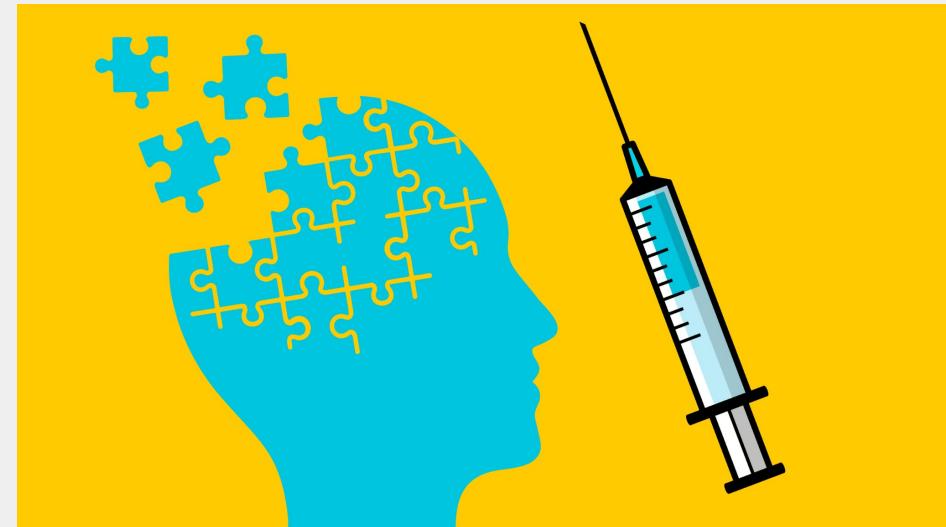


# Opinion Mining on Coronavirus Vaccine Hesitancy Using Twitter

Capstone Defense



# Presentation Layout

- Introduction
- Research Questions
- Specific Aims and Objectives
- Datasets
- Approach and Methodologies
  - Building Sentiment Analysis Models
  - Evaluating the models
  - Constructing Datasets
- Applying the model
- Analysis and Discussion
  - Monthly Trends
  - Reasoning Extraction
- Conclusion
- Future Work



# Introduction

- Vaccine Hesitancy is defined as the delay in acceptance or refusal of vaccination despite the availability of vaccination services.
- The World Health Organization classified vaccine hesitancy as one of the ten threats to global health in 2019.
- Since the start of the Covid-19 pandemic, the negative sentiment over Covid Vaccines have been on the rise.



# Research Questions

- Can we develop a sentiment analysis model that would successfully allow us to capture trends on the hesitancy of the people over taking the coronavirus vaccine?
- Which ML algorithm would work best for this problem?
- Can we analyze the hesitancy over different types of vaccines with high accuracy via ML-based sentiment analysis?
- Can we visualize such trends?



# Aims and Objectives

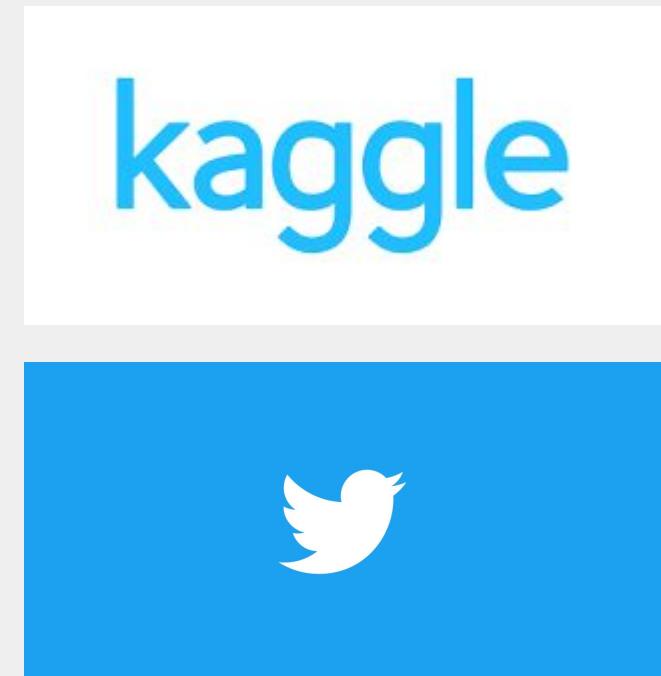
Examining and understanding vaccine hesitancy through twitter is important because:

- Contributes to the control and elimination of the virus and the spread of misinformation.
- Helps the science community by finding direct reasons on reasons behind such sentiment.
- Twitter is important social media platform to understand sentiment and thus knowing the sentiment on twitter could reflect real world sentiment.



# Datasets

1. Annotated DataSet: An existing dataset that consists of 1.6 million tweets including their sentiment, which we used to train and test the ML classifiers.
2. Covid-19 Vaccine-Related Tweets: We constructed a dataset that consist of 628,582 tweets in order to study the hesitancy of taking Covid-19 Vaccines.



# Approach and Methodologies

Building Sentiment Analysis  
Models

Evaluating the  
Models

Constructing  
Dataset

## Approach and Methodologies (Building Sentiment Analysis Models)

## Preprocessing Tweets:

- Replacing full urls with “URL”
  - Replacing Emojis with their sentiment
  - Removing non-alphabets
  - Removing Consecutive letters
  - Removing short words
  - Removing stop words

## Analyzing the Data:

- Figure 1 refer to negative word cloud
  - Figure 2 refers to positive word cloud

## Splitting the Dataset:

- 80% training set and 20% testing set.

