Sentiment Analysis on Covid-19 Vaccines

Yash Brid – 2019130008

Abhishek Chopra – 2019130009

Sumeet Haldipur – 2019130018

Problem Definition:

Problem:

- 2021 marked the commencement of the COVID-19 vaccination drive.
- A plethora of vaccinations are now available in the market that'll help in the fight against COVID-19.
- There is a lot of stigma regarding the vaccines amongst the people.

Our Solution:

- We want to create a platform where the people can search up the name of a vaccine and get the general sentiment associated with it.
- We plan on doing to break the stigma associated with these vaccinations so that the people get to know the general opinion surrounding a vaccine.

Literature Survey:

We went through 6 papers for the purpose of implementing these projects. A brief description of what we found is mentioned below:

Paper Name	Major Work	Drawbacks
Sentiment Analysis on Twitter Data for product evaluation	To analyze the sentiments and opinions of people, on various products and services, posted on various microblogging websites using Machine Learning approach. This paper used the Naïve Bayes Algorithm to train a Movie Review Dataset, it also uses the TextBlob package in python to calculate the sentiments of the tweets	The accuracy of the algorithm is satisfactory but can be improved by training it on a larger dataset and using better approaches and algorithm.
Sentiment Analysis Of Product Reviews – A Survey	This survey paper gives an impact on the ongoing educates in sentiment analysis calculations and applications. This article also offers commitments to numerous feeling examination related fields that utilize	The approach isn't very accurate due to the inherent complexity of the natural language constructs as there are different ways of representing the same meaning.

	sentiment analysis systems for genuine application.	
Sentiment Analysis for Social Media	This paper reveals an approach which is implemented as a tool that can analyze sentiments on twitter social media addressing above issues and then develop an application to generate knowledge that can be useful for business environments using people's attitudes about their products and services.	Some people use different jargon, slang communications and short forms of the words for their ease. Therefore, it is difficult to gauge and measure the sentiments accurately in terms of their polarity such as positive, negative or neutral and the subjectivity of sentiments
Sentiment analysis of twitter data	This paper discusses social network analysis and the importance of it, implements a python program to conduct sentimental analysis and show results based on different queries including movie, politics, fashion, and fake news.	For some queries, the neutral tweets are more than 60% which clearly shows the limitation of the current works and highlights a need to improve twitter sentiment analysis.
A Study on Sentiment Analysis Techniques of Twitter Data	This paper explores the various sentiment analysis applied to Twitter data and their outcomes. It gives an idea about the Naive Bayes, Maximum Entropy and Support Vector Machine Algorithms. It also discusses the Supervised Machine Learning Approaches, Ensemble Approaches, Lexicon Based Approaches and Hybrid methods for sentiment analysis.	Combining various features was found to lead to improve the performance in most cases, but substandard performance in others. Active learning methods weren't utilised to detect Twitter sentiments.
Sentiment Analysis: It's Complicated!	This paper highlights the need to better engage with how humans actually annotate data in short-text sentiment analysis dataset construction by constructing the new McGill Twitter Sentiment Analysis (MTSA) dataset.	Raw human annotations weren't included which limits the extent to which the MTSA dataset can operate.

Scope of Work:

Assumptions:

Considering all the tweets used for the analysis are honest opinions and don't fall under the umbrella of sarcasm.

Constraints:

We won't be able to filter out bots/fake accounts created for defamation/glorification purposes.

Focus:

To develop a website to generate knowledge that can be useful for business environments using people's attitudes about their products and services.

