations' actions to limit individuals' freedom to get vaccinated (Boucher et al, 2021). This, in turn, led many to be distrustful of large businesses such as those in the pharmaceutical industry and the government. Specifically, one longitudinal study conducted by Ariel Fridman looked at the influence of political parties on trust in the media and government. Results revealed that Republicans were more likely to have mistrust in the media and government, which was closely correlated with their lack of confidence in the COVID-19 vaccine, while the opposite was true for Democrats (Fridman et al, 2021). Thus, due to widespread mistrust, which is influenced by the politicization of the vaccine, forceful actions taken by the government to promote the vaccine, may have major negative consequences in regards to public perception of the vaccine.

Individuals' social attributes

Individuals' social attributes also play a big role in vaccine hesitancy. This ranges from one's socioeconomic status, level of education, to political affiliation. A study conducted by Chen Luo analyzed Chinese and American sentiment toward the COVID-19 vaccine through a cultural perspective. In general, US citizens' individualism, as opposed to the collectivism present in China, are one of the main drivers of vaccine hesitancy (Luo et al, 2021). While level of education and finances do not have a strong correlation with vaccine hesitancy, social cohesion and politicization, which are further influenced by the fact that Americans have more individualistic views, play an important role in vaccine hesitancy (Paul et al, 2021). More specifically, it has been revealed that liberals express the least amount of distrust and vaccine hesitancy, while the opposite is true of conservatives (Jiang et al, 2021). Because of the heavy influence of political polarization fueled by the Americans' individualistic mindsets, critical announcements made by political leaders are one of the major contributors to swaying public opinion on vaccines one way or another (Hu et al, 2021). Therefore, the vaccine mandate that President Biden implemented in November of 2021, could have had major impacts on public sentiment of the vaccine.

Although there has been research conducted on vaccine hesitancy in the past, it still needs to be investigated more thoroughly in order to get a more complete understanding of causes behind vaccine hesitancy. According to the Centers for Disease Control and Prevention (CDC) at the time of writing, close to 40% of the US population have not been fully vaccinated while there is still an average of more than 500,000 COVID-19 cases each week (Centers for Disease Control and Prevention, 2022). As such, it can be seen that vaccine hesitancy is still prevalent and needs to be more thoroughly investigated in order to implement efficient interventions to increase vaccination. Furthermore, because of the politicized nature of the COVID-19 vaccines, vaccine mandates, implemented to increase vaccination rates, may not be very effective. Specifically, there has been speculation that vaccine mandates could potentially be related to negative sentiments toward the vaccines themselves (Paul et al, 2021). Despite this, there has been little research exploring vaccine sentiment with respect to the fed-

eral vaccine mandates passed in the US. To address this gap, this study aims to look at American vaccine sentiment on Twitter, and particularly understand the reasons behind vaccine hesitancy at the time of the first private sector federal vaccine mandate in the US. Therefore, the question emerges: What are the reasons behind negative sentiment towards the COVID-19 vaccine as expressed by Americans on Twitter at the time when the first federal vaccine mandate for the private sector was implemented?

Methods

To address the question, similar to previous social media analysis research, this study also makes use of machine learning models in order to understand vaccine sentiment and hesitancy. Particularly, this research uses two pre-trained models which perform sentiment analysis and topic modeling. Sentiment analysis is the process of identifying the overall sentiment, while topic modeling analyzes major themes of a corpus. In accordance with the aim of this study, I used a sentiment analysis model to determine the overall sentiment of each tweet in the dataset and then proceeded to conduct topic modeling on only the negative tweets to understand reasons behind vaccine hesitancy.

Data Extraction and Cleaning

The Twitter data that was collected for this particular study came from between the dates November 04, 2021—when the federal private sector mandate was first passed—to December 04, 2021—a month after the mandate was implemented. Although researchers frequently use the Twitter API to extract their data, many restrictions such as number of tweets and lack of access to historical data, made this a poor choice for the requirements of this particular study. Instead, the data used for this particular study comes from Kaggle, a reputable website which consists of numerous pre-curated datasets that are easily accessible. This specific Kaggle dataset contains just over 730,000 English tweets with the keyword *covid_vaccine* curated from October 27, 2021 to December 31, 2021 (Ozturk, 2021). This study used the information columns of Tweet Content, User Location, and Date that were provided with the dataset. I first filtered out the rows which contained tweets from November 04, 2021 to December 04, 2021 and came from the United States since I am only focusing on American Twitter sentiment. I filtered out the tweets using a combination of the Microsoft Excel filtering function and Python code. Next, I had to clean my data by performing a series of preprocessing techniques—removing all at-mentions, hashtags, links, and special characters from my data—in order to have it in a format that my data analysis models can understand. To perform this cleaning process, I used a few specific python libraries, which include Pandas and Regex, in order to discard the irrelevant information.