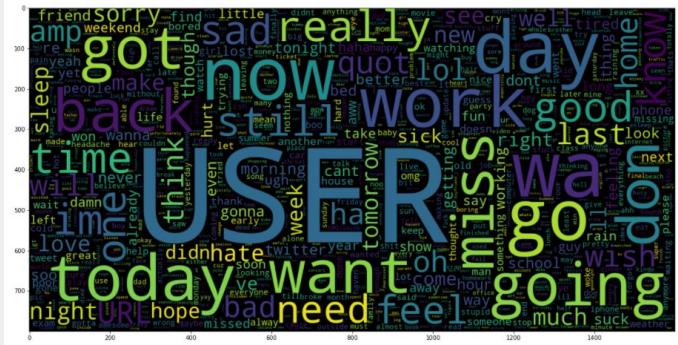


Figure 1: Word Cloud for negative words

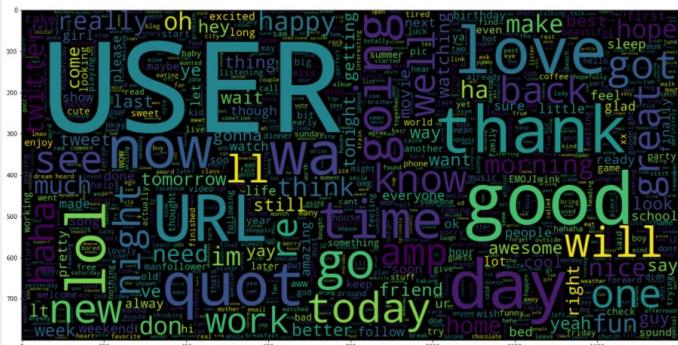


Figure 2: Word cLoud for positive words

# Approach and Methodologies

(Building Sentiment Analysis Models)

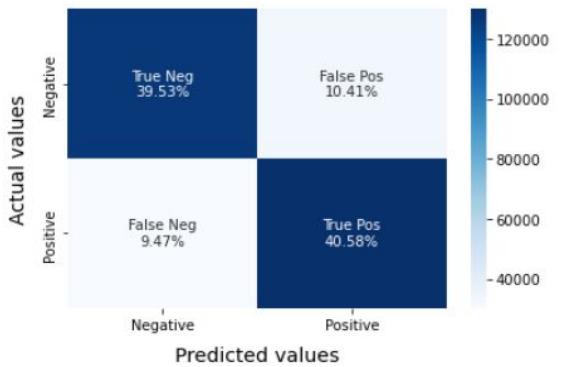
- Extracting Features:
  - TDIF vectorization:
    - Term Frequency: the occurrence of the word in the dataset over the total number of words in the dataset.
    - Word weight: Weight of each word based on its frequency that occurred in the dataset.
- Creating Sentiment Analysis Models:
  - Naive Bayes Algorithm
  - Support Vector Machine Algorithm
  - Logistic Regression Algorithm

# Approach and Methodologies (Models Evaluation)

## Naive Bayes Classification

	precision	recall	f1-score	support
0	0.81	0.79	0.80	159815
1	0.80	0.81	0.80	160185
accuracy			0.80	320000
macro avg	0.80	0.80	0.80	320000
weighted avg	0.80	0.80	0.80	320000

Confusion Matrix



## Support Vector Machine

	precision	recall	f1-score	support
0	0.82	0.81	0.81	159815
1	0.81	0.82	0.82	160185
accuracy			0.82	320000
macro avg	0.82	0.82	0.82	320000
weighted avg	0.82	0.82	0.82	320000

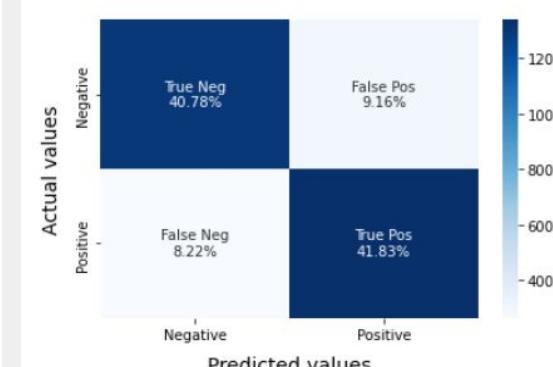
Confusion Matrix



## Logistic Regression

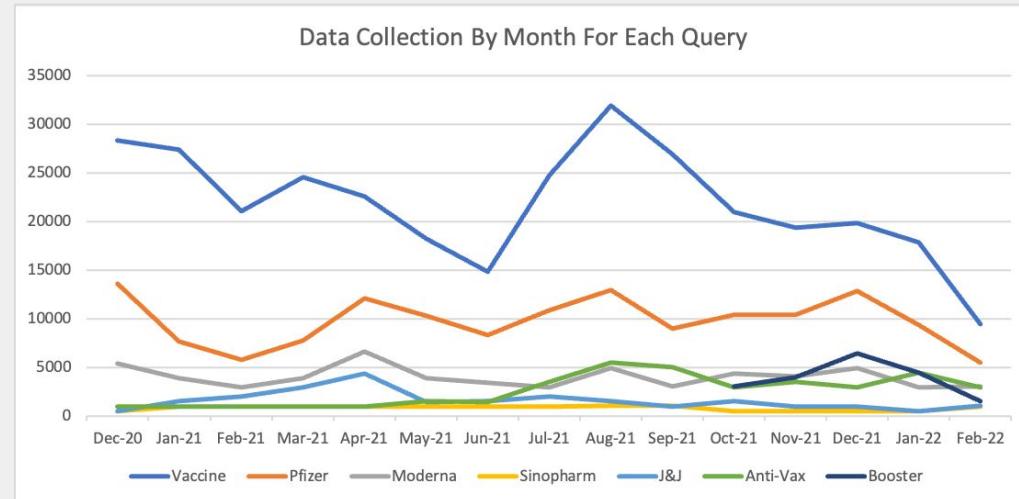
	precision	recall	f1-score	support
0	0.83	0.82	0.82	159815
1	0.82	0.84	0.83	160185
accuracy			0.83	320000
macro avg	0.83	0.83	0.83	320000
weighted avg	0.83	0.83	0.83	320000