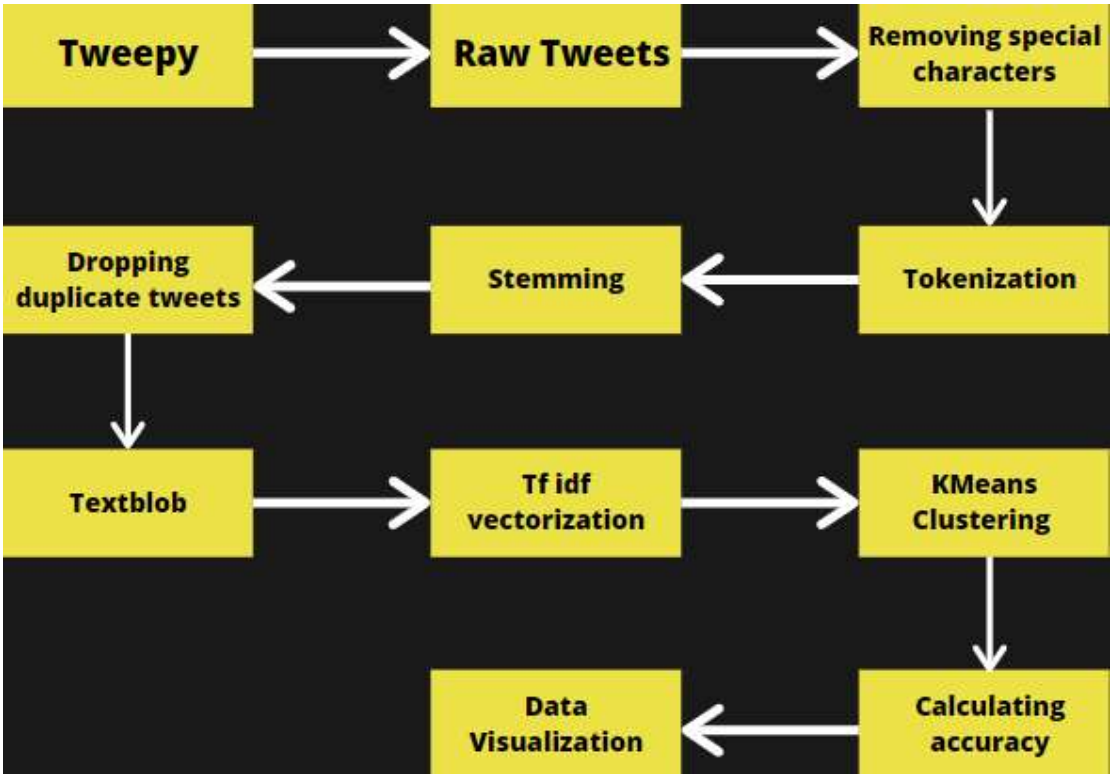
Objectives:

a) Creating an easily accessible interface for the population to get an idea about the sentiments surrounding vaccines.

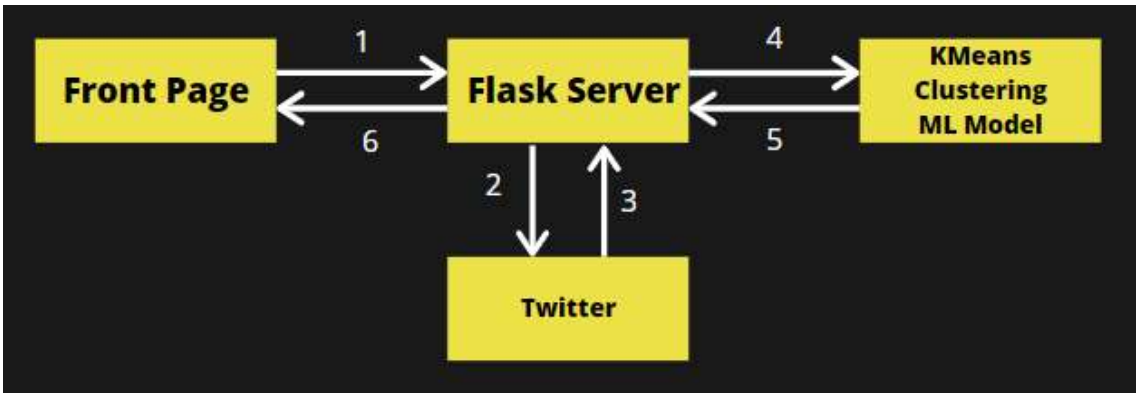
b) To give people a broader perspective about the vaccines.

Block Diagram:

Model:



Application:



1) The user clicks “Analyze”.

2) A “POST” request is made to the flask server.

3) The flask server makes a request to tweepy (Twitter API) using key.

