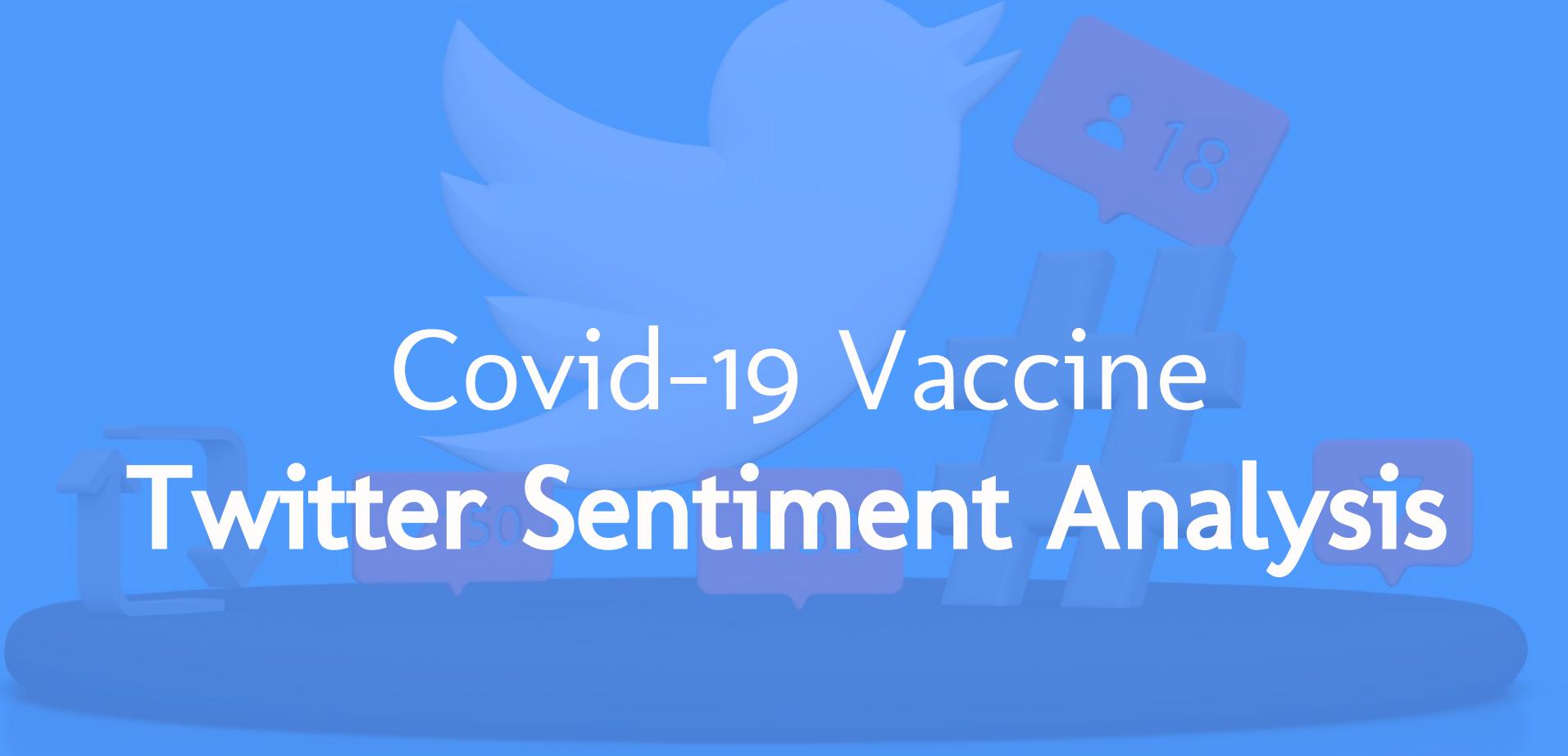


---

A large, semi-transparent Twitter logo is positioned in the center of the slide. It features a blue silhouette of the Twitter bird facing right, with several light blue speech bubbles containing text floating around it. One bubble contains the word 'so', another contains '18', and others contain smaller, less legible text.