Confusion Matrix



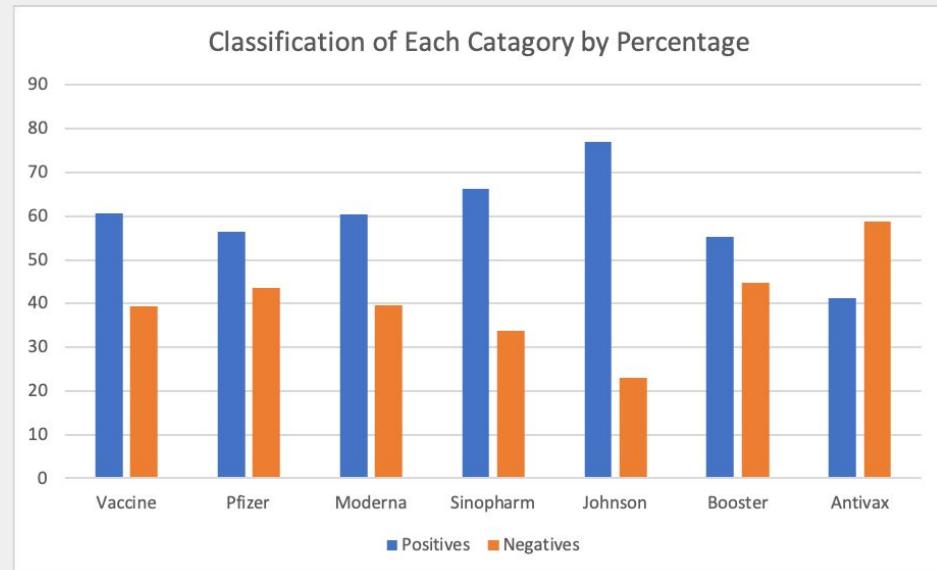
# Approach and Methodologies (Constructing Twitter Datasets)

- Query Parameters:
  - ◆ Timeframe: Dec 2020-Feb 2022
  - ◆ Language: English
  - ◆ Ads: Excluded
  - ◆ Keywords: 7 different categories
- Keywords in the queries passed:
  - ◆ General Covid Vaccine Sentiment
  - ◆ Pfizer Vaccine Sentiment
  - ◆ Moderna Vaccine Sentiment
  - ◆ Sinopharm Vaccine Sentiment
  - ◆ Johnson & Johnson Vaccine Sentiment
  - ◆ Anti Vax Sentiment
  - ◆ Booster Vaccine Sentiment
    - Timeframe: Oct 2021- Feb 2022



# Applying the Model

1. Used Logistic Regression Model
2. Preprocessing the text
3. Extracting Features
4. Classifying using Logistic Regression Model
  - a. Annotating the data to positive and negative based on the sentiment given by the model



# General Vaccine Related Tweets

- 61% positives tweets.
  - 39% negative tweets.
  - Figure 3: Tweets Classification over 15 months
  - Figure 4: Word Cloud for Negative Sentiment.

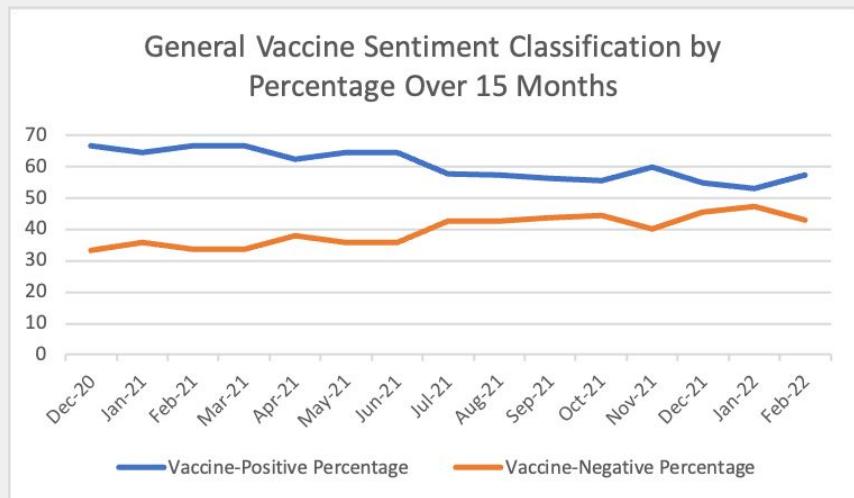


Figure 3: General Vaccine Sentiment Classification by Percentage Over 15 Months.

