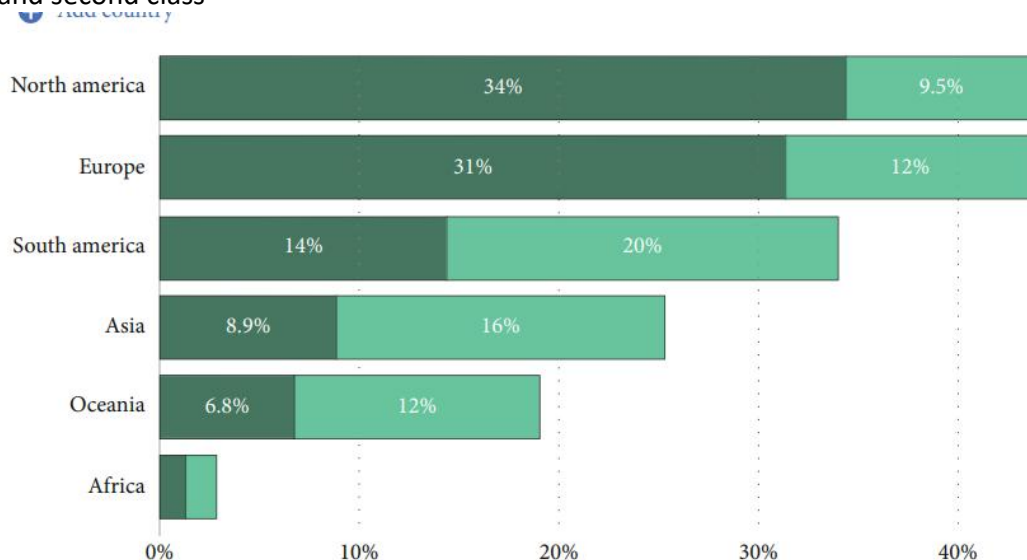


PROBLEM DEFINITION:



The COVID-19 outbreak has brought significant attention to the healthcare sector in recent times, and it has changed the concept of safety in every aspect of our lives. Social distancing is an effective method for reducing the spread of coronavirus. Safety measures such as wearing masks, washing hands regularly, and staying careful regarding intimacy are currently very important. However, these can only reduce the spread of coronavirus, not eradicate it completely. Here, vaccination came into light as the only solution that could fight most effectively against coronavirus and probably eradicate it. Rigorous tests were conducted with the first mRNA vaccines to be introduced; more than 40,000 people participated in a Pfizer vaccine trial and 30,000 in a Moderna vaccine trial. The average efficacy rate of the vaccines in both trials was approximately 94%, and there were no deaths in either of them. Early findings about another viral vector vaccine named Johnson & Johnson that proved to be able to fight against coronavirus and stimulate the recipient's immune response showed a rate of effective action of >85% without serious adverse effects [1]. Vaccination procedures are in full swing worldwide, as shown in Figure 1. There might be some conflicts among regions owing to differences in urgency and economic barriers (which will be explained later in our paper), but mostly, we attempted to present the actual data about people's vaccination status without bias.

DATA COLLECTION:

Share of people vaccinated against COVID-19, Jul 9, 2021
Add country
This data is only available for countries which report the breakdown of doses administered by first and second class



Source: Official data collated by our world in data
Dec27, 2020 · ————— · Jul 9, 2021

 Share of people fully vaccinated against COVID-19 (dark green)
 Share of people only partly vaccinated against COVID-19 (light green).

DATA PREPROCESSING:

First, after dropping some unnecessary columns, all the URLs and emails from the tweets were removed. Then, all the new line characters, double and single quotes, and punctuation signs were deleted. For this type of processing, all the tweets were tokenized before applying all the methods to remove those texts. These were then detokenized and converted into NumPy arrays. After data extraction, the next step is data preprocessing to remove noise and irrelevant information so that the training process of the selected models can be enhanced.

The cleaning process involves the removal of the data elements in the tweets which are not useful for the sentiment analysis process. Such data elements include

@username,

symbols, hyperlinks, punctuation and stop words, and so forth. The preprocessing of tweets was performed using the natural language toolkit (NLTK) library. NLTK incorporates more than 50 corpora, lexical analysis resources, and a collection of libraries for

text processing. These text processing libraries contain the important and fundamental

NLP functions for tagging, parsing, tokenization, as well as semantic reasoning [28].

The

preprocessing steps are explained in subsequent subsections, and some tweet samples

before and after preprocessing are shown to show the output of these steps.

3.2.1. Removal of Username, Hashtags, and Hyperlinks

People mostly tag their friends and related persons in their tweets using '@username' on Twitter to refer to or tag them, and also use hashtags and hyperlinks in their tweets.

These elements in tweets are not useful for the sentiment analysis, so 'username' , 'hashtags' ,

and 'hyperlinks' were removed from the tweets. Table 2 shows the sample text of tweets

before and after preprocessing.