Procedure/Analysis

After filtering and cleaning the data, the final dataset consisted of a little over 57,000 tweets. Due to the large amount of data, it was impossible to analyze it by hand. As such, machine learning and artificial intelligence (AI) was needed to analyze the data. Specifically, this research made use of natural language processing (NLP), a subfield of AI that deals with the interaction between computers and human language. Two different tasks within NLP were performed: sentiment analysis and topic modeling. Sentiment analysis is looking at the overall sentiment of each tweet, specifically understanding whether the tweet is positive, negative, or neutral. Afterwards, particularly focusing on the negative sentiment tweets, topic modeling was conducted to understand major topics being discussed amongst those who are vaccine hesitant. Topic modeling is a model that is used to discern prevalent themes or topics being discussed within large bodies of text.

The method used in this study was modeled after Chad A. Melton and others' research who analyze vaccine hesitancy on Reddit by sentiment analysis and topic modeling

(Melton et al). However, in contrast to to their use of the Textblob sentiment analysis model, this study uses the roBERTa pre-trained sentiment analysis model which was trained specifically on Twitter data and thus, has a high accuracy rate. Additionally, Latent Dirichlet Allocation (LDA) Topic Modeling through the Gensim library on Python was used as previous research has identified it to have the most relevant results for short-text data (Albalwi et al).

Results

Sentiment Analysis

After conducting Sentiment Analysis using the roBERTa model, there were three output values for each tweet. The output consisted of the percent positive, negative, and neutral sentiment that each tweet has on a scale from 0 to 1 (Figure 1).

```
1) neutral 0.5942
2) negative 0.3839
3) positive 0.0219
```

Figure 1: This is a sample output for a tweet. Each value is on a scale from 0 to 1. Neutral+Negative+Positive=1

Sentiment Analysis revealed that the majority of tweets in my dataset, 58.3% (n=33,679), were negative. The second largest group was neutral at 33.2% (n=19,166), and the smallest was positive at only 8.5% (n=4,937) (Figure 2).

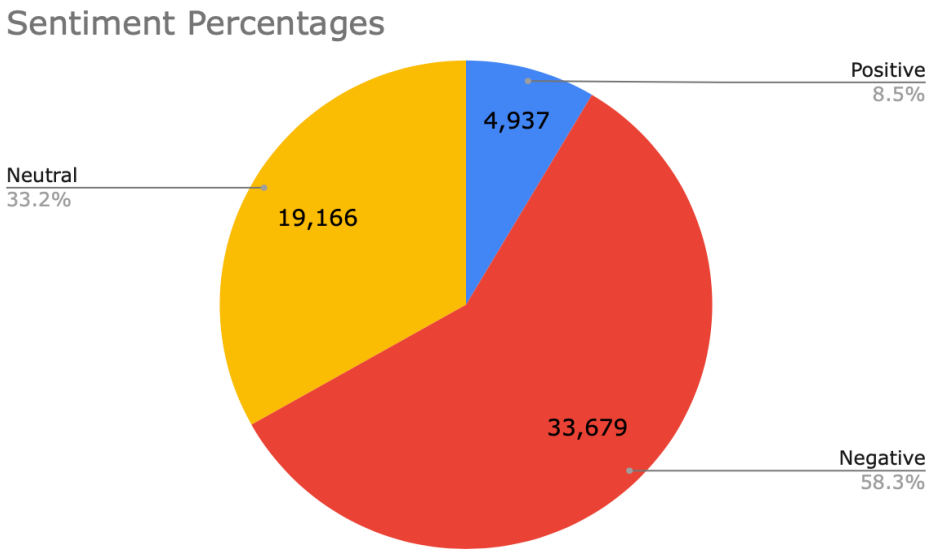


Figure 2. Using the roBERTa model, sentiment analysis results showed that a majority of tweets were negative, with neutral being the second largest, and positive being the smallest group.

The mean and variance of scores for positive, negative, and neutral tweets were also calculated (Table 1). It can be seen that negative and positive tweets had similar score means which indicate that they had relatively the same strength of sentiment while it was a little lower for neutral. It also seems that positive and negative tweets have similar variances while, again, the neutral tweet scores have a lower variance.

Table 1

	Negative	Positive	Neutral
Mean	0.74	0.72	0.63
Variance	0.020	0.026	0.016

Table 1. Means for the negative and positive tweets were relatively similar with the score for neutral being a little lower. Variance was overall very similar for all three groups, but with neutral again being a little lower than the other two.