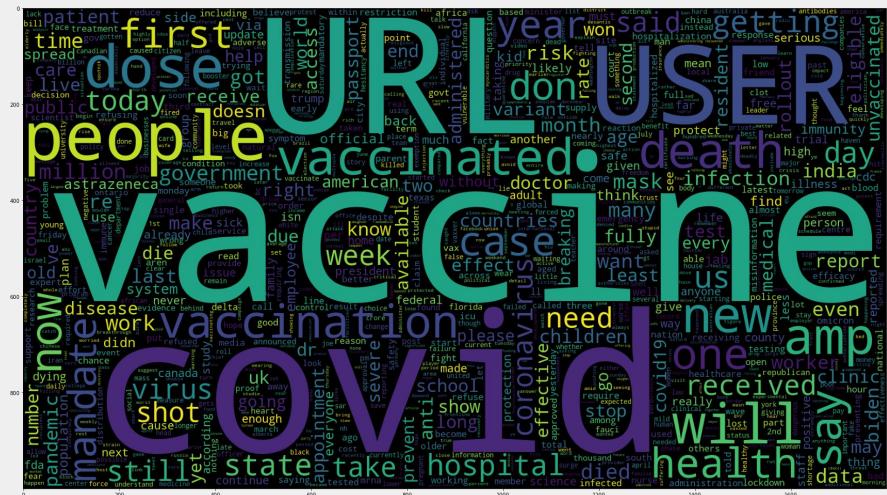


Figure 4: Word Cloud for Negative General Vaccine Sentiment

# Pfizer Vaccine Related Tweets

- 56% positives tweets.
  - 44% negative tweets.
  - Figure 5: Tweets Classification over 15 months
  - Figure 6: Word Cloud for Negative Sentiment.

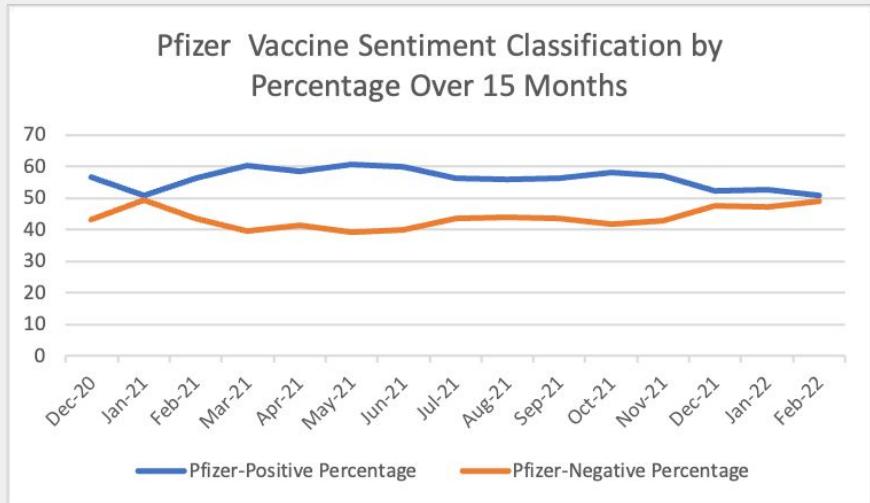


Figure 5: Pfizer Vaccine Sentiment Classification by Percentage Over 15 Months.

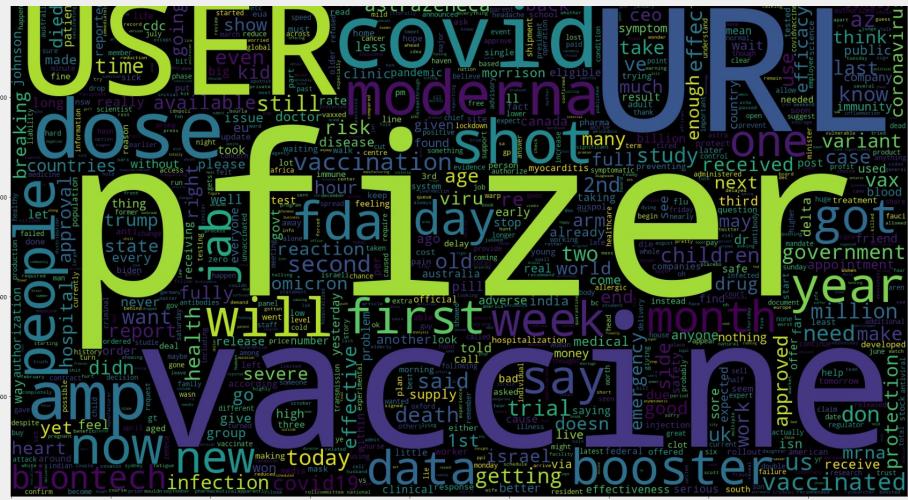


Figure 6: Word Cloud for Negative Pfizer Vaccine Sentiment

# Moderna Vaccine Related Tweets

- 60% positives tweets.
  - 40% negative tweets.
  - Figure 7: Tweets Classification over 15 months
  - Figure 8: Word Cloud for Negative Sentiment.

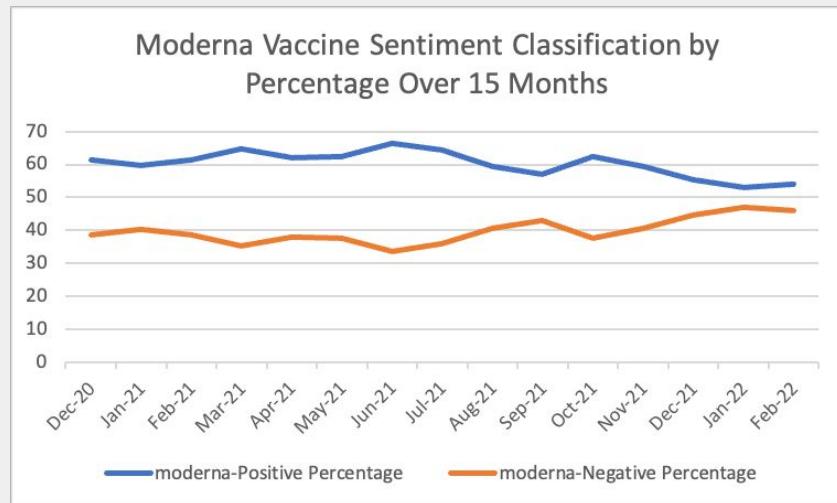


Figure 7: Moderna Vaccine Sentiment Classification by Percentage Over 15 Months.