Data before and after preprocessing.

Before Preprocessing	After Preprocessing
Many thyroid and autoimmune patients are wondering whether they should get the COVID-19 vaccine. Thyroid Expert Mary Sâ€¦ https://t.co/8OHcyR5kQ7 (accessed on: 20 May 2021)	many thyroid autoimmune patient wonder whether get covid vaccine thyroid expert mary
As expected, @WHO celebrated the return of #USA to the organization during the surge of the covid #pandemic #COVID19â€¦ https://t.co/TbVcBF3Nxr (accessed on: 20 May 2021)	expect celebrate return organization surge covid

EXPLORATORY DATA ANALYSIS:

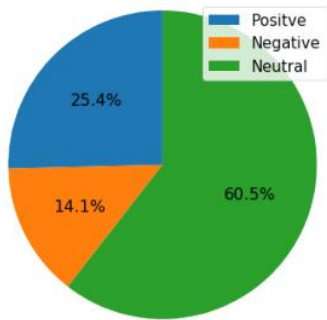
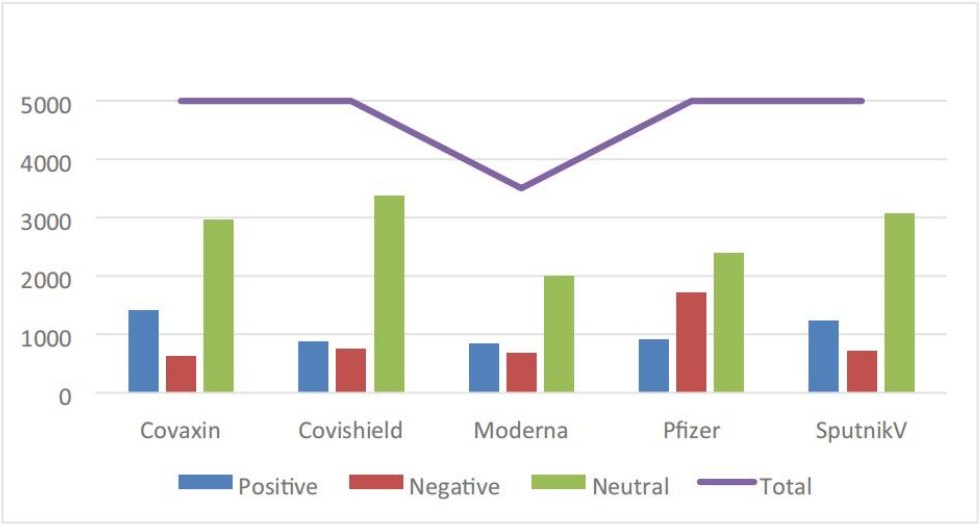
Sentiment Analyzer Tool (VADER). VADER, proposed by C.J. Hutto [18] in 2014, is an NLP-based sentiment analyzer and a pretrained model that uses rule-based values tailored to the perceptions of social media expressions and works well on texts from other fields. It has impeccable performance in the area of social media text. Based on its comprehensive rules, VADER can perform a sentiment analysis of assorted lexical characteristics, as shown in Figure 3. Looking at the valence values for each word in the lexicon, VADER provides a percentage of text ratios that crumble into a positive, negative, or neutral category and sums up a probability value of 1. The compound score for sentiment analysis is the most frequently used measure; a float value in the interval $[-1, +1]$ is a compound score, whose index is determined by adding the values of each word in the lexicon, adapted according to rules, and then standardized to its range.

STATISTICAL ANALYSIS:

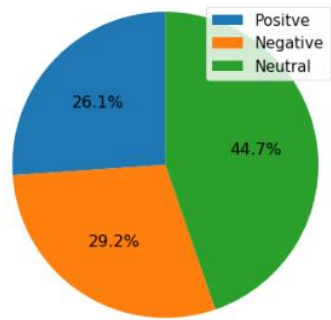
The sentiment analysis is techniques it is used to identifying the users expressions and for that Sentiment Analysis, Machine learning, Natural language processing are popular techniques are effectively used. In this research article here perform the twitter application management and collect real time hashtags discussion on covid-19 vaccination. The twitter general public data collection and preprocessing techniques are applied. Our investigation detected that unigram Sentiment Analysis for all five datasets. Lexicons are used Bing Liu and Sentiment140 are used for interpreting the data. The study is completed on the tweets which are related to the COVID-19 vaccination. Also focused on closest the users approaches of the COVID-19 vaccination process on twitter as a social media platform using machine learning. Maximum of the described sentiments that debated the vaccines effectiveness, security, and the distribution plans of Governments and the plans to safe the dosages for their people. This research analyzed the users opinions since the vaccination drives was started.