# Covid-19 Vaccine Twitter Sentiment Analysis

Bekah McLaughlin

# Overview



# Public Health Problem - Covid-19 Vaccination for 2024



- Covid-19 is no longer an acute crisis
- It is still prevalent and surges can put vulnerable people at risk
- Continued Barriers to Vaccination

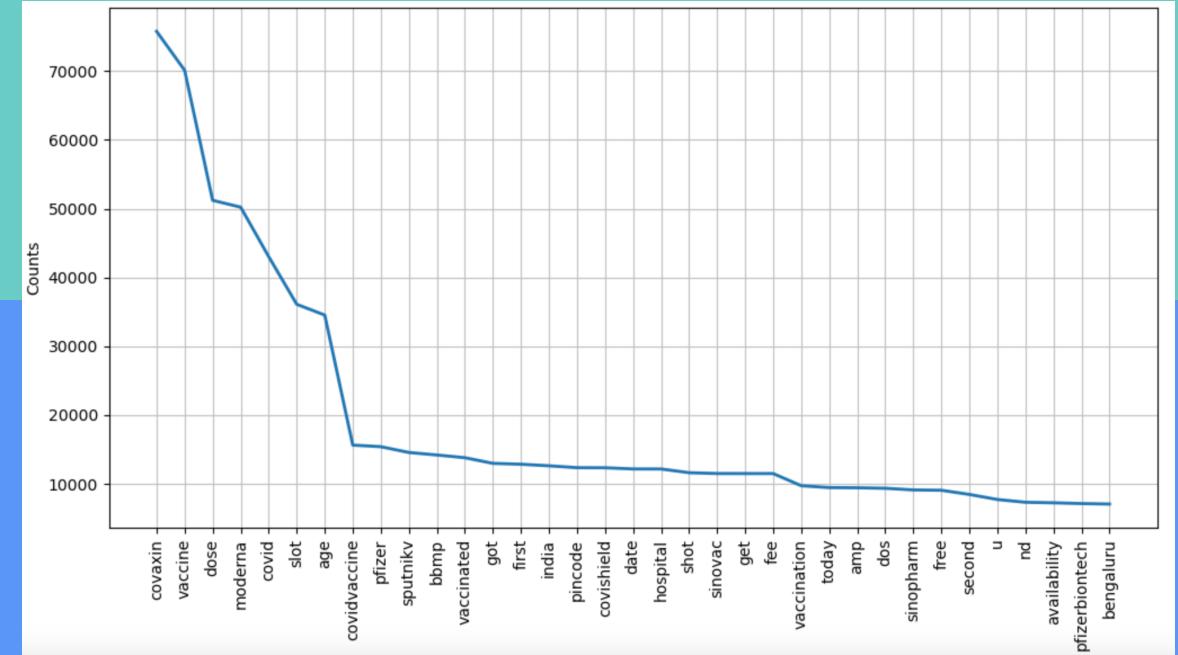
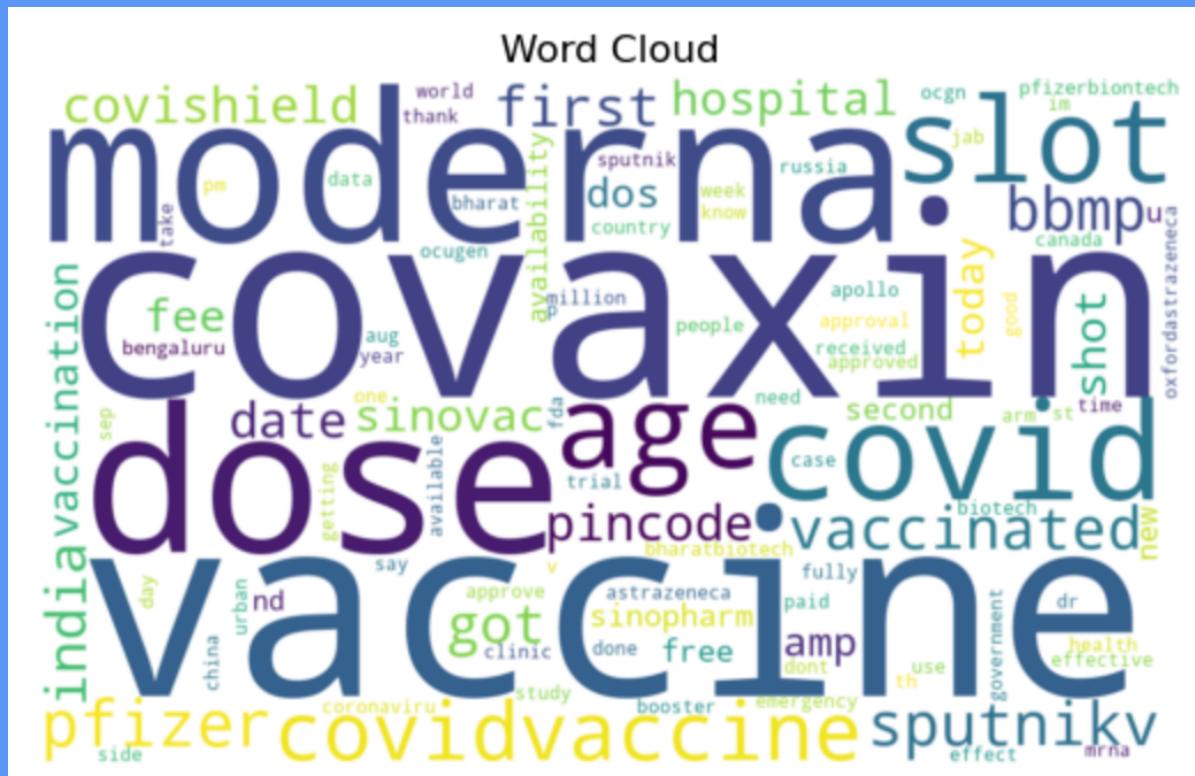
## THE GOAL:

Create a machine learning model that can quickly assess tweets about the Covid-19 vaccines for their sentiment



# Data

- Large dataset from Twitter
  - 228,207 rows each containing a raw tweet's text
  - Every tweet in this dataset pertains to the COVID-19 vaccines.



- Normal text cleaning + removing emojis, usernames, hyperlinks
  - Used large database of positive and negative words to score each tweet and label with sentiment
  - Preprocessing: tokenization, stopword removal, lemmatization, TF-IDF Vectorization

# Models

Logistic  
Regression

Naive  
Bayes

Random  
Forest

Support Vector  
Machine

	precision	recall	f1-score	support
Negative	0.94	0.92	0.93	5359
Neutral	0.97	0.96	0.97	19727
Positive	0.95	0.97	0.96	9146
accuracy			0.96	34232

	precision	recall	f1-score	support
Negative	0.70	0.65	0.67	5359
Neutral	0.81	0.86	0.83	19727
Positive	0.79	0.72	0.75	9146
accuracy			0.79	34232

	precision	recall	f1-score	support
Negative	0.91	0.68	0.77	5359
Neutral	0.87	0.95	0.91	19727
Positive	0.90	0.87	0.88	9146
accuracy			0.88	34232

	precision	recall	f1-score	support
Negative	0.97	0.97	0.97	5359
Neutral	0.99	0.98	0.99	19727
Positive	0.98	0.99	0.98	9146
accuracy			0.98	34232