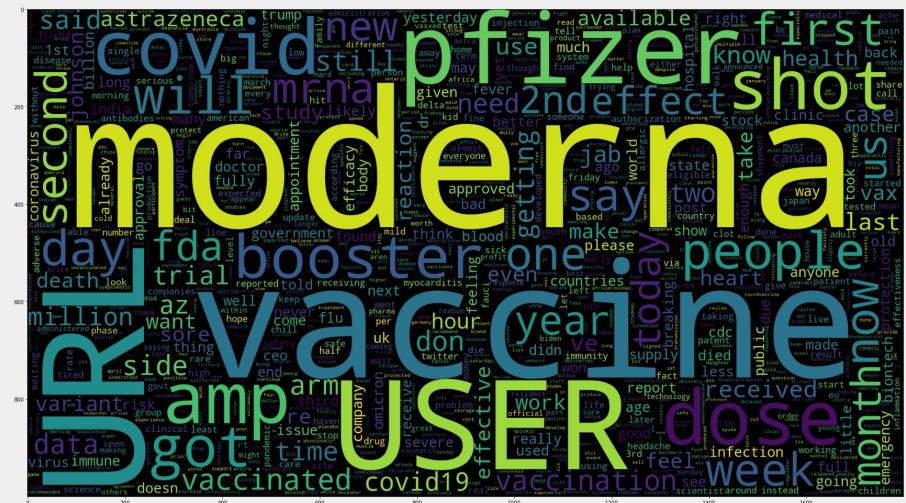


Figure 8: Word Cloud for Negative Moderna Vaccine Sentiment

# Sinopharm Vaccine Related Tweets

- 66% positives tweets.
  - 34% negative tweets.
  - Figure 9: Tweets Classification over 15 months
  - Figure 10: Word Cloud for Negative Sentiment.

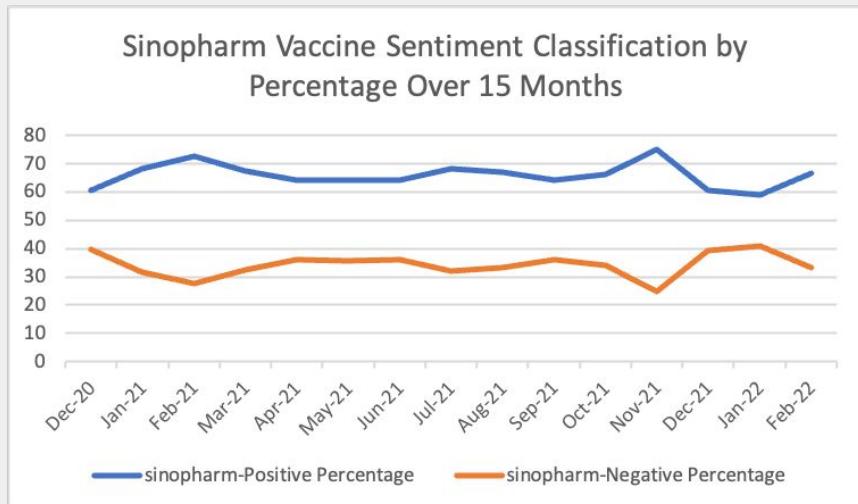


Figure 9: Sinopharm Vaccine Sentiment Classification by Percentage Over 15 Months.

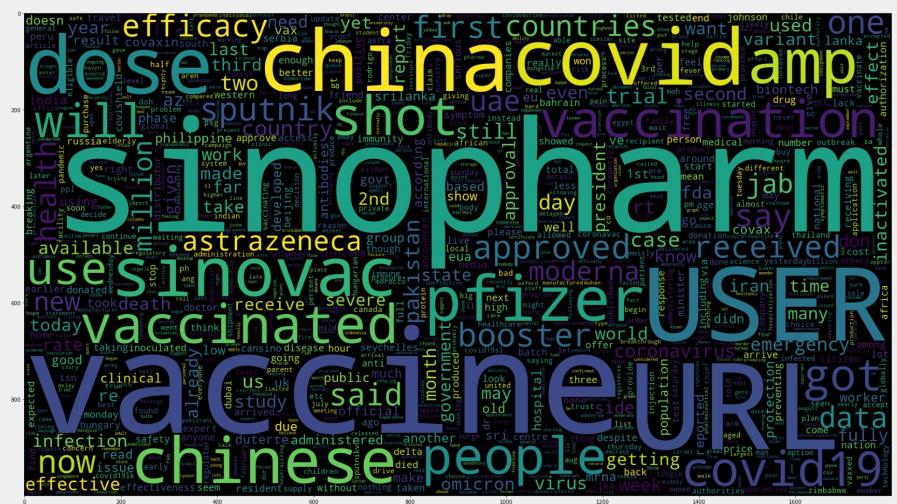


Figure 10: Word Cloud for Negative Sinopharm Vaccine Sentiment

# Johnson & Johnson Vaccine Related Tweets

- 77% positives tweets.
  - 23% negative tweets.
  - Figure 11: Tweets Classification over 15 months
  - Figure 12: Word Cloud for Negative Sentiment.

