Process Book

Abstract

(introduction to the dataset and our main research question)

Image(vaccine)

1. # Overview of Tweet Content

What Do People Tweet about Covid-19 Vaccines?

Data: Text content of all the tweets about Covid-19 vaccines.

Process: Clean and Lemmatization text data of tweets about vaccines, and use the hunspell package to get stems as complete words. Make a word cloud to present most commonly mentioned keywords.

Conclusion: This word cloud includes the popular keywords (appeared more than 600 times) used in tweeting about Covid-19 vaccine.

The most commonly mentioned word is of course “vaccine”, followed by “moderna” and “covid”, while “pfizer” and “pfizerbiontech” are much smaller.

We can also see the variants of alias of Covid-19 vaccine (“covaxin”, “covidvaccine”), and keywords used commonly in describing experiences (“dose”, “receive”, “today”) mentioned for many times.

A Co-occurrence Network of Keywords

Data: Keywords extracted from last step.

Process: Keep only the top 1000 keywords in terms of occurrence and calculate the times of co-occurrence for each pair of word. Keep only the ties indicating more than 100 co-occurrences, and make a network to exhibit the connections between those commonly mentioned keywords. Walk trap algorithm is used to detect word clusters.

Conclusion: There are three major clusters detected – one is the major group with two central points – “vaccine” and “covid”; another is one around “moderna” the manufacturer, which may come from tweets reporting new progresses of moderna vaccine; the other one is more dispersed with three centers – “today”, russia” and “antario”, which may come from those focus on vaccine exportation news. The clusters are interweaved together, but can offer some hits on different popular topics. Readers can freely explore the network and look for the relevant words they are interested in.

2. #Sentiment Analysis

Sentiment State Distribution of Tweets

Process: Clean and Lemmatization text data of tweets about vaccines. Vader sentiment analysis.

Conclusion: 1.Most tweets about vaccines are neutral one or positive one. Negative sentiment is not widely available.

2.Trends over time of numbers of tweets posted of three sentiment types are similar.

3.After seeing the common words of positive, neutral and negative tweets, we find people share their happiness about the arrival of vaccines and give positive feedback after receiving a shot in positive tweets; neutral tweets are just objective statements of vaccines news or information; people worry about the side effect of vaccines and whether vaccines will work in negative tweets.

Portrait of Popular Tweets and Users

To study the possible influence of sentiment attribute of popular tweets and users on twitter, we made these portraits.

Data: Base on classification result of sentiment analysis

Conclusion: 1.We can see sentiment attributes of popular tweets(based on favorite times). Most of the top 15 popular tweets are neutral one or positive one, which means that people didn't show a preference for negative tweets.

2.We can see the sentiment attribute of popular tweets(based on retweeted times). Most of the top 15 popular tweets still are neutral one or positive one, which means that people didn't show a preference for retweeting negative tweets and maybe kept a positive attitude towards the effect of vaccines.

3.We also check tweets of popular users(based on number of followers they have), because they have great influence among the public. Most of these users mainly post objective statements of vaccines news or information and they show more positive sentiment than negative sentiment.

3. # Tweets about Covid-19 Vaccine’s Side Effects

Proportion of tweets about side effects

Data: The whole tweet dataset.

Process: Tweets that mention “side effect” or “side effect” are selected to make a comparison between its number and that of all the tweets.

Conclusion: Side effect is not really a heated topic among the discussions on twitter. Only 1/50 of all the tweets about Covid-19 vaccine mention side effects.

What Do They Tweet about Covid-19 Vaccine’s Side Effects?

Data: Text content of tweets about side effects.

Process: Tweets that mention “side effect” or “side effect” are selected, whose text content is cleaned and stemmed using the hunspell method, then form a word cloud to present the most commonly mentioned keywords. Given the dataset shrinks a lot when we restrict it to side effect discussions, word cloud covers words appeared more than 20 times.

Conclusion: The most commonly mentioned word, in this case, is of course “side effect”, followed by “moderna” and “vaccine”, while “pfizer” and “pfizerbiontech” are much smaller, in line with what we found in all of the Covid-19 vaccine tweets.