Second best  
performing model

Lowest performing  
model

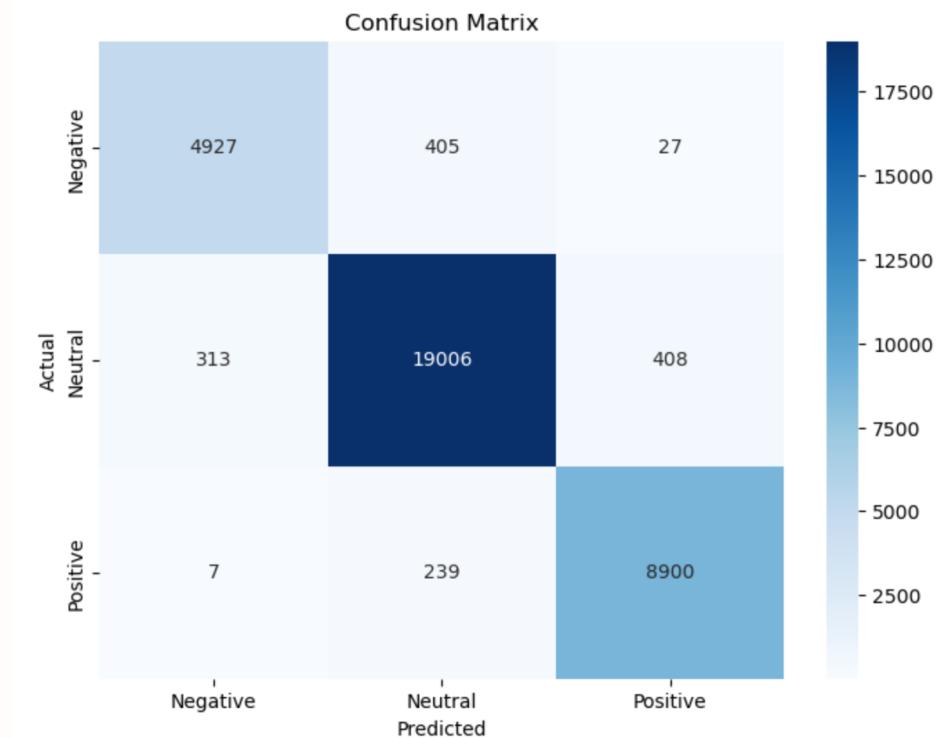
This model did well,  
but not the best

Model chose for  
deployment

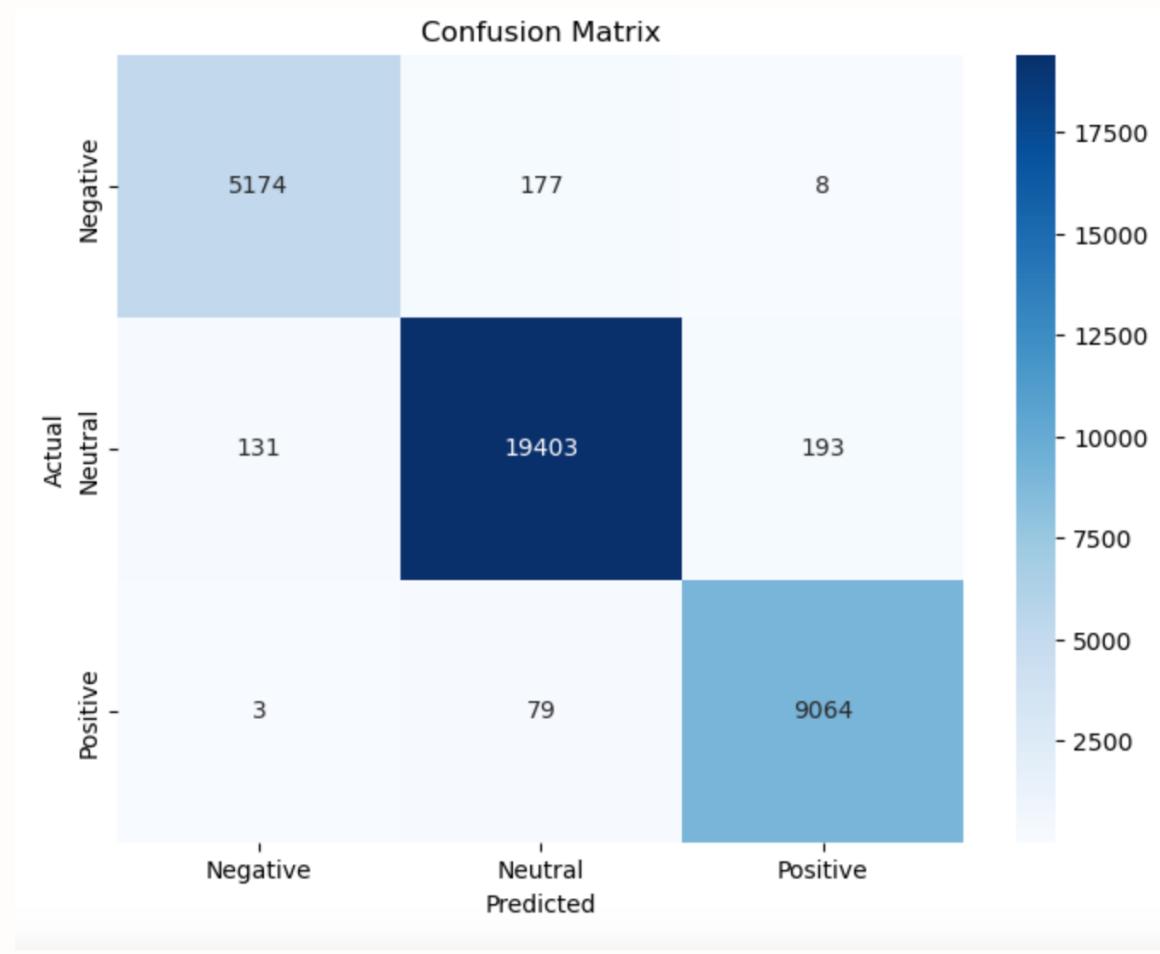
# Logistic Regression

Test Set:

	precision	recall	f1-score	support
Negative	0.94	0.92	0.93	5359
Neutral	0.97	0.96	0.97	19727
Positive	0.95	0.97	0.96	9146
accuracy			0.96	34232
macro avg	0.95	0.95	0.95	34232
weighted avg	0.96	0.96	0.96	34232



# Support Vector Machine



Test Set:	precision	recall	f1-score	support
Negative	0.97	0.97	0.97	5359
Neutral	0.99	0.98	0.99	19727
Positive	0.98	0.99	0.98	9146
accuracy			0.98	34232
macro avg	0.98	0.98	0.98	34232
weighted avg	0.98	0.98	0.98	34232

# Results

## APP DEPLOYMENT

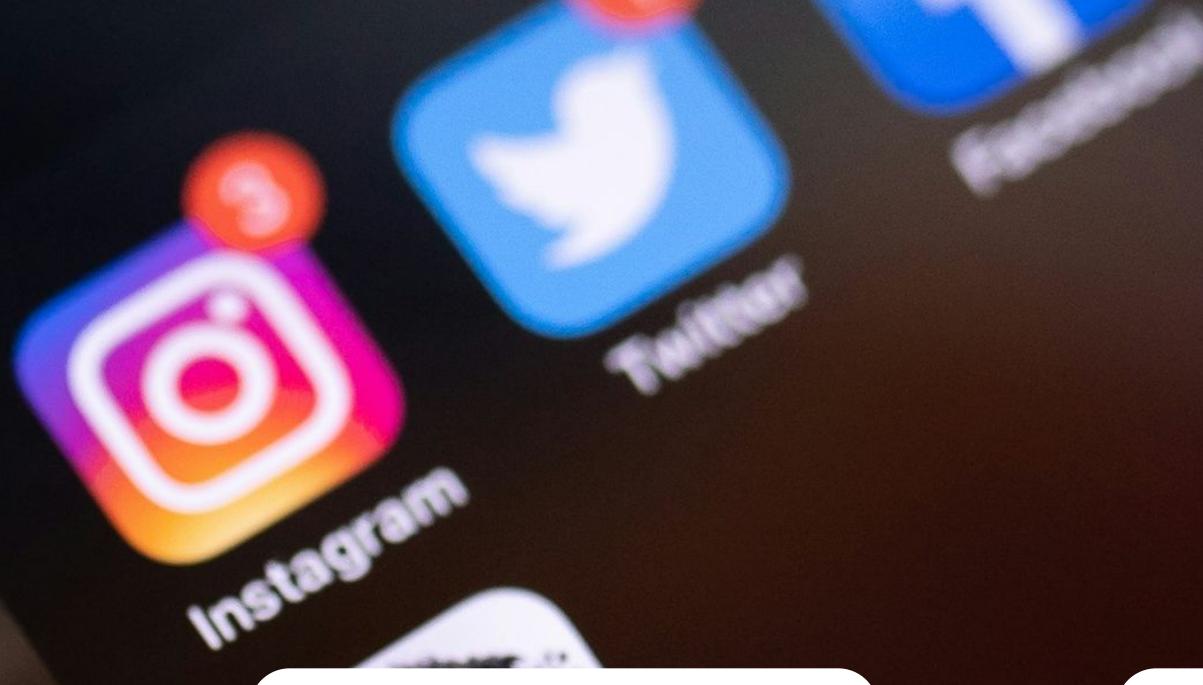
[www.streamlit.deploy](http://www.streamlit.deploy)

## LIMITATIONS / CONSIDERATIONS

- Time
- Different sentiment analysis approaches
- Different classification approaches

## FUTURE GOALS

- Topic Modeling
- Different Sentiment Labeling methods
- Different Models





# Questions?



<https://www.linkedin.com/in/bekah-mclaughlin-8a86b02b4/>



rebekah.l.mclaughlin@gmail.com