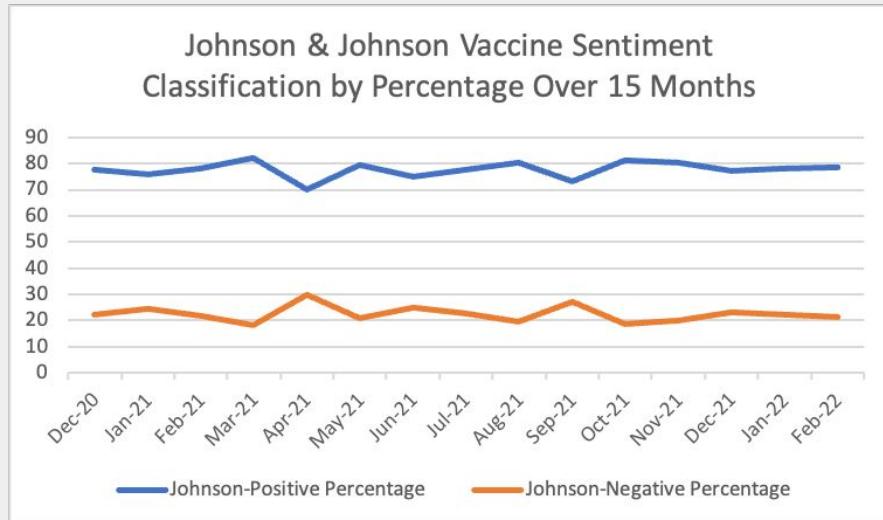


Figure 11: Johnson & Johnson Vaccine Sentiment Classification by Percentage Over 15 Months.

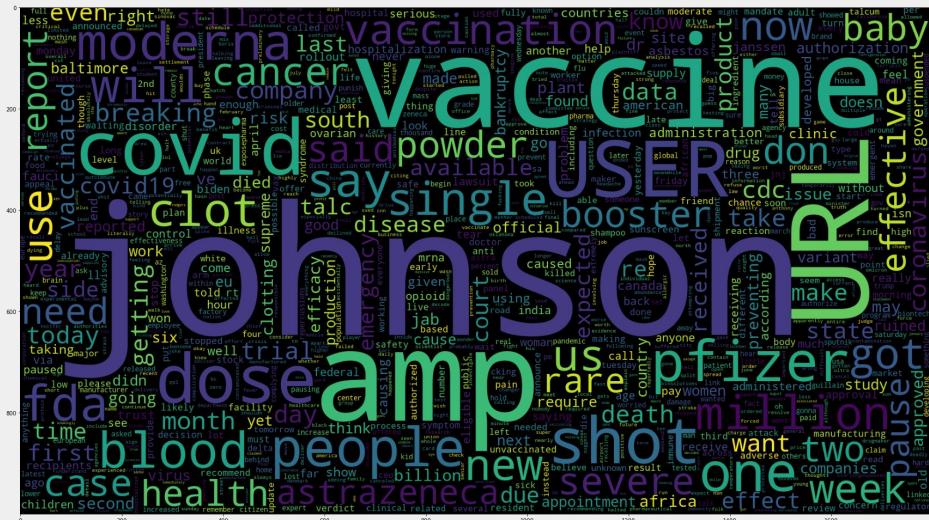


Figure 12: Word Cloud for Negative Johnson & Johnson Vaccine Sentiment

# Anti Vax Vaccine Related Tweets

- 41% positives tweets.
  - 59% negative tweets.
  - Figure 13: Tweets Classification over 15 months
  - Figure 14: Word Cloud for Negative Sentiment.

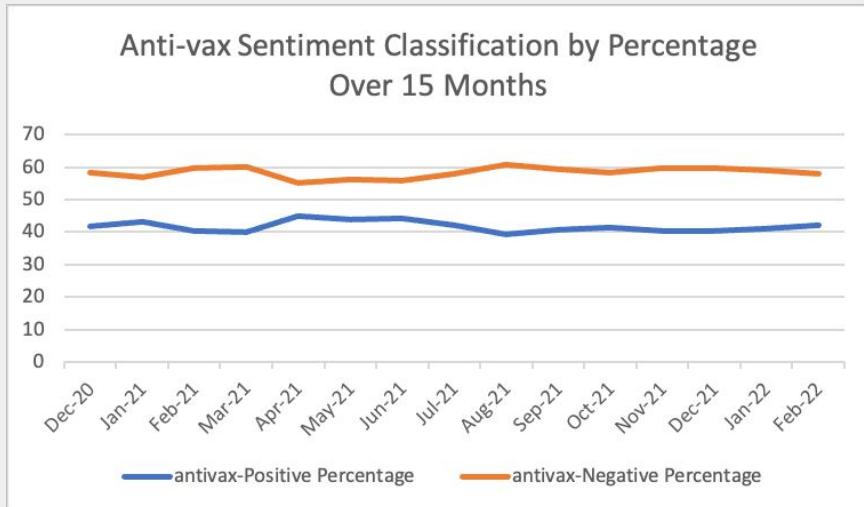


Figure 13: Anti-Vax Vaccine Sentiment Classification by Percentage Over 15 Months.