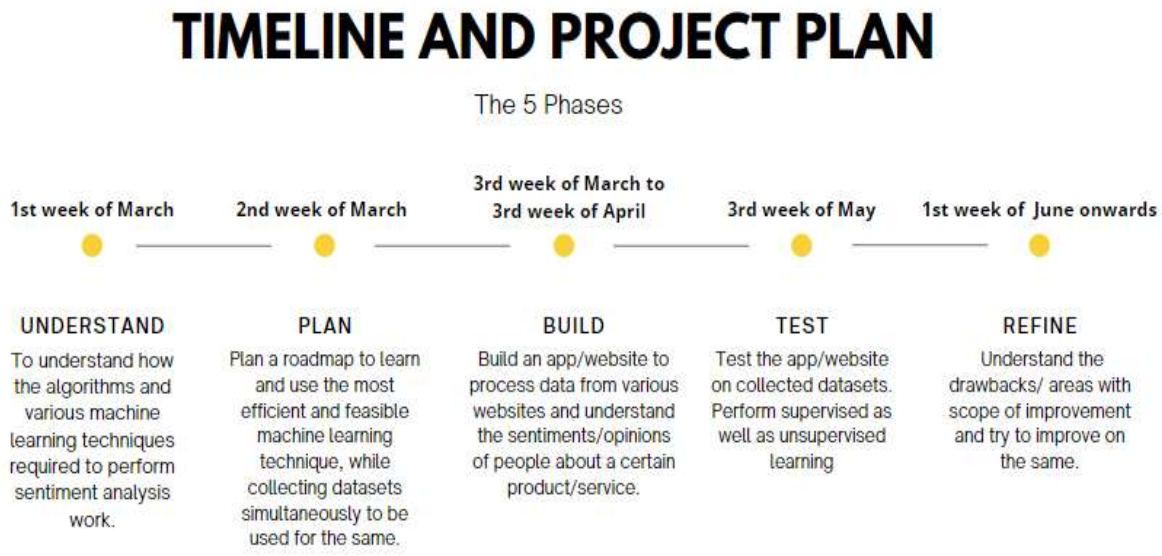
4) After authentication the twitter API returns the dataset (tweets and valuable attributes) on the key words (i.e. vaccine names).

5) Now, the data is passed to the ML model for computation.

6) K Means Clustering algorithm clusters the tweets in 3 categories (Positive, Negative, Neutral).

7) Finally Flask displays the results on the web page and provides an analysis report using google charts.

Project Plan And Timeline:



Implementation Details:

Implementation Tools:

Tech Stack Used:

- Python
- Jupyter Notebook
- HTML, CSS, JavaScript
- Flask

Packages:

- Pandas, Numpy, nlkt (NLP Toolkit) , re (Regex Expression), matplotlib, vaderSentiment, sklearn.

APIs:

- Tweepy
  https://www.tweepy.org/

Implementation:

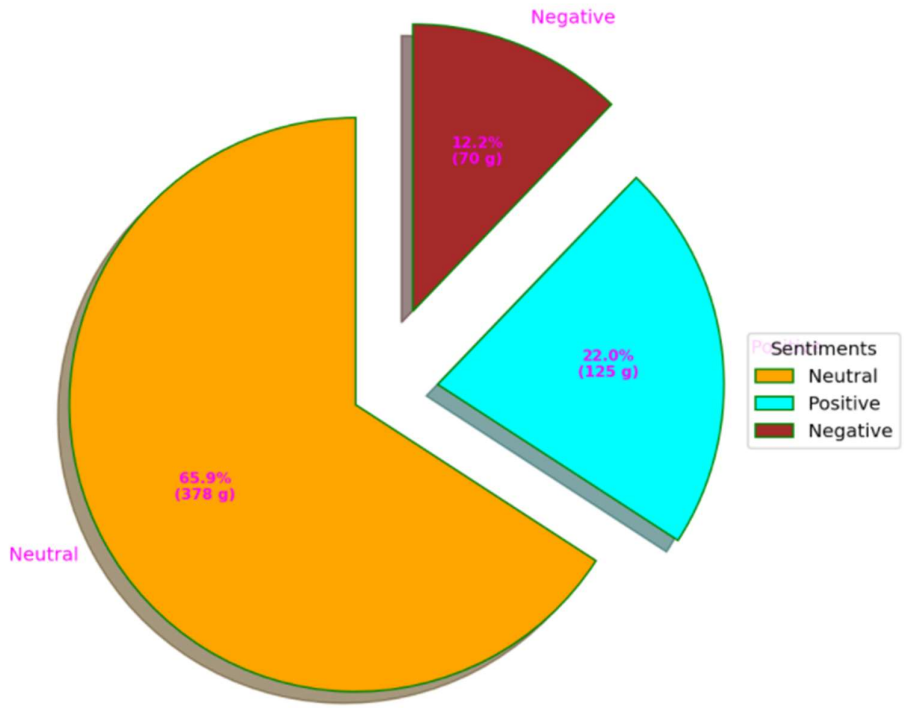
- We have successfully implemented the model using K-Mean's clustering algorithm to classify the Tweets into the 3 clusters.
- These 3 clusters can be further broken down into Pie Charts to give a better idea to the users regarding the sentiment surrounding a particular vaccine.
- We have deployed our website by integrating it with our model using Flask.