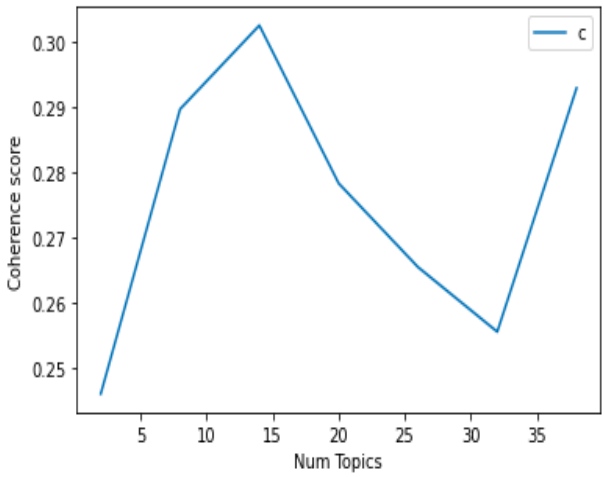


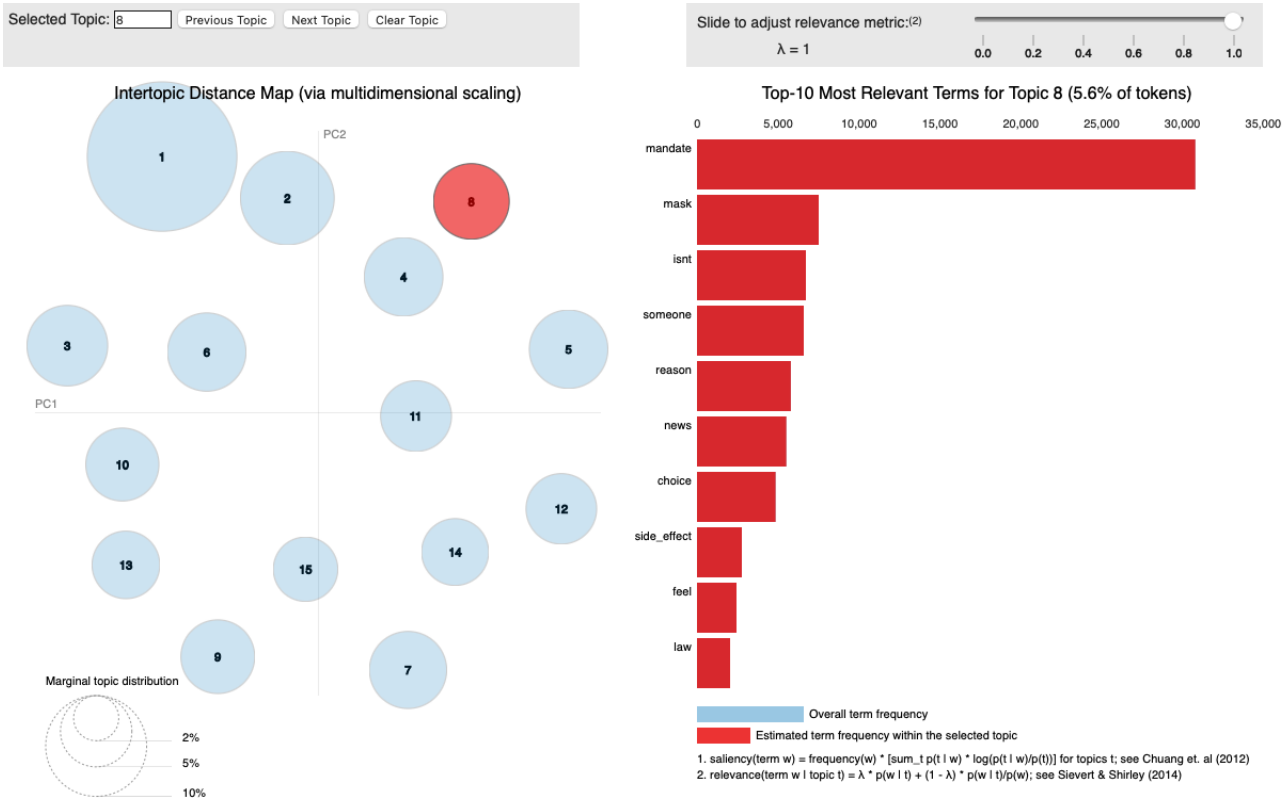
Figure 3. Graph of number of topics vs. topic coherence score. The optimal number of topics can be seen at 15, the highest peak. The coherence score at this point is around 0.30.