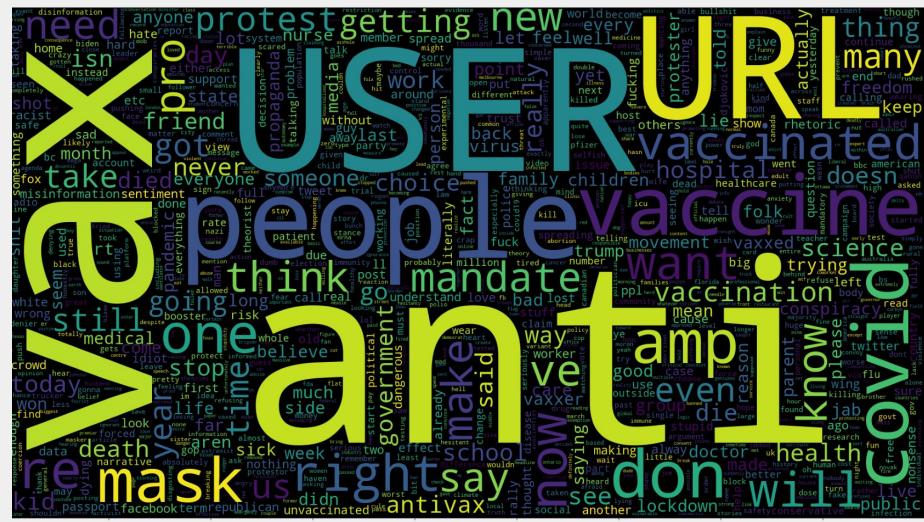


Figure 14: Word Cloud for Negative Anti-Vax Vaccine Sentiment

# Booster Vaccine Related Tweets

- 55% positives tweets.
  - 45% negative tweets.
  - Figure 15: Tweets Classification over 5 months
  - Figure 16: Word Cloud for Negative Sentiment.

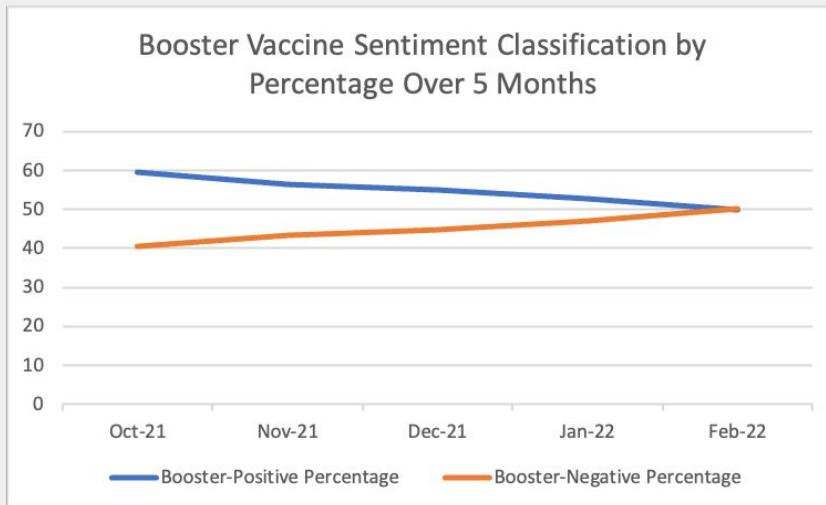


Figure 15: Booster Vaccine Sentiment Classification by Percentage Over 5 Months.

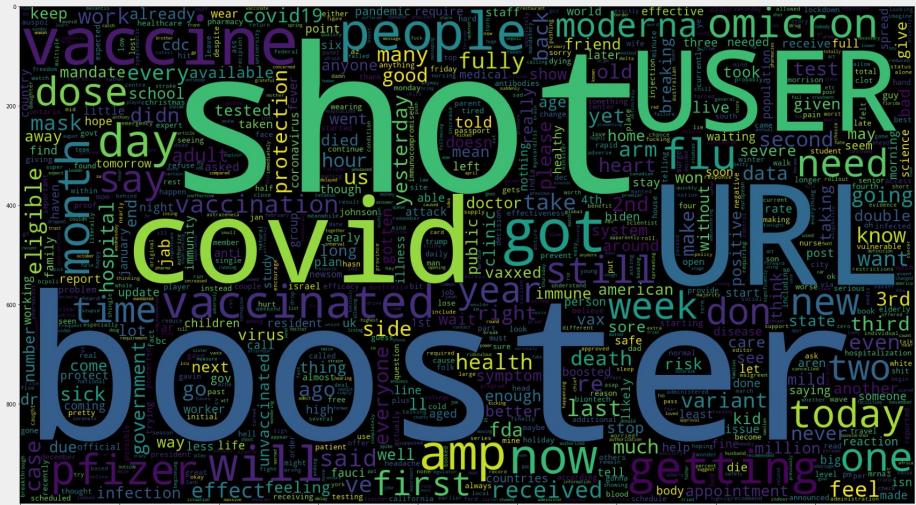


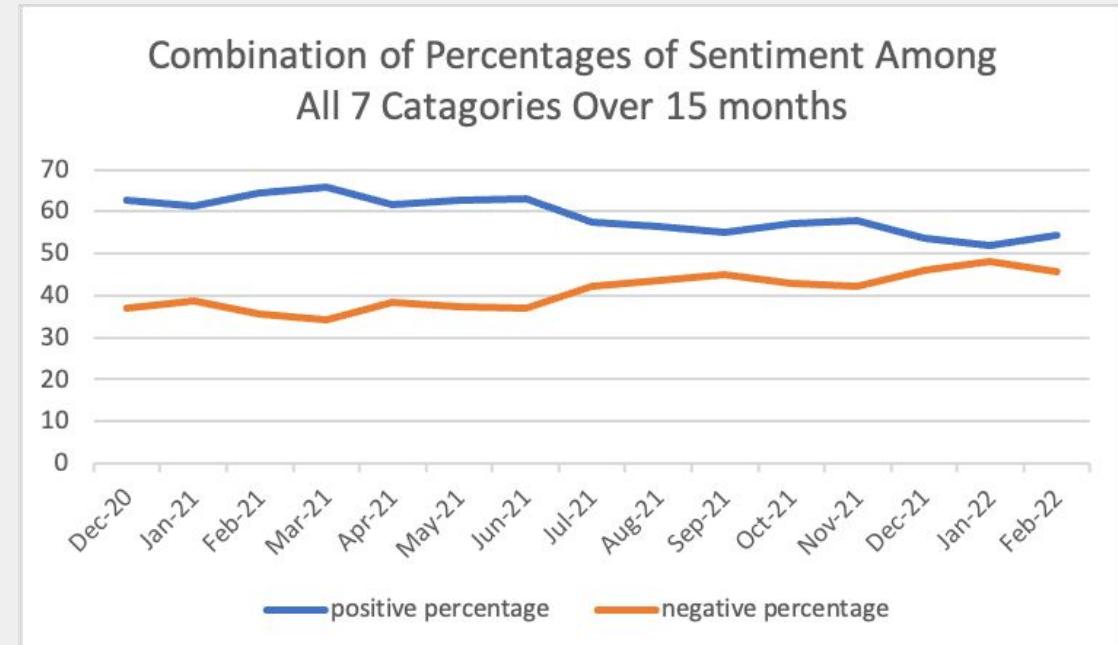
Figure 16: Word Cloud for Negative Booster Vaccine Sentiment