We can also see the keywords “covid”, “day”, “shot”, “arm”, “dose”, which indicates many of these tweets may be records of people’s vaccination experience.

Words like “sore”, “fatigue”, “pain” and “headache” indicate the commonly mentioned side effects in tweets.

A closer look at the two manufacturers

Data: Occurrences of the manufacturer

Process: Add up the occurrences of “pfizer” and “pfizerbiontech” to indicate the volume of Pfizer vaccines, and compare with the volume of Moderna vaccines. Proportion of side-effect-relevant tweets in all tweets mentioning them are calculated and compared.

Conclusion: Moderna has larger volume in both all tweets and tweets about side effects, compared to Pfizer. Looking at the proportion of side-effect-relevant tweets in all tweets for each manufacturer, moderna still scores higher than pfizer, which means it may be more likely to trigger side effects. Just a thought to be tested for in later analyses.

4. # Who are the people reporting adverse reactions?

AGE AND GENDER

Do elders suffer more from side effects? Not exactly

Conclusion: Women and Younger people tend to report more cases. Is it possible that fewer elders got vaccinated thus fewer reports?

We decided to dive deeper into who got vaccinated by looking at different age groups.

Vaccinated Rate by Different Age Group

Figure:(by age group) Percentage of People that Have Received at Least One Dose of Cov Vaccine , by Mar 31, 2021

Type: plotly interactive line chart

source : https://covid.cdc.gov/covid-data-tracker/#datatracker-home

Report rate by different age group (animated bar? to show the changes by time)

Pre-illness

most common illness (wordcloud) (allergy is actually a big category, what are the common allergies mentioned?)

common allergies (there are all kinds of allergies containing other allergy, eg: food allergy, nut allergy or words in reverse order, eg: penicillin allergy, allergy penicillin. Don't really know how to handle it .give up ⚠️)

5.#When did side effects kick in ?

(bar chart, taking average value for different age group and sex)

(tried line chart, but it took extreme values into consideration, the displaying is not very nice. I guess average makes more sense, instead of showing value for every age)

6.#What are the side effects symptoms?

top 10 common symptoms (bar chart)

##emotional words in symptoms description(how do people feel) (facet)

7.# Where are the reports from?

allocation rate by state

This part uses a bar chart and map to show the allocation of vaccines of different brands in different states before 2021-03-31.

Source:

https://data.cdc.gov/Vaccinations/COVID-19-Vaccine-Distribution-Allocations-by-Juris/saz5-9hgg

https://data.cdc.gov/Vaccinations/COVID-19-Vaccine-Distribution-Allocations-by-Juris/b7pe-5nws

Conclusion:

The number of vaccine allocations in each state does not have a brand tendency. In every state, the number of vaccine allocations for the two brands is basically the same.

report rate by state

This part uses a bar chart and map to show the side effect case report rate (from 2020-12-14 to 2021-03-31) in 50 states and the District of Columbia and the relationship between report rate and 2020 election result.

The report rate was calculated by dividing the number of cases in the VARES data set by the number of people vaccinated in the daily administered data set.

Source:

https://covid.cdc.gov/covid-data-tracker/#datatracker-home

VARES data set: https://www.kaggle.com/ayushggarg/covid19-vaccine-adverse-reactions

The results of the state elections come from the data provided in the Week 5 lecture.

Conclusion:

The report rate in the northeast and northwest regions is higher than other regions. The reporting rate of vaccine side effects in each state does not seem to be significantly

related to the party’s victory in the 2020 election. But New York has the highest reporting rate, more than double that of Montana, the second highest.

Cases by manufacturer

This part uses bar chart and map to show the number of reported side effect cases of different brands of vaccines (from 2020-12-14 to 2021-03-31) in 50 states and the District of Columbia

Source:

VARES data set: https://www.kaggle.com/ayushggarg/covid19-vaccine-adverse-reactions

Conclusion:

According to previous visulizations, there is no significant difference in the number of vaccine allocations between Moderna and Pfizer in each state. However, Pfizer’s vaccine has more reported cases of side effects.