Figure 4. Example of "Topic 8" is shown. To the left, the circles represent each topic with the size of the circle denoting its relevance to the corpus; the larger the circle, the more relevant. To the right, since we have selected Topic 8, the top 10 words in this topic are displayed.

Latent Dirichlet Allocation (LDA) Topic Modeling

The sentiment analysis was followed by LDA topic modeling. This was used to understand specific topics being discussed in the negative tweets to better analyze vaccine hesitancy. The LDA model yielded a total of 15 latent topics. This number, which was a parameter passed into the LDA model, was determined using the topic coherence score. Topic coherence measures the semantic similarity between top words in a topic to determine how closely related they are. Figure 3 shows a graph which plots the number of topics against the coherence score. From this, it can be seen that the highest topic coherence score is at 15 topics and therefore, this was the optimal number to use for the LDA model.

The LDA model essentially generates multiple topics, each of which represent a group of words within the corpus. In this case, for each of the 15 topics, the first 10 words were also displayed (Figure 4). However, the overarching themes for each were not immediately recognizable. To interpret the themes, I looked at specific tweets that contained a few of the



	Overall Interpreted Theme	Top 10 Words
Topic 1	Child immunization and side-effects	vaccine, covid, death, child, keep, risk, state, trump, myocarditis, injury
Topic 2	Side-effects and mistrust in pharmaceutical companies	die, anyone, number, show, flu, money, lot, vax, school, drug
Topic 3	Efficacy and mistrust in pharmaceutical companies	fact, spread, life, country, family, nothing, month, booster, world, know
Topic 4	Efficacy and mistrust in government and pharmaceutical companies	don't, year, infection, vaccination, see, protection, end, study, natural_immunity, governor
Topic 5	Mistrust in government and Work-related	get, everyone, didn't, cause, let, they're, age, worker, tweet, kind
Topic 6	Mistrust in government and Work-related	virus, prevent, case, anything, lie, kill, won't, tell, guy, employee
Topic 7	Efficacy	doesn't, work, you're, person, protect, help, source, hospital, start, talk
Topic 8	Side-effects and Anti-vaccine mandates	mandate, mask, isn't, someone, reason, news, choice, side_effect, feel, law
Topic 9	Mistrust in pharmaceutical companies and work-related	amp, pfizer, health, immunity, rate, information, business, jab, question, folk
Topic 10	Mistrust in science and news	time, way, issue, problem, look, medium, treatment, article, ppl, research
Topic 11	Side-effects and Mistrust in government	thing, can't, go, today, trust, cdc, report, cancer, week, none
Topic 12	Child immunization and Side-effects	kid, government, doctor, body, need, hospitalization, parent, remember, happen, response
Topic 13	Mistrust in scientific community and Anti-vaccine mandates	take, day, something, think, care, science, freedom, refuse, community, line
Topic 14	Mistrust in news and scientific community	stop, shot, disease, point, misinformation, post, scientist, yesterday, gonna, consequence
Topic 15	Efficacy	i'm, part, make, word, bc, god, dr, wouldn't, save_life, lack

Table 2. This table lists the interpreted theme and the first 10 words for each of the 15 topics. Efficacy, mistrust, side-effects, and anti-vaccine mandate sentiment themes seem to be the most common.

	Tweet Examples
Efficacy	"For the millionth time, vaccines do not prevent you from getting Covid. Vaccines do not prevent you from spreading Covid & it is becoming very clear that the vaccines don't prevent you from dying from Covid. Don't believe me? Look at the number of breakthrough cases & deaths in MA."
Side-effects	"Yet there are people dying from the vaccine and getting serious side effects and we do not know the long term side effects."
Mistrust	"The main data from which [vaccine proponents] form their opinion are all from CDC, which is in the business of pushing vaccines at the behest of Big Pharma under the guise of a public health agency. Do you not see the conflict of interest?"
Anti-mandate sentiment	"Everytime a new mandate or mask/vax policy is introduced the longer I will refuse to take the vaccine." "End restrictions. All of them. It's over. Covid is endemic. The vaccines didn't work. It is long past time to shift from failed government authoritarianism to personal choice."

Table 3. The four major themes that were displayed from the 15 topics were efficacy, side-effects, mistrust, and anti-mandate sentiment. This table provides specific examples of tweets illustrating each of these themes.

top words (as generated by the model) in each topic. Table 2 displays the results for each topic. From the interpreted themes of all 15 topics, four major concepts emerged. These were: concerns regarding efficacy and side-effects, mistrust in institutions, and strong anti-mandate sentiments. This was evidenced by multiple tweets in the dataset during qualitative analysis. Examples of tweets that portray each of the four ideas are given in Table 3.