# Tweets Misclassification

- Sarcasm, irony, and humor
  - "USER awesome so guys who pretend to be physicians while they spend their days doing donald bidding get vaccines while front line healthcare workers who battle covid 19 everyday are still waiting not good look greg" Classified as positive
- Negations
  - "This week was not bad for us! Our covid-19 cases are going down. Many ugandans have still not accessed the vaccine. We are also still under lockdown In Uganda. Most of the work we are doing is virtual #SheLeads #Nyayadzedu." Classified as Negative
- Multi-polarity
  - "will say the one good thing about me and nick getting covid and him getting really sick is that it convinced my granny to get the vaccine and think if he didn get sick she still would be unvaccinated " classified as negative
  - "just recovered from pretty bad case of covid and if given the choice now between covid and the vaccine would choose covid again " classified as positive

# Analysis and Discussion (Monthly Trends)

- Method:
  - Finding percentage for each month for each category.
  - Summed up the percentages and divided by 7.
  - The following figure demonstrates the trends.



## Analysis and Discussion (Reasons Extraction)

- Methods:
  - Extracted the most frequent, common and relevant keywords used in the negatively classified tweets across the 7 categories.
  - Looking at different tweets that contain these keywords and extract reasons for the negative sentiments.
  - Figure 17 shows list of most frequent, common and relevant keywords.



Figure 17: Word Cloud for Most Frequent, Common and Relevant Words Found in Negatively Classified Tweets Across All Categories.

# Analysis (Reasons Extraction)

## 1. Government

- “turned it off heard absolutely nothing that makes me worried or concerned except possibly the government response the vaccine data shows it works against variants the amount of past covid infections amp vaccine data means outbreak would be near impossible enjoy weekend”

## 2. Effectiveness, Strains, and Variants

- “South Africa has just rejected the AstraZeneca vaccine against COVID-19. It is not effective for the South African strain of the virus. Why should the rest of Africa receive millions of doses when virus strains are undetermined?”
- “New strain of Covid-19 virus found in England. Vaccine seems to be not effective”

## 3. Safety

- “Why is our government pushing a partially tested vaccine which may have dangerous side effects versus using a safe, cheap, and extremely effective prophylaxis and treatment for covid-19 that is ivermectin? ”

## 4. Priority and Eligibility

- “Updated statement from the JCVI: unvaccinated adults aged 30 to 39 years who are not in a clinical priority group at higher risk of severe COVID-19 disease, should be preferentially offered an alternative to the AstraZeneca COVID-19 (AZD1222) vaccine”

# Analysis (Reasons Extraction)

## 5. Infections

- "the covid cultists wants it both ways with the vaccines they say they are effective amp everyone has to get one amp they want vaccine passports but we can never ever go back to normal because the vaccines don prevent infections or transmission they only lessen symptoms"

## 6. Healthcare

- "The speed at which we got a vaccine for COVID-19 is what a quality market-driven health care system could do for America."

## 7. Misinformation

- "One of the biggest barriers standing in the way of ending the pandemic is not medical or logistical. It's the misinformation about the COVID-19 vaccines."

## 8. Republicans

- "rachel maddow account of the dire need many countries have to get covid vaccines and how the Biden administration wants to break patents so these countries can get produce them who is opposed to that republican they are really sociopaths it global health crisis"

## Analysis (Reasons Extraction)

### 9. Mandatory, freedom and Covid Passport

- "We must take note of every politician of any party who expresses support for mandatory covid vaccination or vaccine passports or digital credentials. They must never again be allowed to hold public office."

### 10. Pregnant

- "not sure how to say this forcefully enough without being uncharacteristically profane but to be very clear the covid 19 vaccines do not shed it utter bollocks this idea that vaccinated person can somehow harm pregnant woman this isn rocket science"

### 11. Election

- "Domestic vaccine passports won't be temporary and they won't stop at COVID. They are the thin end of the wedge. No Pass? No gym, no beer, no theatre... The elections tomorrow may be your last chance to be heard for a long time. Please use your vote wisely."

# Conclusion and Future Work

- We develop a sentiment analysis model that would successfully allow us to capture trends on the hesitancy of the people over taking the coronavirus vaccine.
- We found out which ML algorithm would work best for this problem based on evaluation of three different algorithms.
- We analyzed the hesitancy over different types of vaccines and extracted reasons for vaccine hesitancy over Covid-19 vaccination.
- We visualized trends for hesitancy over month and frequent reasons on the hesitancy.



## Future Work

- Enhancing the performance for the model.
- And adding another category of classification that is Neutral to address this category of sentiment.
- Narrowing the search further and making sure to eliminate the propaganda and tweets produced by bots.



**Thank you!  
Questions?**