Results:

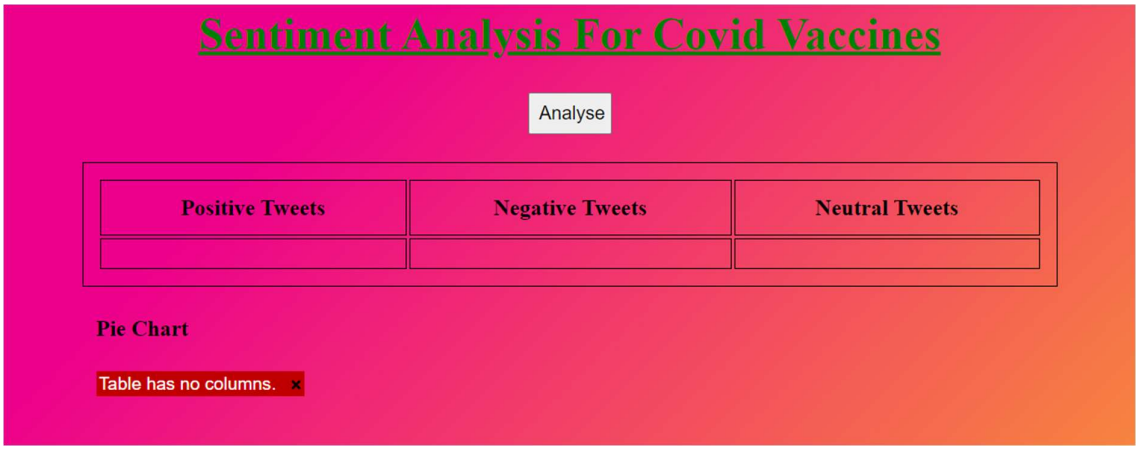
We deployed our website that was successfully able to analyze the sentiments regarding vaccines using live data that was fetched from twitter with an accuracy of 81%.

Screenshots:

Phase 2:

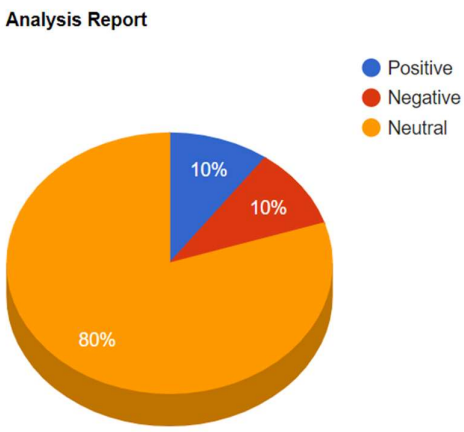


Phase 3:



Positive Tweets	Negative Tweets	Neutral Tweets
3	3	24

Pie Chart



Conclusion:

- We successfully implemented a model on Sentiment Analysis for Covid-19 Vaccines using unsupervised learning.
- We learnt about the K-Means clustering algorithm to implement our model.
- We integrated our model with a web app using Flask and deployed the same.
- We got familiar with various Python libraries and utilized the same for our project.

Further Work:

- Fetching more tweets to analyse on a broader level.
- Improving the UI of our website.
- Letting the users enter the vaccine of their choice and analyse the sentiments regarding it.
- Sensing sarcastic tweets.

References:

K. P. Sinaga and M. Yang, "Unsupervised K-Means Clustering Algorithm," in IEEE Access, vol. 8, pp. 80716-80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

M. Capo, A. Perez and J. A. A. Lozano, "An efficient Split-Merge re-start for the K-means algorithm," in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2020.3002926.

T. Handhayani and I. Wasito, "Fully unsupervised clustering in nonlinearly separable data using intelligent Kernel K-Means," 2014 International Conference on Advanced Computer Science and Information System, Jakarta, Indonesia, 2014, pp. 450-453, doi: 10.1109/ICACSIS.2014.7065891.