From the fifteen topics, four major themes emerged. The first theme was composed of four topics out of the fifteen (26.67%) and discussed concerns regarding efficacy. The second theme of side-effects consisted of four topics (26.67%) with two of these specifically focused on child immunization. The third major theme, made up of 10 topics (66.67%) was mistrust in institutions. Within this theme, 40% of topics were mistrustful of the government, 40% mistrustful of the pharmaceutical industry, and 30% were mistrustful of the news and scientific community at large. The final theme of anti-vaccine mandate sentiment was composed of five topics (33.33%) with three of these focused on specifically work-related mandates. Percentages do not add up to one hundred since some topics reflected multiple themes.

Discussion

Results indicate that there was an overwhelming amount of negative sentiment regarding the COVID-19 vaccine at the time of the mandate on Twitter whereas positive sentiment made up only a small fraction of tweets. Prior research conducted from March of 2020—the beginning of the pandemic—to January of 2021 has indicated that negative vaccine sentiment displayed on Twitter has generally declined as time progressed (Lyu et al, 2021). However, this study suggests this may not really be the case and instead, negative vaccine sentiment may still be prevalent in the later stages of the pandemic. Therefore, the phenomenon of vaccine hesitancy is only becoming more important to understand with regards to the mandate and how discourse over the mandate manifests on social media platforms like Twitter.

Qualitative analysis of the LDA topics revealed 4 overarching themes. These were concerns regarding the efficacy of the vaccine, side-effects, mistrust in institutions, and anti-vaccine mandate sentiments. Most previous literature on vaccine sentiment has been conducted in the earlier stages of vaccine rollout (Hu et al, 2021). Because of the rapid development of the vaccine, past research has identified efficacy to be one of the biggest drivers of vaccine hesitancy (Paul et al). This study, which used Twitter data from the later months of 2021, almost a year after vaccines were first distributed, still found that efficacy remains one of the major drivers of vaccine hesitancy. Previous research found that efficacy was a reason behind vaccine hesitancy, but was unable to identify explicitly where this sentiment stemmed from. However, further analysis of data from this study revealed breakthrough infections to be a leading cause of vaccine hesitancy. Because there have been many particular instances where people who were vaccinated still contracted the virus, there has been increased speculation regarding efficacy. Furthermore, earlier research has indicated uncertainty regarding side-effects to be a factor for anti-vaccine sentiments (Boucher et al, 2021). As evidenced by the tweet associated with side-effects in Table 3, this concern over side-effects has also been prevalent in this study's results. In addition, the results of this study are unique in that they show there is an increased emphasis on potential side-effects for children who are vaccinated. This is particularly evident in topics 1 and 12. A potential reason for this could be because of the increasing pressure to get children vaccinated and more widespread availability of the vaccine for younger ages. As news sources have highlighted, later variants of COVID-19 such as Omicron have resulted in much greater hospitalization rates for children (Kozlov, 2022). As such, a larger emphasis on child vaccination could have ensued due to this risk.

Mistrust in institutions that include the government, the news, pharmaceutical companies, and the scientific community has played a large part in the negative sentiments. From analysis, it seems that mistrust of the government and the pharmaceutical industry is most prevalent. Because there has been an increased association between the private sector and the federal government through funding of vaccine development in addition to the mandates being imposed, this mistrust has only grown. The tweet associated with the theme of mistrust in Table 3 clearly illustrates this issue by discussing the possibility of a conflict of interest between the government and pharmaceutical companies. While in the past, there has been a greater emphasis on mistrust of the government and the pharmaceutical industry, this study identifies mistrust of the media and the scientific community to be present as well, although to a lesser extent.

Since this study was conducted in the one-month time frame right after the vaccine mandate for the private sector was announced, there is a larger emphasis on personal freedom and choice when it comes to the vaccine. In particular, topics eight (Figure 4) and thirteen are particularly interesting. The overarching theme for those topics conveyed that many are against the vaccine *mandates* in particular, which could potentially be a driving force behind their anti-vaccine sentiments. Prior studies have speculated that negative vaccine sentiment could potentially be correlated with negative *mandate* sentiment (Paul et al, 2021). This has been explicitly identified in my results and can be evidenced through the example tweets for anti-mandate sentiment in Table 3. Moreover, research has indicated that confidence in the vaccine itself can influence support for mandates (Ashwell et al, 2021). Therefore, it comes as no surprise that there has been such strong negative sentiment towards the federal mandate, since there has been a lot of uncertainty regarding the efficacy and many concerns regarding side-effects.