VISUALIZATION:

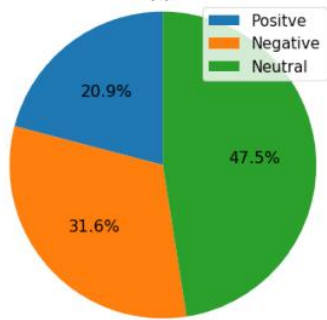
In this study, various types of analyses were performed to visually observe how data correlated with one another. Different types of diagrams, such as bar plots, line graphs, and Word Cloud, were implemented to understand patterns between our datasets. For visualization, we used many prebuilt libraries available in Python.



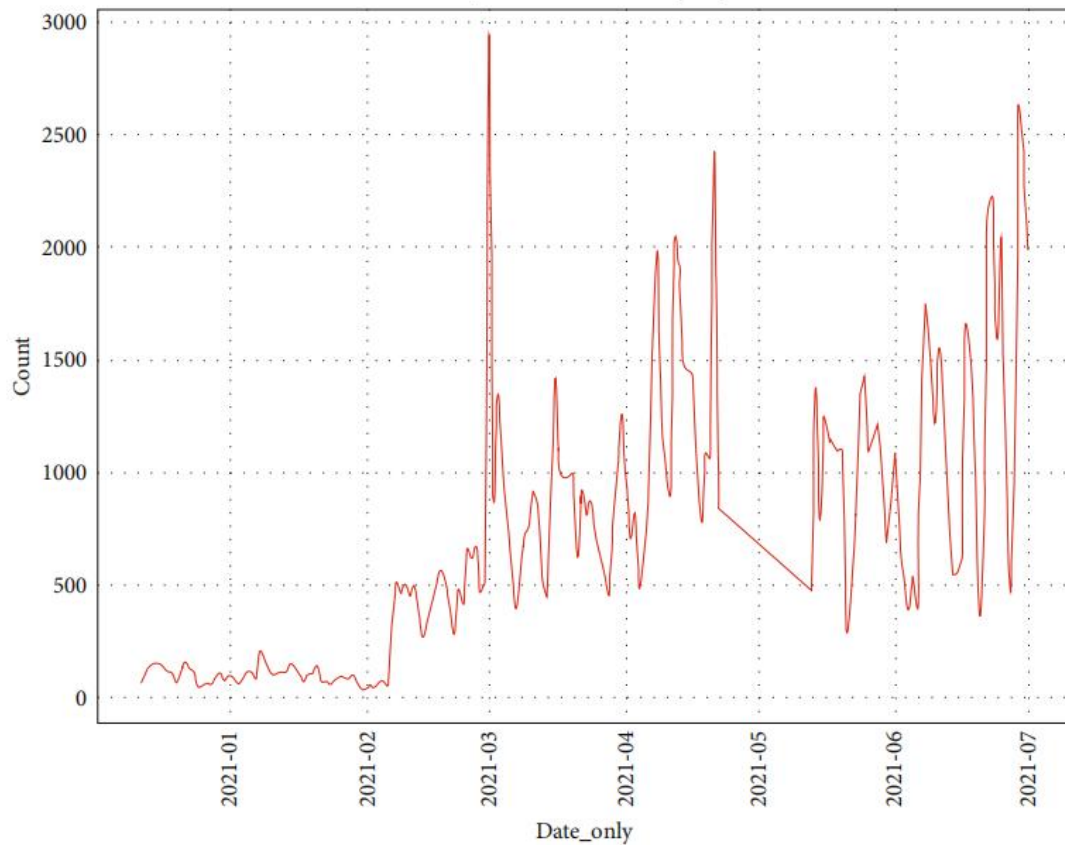
(a)



(b)



(c)



INSIGHTS AND RECOMMENDATIONS:

Vaccination of the whole population at a fast pace is encouraged by the WHO to minimize the spread and fatality risks, and governments are utilizing all available resources to accelerate COVID-19 vaccinations. Despite recommendations to take the vaccine from government officials, medical experts, and social workers, people show concerns and reservations regarding the side effects and other medical complications that may arise when vaccinated. This study proposes a methodology to analyze the global perceptions and perspectives of people towards COVID-19 vaccinations using a worldwide Twitter dataset. Dataset analysis indicates that the majority of the tweets in the collected dataset belongs to the neutral and positive classes regarding the COVID-19 vaccination. The study relies on two techniques: the NLP lexicon-based method for annotating the sentiments, and machine and deep learning models for sentiment analysis. Experimental results using TextBlob, VADER, and AFINN show that machine learning models show good performance with a TextBlob-labeled dataset with a 93% accuracy score using DT and LR. For increasing the sentiment classification accuracy, LSTM-GRNN, the ensemble of LSTM, GRU, and performance comparison with state-of-the-art models proves the model's superiority for sentiment classification with a 95% accuracy score. The decision-making process towards an effective and successful vaccination drive may be guided by engagement with the target population by listening and responding to their concerns, expectations, and difficulties related to the vaccination. Time-based sentiment analysis shows that the ratio of negative sentiments for 2022 was increased as compared to 2021.