Moreover, one study wherein the researchers analyzed twitter data proposes that the individualistic mindset of Americans as opposed to a communal mindset in countries like China have played an important role in whether the people got vaccinated or not (Luo et al, 2021). Americans tend to be much more self-reliant and in health-related contexts such as this one, they wanted to have personal control over health risks. This could explain the results from this research because this particular study was solely focused on American sentiment particularly when the mandate was imposed. Since Americans have a more individualistic mindset, they value personal choice and by imposing the mandate it could have potentially had a negative emotional impact on many.

Topics 6 and 9 are interesting in that they encompass sentiments wherein vaccine hesitancy has been displayed in the context of the workplace. Specific references to employers, workers, and businesses have shown that vaccination requirements at work could be potential reasons for vaccine hesitancy as well. Previous research has not identified implications of mandates in businesses. Because this federal mandate has been centered around the private sector, it has likely fueled concerns among the vaccine hesitant with regards to how their income and work life will be affected.

Conclusion

Sentiment analysis of Twitter data following the announcement of the first federal private sector COVID-19 vaccine mandate in the US showed that there was overwhelming negative sentiment towards the vaccine. LDA portrayed 15 topics that were then qualitatively analyzed and broken down into four major themes which consisted of concerns regarding efficacy, side-effects, mistrust, and personal freedom. Overall, my results are in alignment with Almoodi and others' identified themes of the causes of negative vaccine sentiment

(Almoodi et al, 2021). Ultimately, this research has been effective in building a greater understanding of the unvaccinated perspective which will aid in targeted efforts at increasing vaccination.

Limitations

In conducting this research study, it is important to acknowledge the limitations that may have impacted the work done. For this study, twitter data taken from a Kaggle dataset was used. Kaggle is a very reputable source from which many data scientists and researchers get their data. However, it is pre-curated by others and as such, I do not have the means to definitively identify the methods through which the data was scraped. As such, there could be implicit biases inherent in the types of tweets collected. Furthermore, the specific sentiment analysis model that was used in this study called roBERTa has one of the highest accuracy rates when it comes to pre-trained sentiment analysis models. Despite this, since I had neither the capabilities nor the time to train my own model, sentiment analysis was not completely accurate. Specifically, in conducting the qualitative analysis, it was revealed that in certain cases, sentiment analysis could have reflected the underlying connotation of the tweets rather than their attitude towards the COVID-19 vaccine. Thus, this resulted in some negatively labeled tweets actually being positive. For example, one tweet reads: "*I don't push vaccines because I enjoy the fight. I don't push them because I hate freedom. I push them so I don't have to see another person I know or take care of die from something easily preventable.*" This tweet received a negative score of 0.7482.

Since the Natural Language Processing sector of AI and Machine Learning is in its beginning phases of development, the LDA model, although very useful in analyzing topics, can be difficult to interpret using only the top words in each topic. Specifically, the ability to understand the overarching theme of a topic right away is sometimes difficult. Due to this, qualitative analysis was also conducted. Because it would be physically impossible to go through each tweet and qualitatively analyze it, I was only able to look at a sample of a few of these. Therefore, it may not capture the full picture in regards to topics relating to vaccine hesitancy.

Implications

The findings of this research have assisted in furthering the understanding of vaccine hesitancy. As

mentioned previously, research such as Boucher and others' study has focused on the earlier stages of vaccine rollout (Boucher et al, 2021). However, only around 65% of the US population have been fully vaccinated to date while herd immunity is estimated to be achieved only when 70-85% of the population are vaccinated. As such, vaccine hesitancy remains a large issue. The findings of this study are important to understand anti-vaccine perspectives. Only through recognizing the opposing points of views, will government officials, scientists, and news agencies be able to take effective measures to further combat low vaccination rates. Aside from vaccine mandates, other interventions have also been implemented to varying degrees of success. Previous research has revealed interventions such as offering monetary incentives to get vaccinated have had positive impacts on vaccination rates (Mercade et al, 2021). On the other hand, news sources have highlighted the fact that advertisements regarding vaccination on television have not been very successful in increasing vaccination rates nationally (Christensen). Based on the results from this study, it seems that vaccine mandates are not very effective in encouraging people to get vaccinated. Previous research has indicated that implementing mandates in the workplace could normalize the vaccine, thus increasing vaccination rates (Ashwell et al). However, it seems that in reality, this may not be the case, based on the results of this study. Therefore, mandates may not be the best approach to tackle the issue of vaccine hesitancy. Although the federal vaccine mandate has been blocked by the Supreme Court, many states and employers have enforced rules regarding vaccination. Since there has been strong opposition to federal vaccine mandates, there is a very likely possibility that there will also be negative sentiments towards state government and workplace mandates as well. In addition, based on the results, it is important to note that continued efforts in displaying accurate information regarding side-effects and efficacy rates are important to combat rampant disinformation. Moreover, it is crucial that greater attempts be made to build more trust between the people and government, pharmaceutical companies, and the scientific community. It is likely that decreased political polarization and more affirmations of unity are essential to this endeavor.

Future Directions

While this study focused on Twitter data from the US nationally, future research could focus on a comparison of vaccine hesitancy between states with lower vaccination rates and those that have higher vaccination rates. This could allow for a more focused analysis of causes behind anti-vaccine sentiment. It would aid in improving vaccination rates of states with lower vaccination rates through implementing effective methods that have been used in states with higher vaccination rates. Further, this study used pre-curated Twitter data from Kaggle. Future works could use Twitter data scraped directly from the Twitter API to get a more comprehensive and specified list of tweets to analyze. Not only that, but researchers could compare tweets from different demographics which could provide greater insight into views of different age groups. Overall, more research will need to be conducted in order to develop effective, targeted outreach programs to combat low vaccination rates.

References

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, *3*, 42. https://doi.org/10.3389/frai.2020.00042

Alamoodi, A. H., Zaidan, B. B., Al-Masawa, M., Taresh, S. M., Noman, S., Ahmaro, I. Y. Y., Garfan, S., Chen, J., Ahmed, M. A., Zaidan, A. A., Albahri, O. S., Aickelin, U., Thamir, N. N., Fadhil, J. A., & Salahaldin, A. (2021). Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy. *Computers in Biology and Medicine*, *139*, 104957. https://doi.org/10.1016/j.compbiomed.2021.104957

Ashwell, D., Cullinane, J., & Croucher, S. M. (2021). Vaccine hesitancy and support for employer vaccine mandates. *Frontiers in Communication*, *6,* 10. https://doi.org/10.3389/fcomm.2021.780415

Boucher, J.-C., Cornelson, K., Benham, J. L., Fullerton, M. M., Tang, T., Constantinescu, C., Mourali, M., Oxoby, R. J., Marshall, D. A., Hemmati, H., Badami, A., Hu, J., & Lang, R. (2021). Analyzing social media to explore the attitudes and behaviors following the announcement of successful COVID-19 vaccine trials: Infodemiology study. *JMIR Infodemiology*, *1*(1), e2880. https://doi.org/10.2196/28800

Breuninger, K., & Kimball, S. (2022, January 14). *Supreme Court blocks Biden Covid vaccine mandate for businesses, allows health-care worker rule*. CNBC. Retrieved April 22, 2022, from https://www.cnbc.com/2022/01/13/supreme-court-ruling-biden-covid-vaccine-mandates.html

Campos-Mercade, P., Meier, A. N., Schneider, F. H., Meier, S., Pope, D., & Wengström, E. (2021). Monetary incentives increase covid-19 vaccinations. *Science*, *374*(6569), 879–882. https://doi.org/10.1126/science.abm0475

Centers for Disease Control and Prevention. (2022). *CDC Covid Data tracker*. Centers for Disease Control and Prevention. Retrieved January 12, 2022, from https://covid.cdc.gov/covid-data-tracker/?utm_campaign=Private%20Equity%20Newsletter&utm_source=hs_email&utm_medium=email&_hsenc=p2ANqtz-_9OSt7jgR0ag2vPQVMdjmHOJhtBfAKqb3OsX-S0Pr1RdsT3Wm9YvOe0XcLivUaeChn93xw#vaccinations_vacc-people-onedose-pop-5yr

Chandrasekaran, R., Mehta, V., Valkunde, T., & Moustakas, E. (2020). Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. *Journal of Medical Internet Research*, *22*(10), e22624. https://doi.org/10.2196/22624

Christensen, J. (2021, August 6). *Sharp decline in ads for covid-19 vaccines, as the number of 'persuadable' Americans diminishes*. CNN. Retrieved May 2, 2022, from https://www.cnn.com/2021/08/06/health/covid-19-ads-why-theyve-gone-away/index.html

Fazel, S., Zhang, L., Javid, B., Brikell, I., & Chang, Z. (2021). Harnessing twitter data to survey public attention and attitudes towards covid-19 vaccines in the UK. *Scientific Reports*, *11*(1), 1-5. https://doi.org/10.1038/s41598-021-02710-4

Featherstone, J. D., Barnett, G. A., Ruiz, J. B., Zhuang, Y., & Millam, B. J. (2020). Exploring childhood anti-vaccine and pro-vaccine communities on Twitter – a perspective from influential users. *Online Social Networks and Media*, *20*, 100105. https://doi.org/10.1016/j.osnem.2020.100105

Featherstone, J. D., Ruiz, J. B., Barnett, G. A., & Millam, B. J. (2020). Exploring childhood vaccination themes and public opinions on Twitter: A Semantic Network analysis. *Telematics and Informatics*, *54*, 101474. https://doi.org/10.1016/j.tele.2020.101474

Fridman, A., Gershon, R., & Gneezy, A. (2021). Covid-19 and vaccine hesitancy: A longitudinal study. *PLOS ONE*, *16*(4), e0250123. https://doi.org/10.1371/journal.pone.0250123

Hu, T., Wang, S., Luo, W., Zhang, M., Huang, X., Yan, Y., Liu, R., Ly, K., Kacker, V., She, B., & Li, Z. (2021). Revealing public opinion towards covid-19 vaccines with Twitter data in the United States: Spatiotemporal Perspective. *Journal of Medical Internet Research*, *23*(9), e30854. https://doi.org/10.2196/30854

Jiang, J., Chen, E., Yan, S., Lerman, K., & Ferrara, E. (2020). Political polarization drives online conversations about COVID -19 in the United States. *Human Behavior and Emerging Technologies*, *2*(3), 200–211. https://doi.org/10.1002/hbe2.202

Koslap-Petraco, M. (2019). Vaccine hesitancy: Not a new phenomenon, but a new threat. *Journal of the American Association of Nurse Practitioners*, *31*(11), 624–626. https://doi.org/10.1097/jxx.0000000000000342

Kozlov, M. (2022, February 4). *Does Omicron Hit Kids harder? scientists are trying to find out*. Nature News. Retrieved May 2, 2022, from https://www.nature.com/articles/d41586-022-00309-x

Kuter, B. J., Offit, P. A., & Poland, G. A. (2021). The development of covid-19 vaccines in the United States: Why and how so fast? *Vaccine*, *39*(18), 2491–2495. https://doi.org/10.1016/j.vaccine.2021.03.077

Luo, C., Chen, A., Cui, B., & Liao, W. (2021). Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective: Semantic network analysis of two social media platforms in the United States and China. *Telematics and Informatics*, *65*, 101712. https://doi.org/10.1016/j.tele.2021.101712

Lyu, J. C., Han, E. L., & Luli, G. K. (2021). Covid-19 vaccine–related discussion on Twitter: Topic Modeling and sentiment analysis. *Journal of Medical Internet Research*, *23*(6), e24435. https://doi.org/10.2196/24435

MacDonald, N. E. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, *33*(34), 4161–4164. https://doi.org/10.1016/j.vaccine.2015.04.036

Melton, C. A., Olusanya, O. A., Ammar, N., & Shaban-Nejad, A. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *Journal of Infection and Public Health*, *14*(10), 1505–1512. https://doi.org/10.1016/j.jiph.2021.08.010

Oktay Ozturk. (2021). 800k+ COVID-19 Vaccine Tweets [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/2992050

Paul, K. T., Eberl, J.-M., & Partheymüller, J. (2021). Policy-relevant attitudes toward covid-19 vaccination: Associations with demography, Health Risk, and social and political factors. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.671896

Pew Research Center. (2021, April 7). *Demographics of social media users and adoption in the United States*. Pew Research Center: Internet, Science & Tech. Retrieved January 12, 2022, from https://www.pewresearch.org/internet/fact-sheet/social-media/#social-media-use-over-time

Saha, K., Torous, J., Caine, E. D., & De Choudhury, M. (2020). Psychosocial effects of the COVID-19 pandemic: Large-scale quasi-experimental study on social media. *Journal of Medical Internet Research*, *22*(11), e22600. https://doi.org/10.2196/22600

Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for Health Research: A Systematic Review. *American Journal of Public Health*, *107*(1), e1-e8. https://doi.org/10.2105/ajph.2016.303512

The United States Government. (2021, November 4). *Fact Sheet: Biden Administration Announces Details of Two Major Vaccination Policies.* The White House. Retrieved January 12, 2022, from https://www.whitehouse.gov/briefing-room/statements-releases/2021/11/04/fact-sheet-biden-administration-announces-details-of-two-major-vaccination-policies/

World Health Organization. (2021). *Who coronavirus (COVID-19) dashboard*. World Health Organization. Retrieved January 12, 2022, from https://covid19.who.int/