VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

"Sentiment Analysis Of Lockdown In USA During Covid-19 A Case Study On Twitter using ML"

Submitted in partial fulfilment for the award of degree(18CSI85)
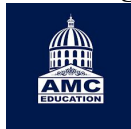
**BACHELOR OF ENGINEERING IN
YOUR BRANCH**

Submitted by:

JAZIB AHMED DANISH

1AM19IS035

Conducted at
Varcons Technologies Pvt Ltd

**AMC ENGINEERING COLLEGE
DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING**

Campus, Bannerghatta Main Rd, Bengaluru, Karnataka 560083

DECLARATION

I, **JAZIB AHMED DANISH (1AM19IS035),** final year student of Information Science and Engineering, **AMC ENGINEERING COLLEGE**-560 083, declare that the Internship has been successfully completed, in **Varcons Technologies Pvt Ltd**. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Information Science and Engineering, during the academic year 2022-2023.

Date:

Place: Bengaluru

JAZIB AHMED DANISH
(1AM19IS035)

OFFER LETTER

INTERNSHIP OFFER LETTER

Date: **2ⁿᵈ September, 2022**

Name: **Jazib Ahmed Danish**
USN: **1AM19IS035**

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With-Python(Research Based)** Internship position with **Varcons Technologies Pvt Ltd**, effective Start Date **2ⁿᵈ September, 2022**, All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With Python(Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!

Sincerely,

Spoorthi H C
Director
VARCONS TECHNOLOGIES PVT LTD
213, 2ˢᵗ Floor,
18 M G Road, Ulsoor,
Bangalore-560001

ACKNOWLEDGEMENT

It gives me immense pleasure to present before you our project titled " **Sentiment Analysis Of Lockdown In USA During Covid-19 A Case Study On Twitter using ML** ". The joy and satisfaction that accompany the successful completion of any task would be incomplete without the mention of those who made it possible. I am glad to express my gratitude towards my prestigious institution **AMC ENGINEERING COLLEGE** for providing us with utmost knowledge, encouragement and the maximum facilities in undertaking this mini project.

I wish to express sincere thanks to our respected chairman **Dr. K. R Paramahamsa** and beloved principal **Dr. Girisha C** for all their support.

I express my deepest gratitude and special thanks to **Dr. M Sathya, Professor & H.O.D, Dept. Of Information Science and Engineering**, for all her guidance and encouragement.

We would like to thank all the faculty members of our department for the support extended during the course of Internship.

We would like to thank the non-teaching members of our dept, for helping us during the Internship.

Last but not the least, we would like to thank our parents and friends without whose constant help, the completion of Internship would have not been possible.

<div align="right">

JAZIB AHMED DANISH
(1AM19IS035)

</div>

ABSTRACT

In this era of flourishing technology, Social Media has become a powerful platform for the public to voice their concerns and beliefs. Among them one such platform is Twitter. Twitter has been a popular platform for microblogging in the past few years. In this context, Sentiment Analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. Across the past few years, as the organizations and governments across the world start to adopt the ability to extract insights from social data, the applications of sentiment analysis are broad and powerful. There has been a clear implication that shifts in sentiment on social media correlate with shifts in the economics of a country and also the common notion among the public.

Due to the recent COVID-19 pandemic, there has been a wide change in sentiments of various sectors of the Indian public towards the government policies/actions. Studying the sentiment of the people on the epidemic and government decisions is very important as it acts as a sanity check for the effectiveness of the adopted government policies. This study also provides insight into the business models required to be adopted to suit this new age of post-COVID-19 where people's sentiments have widely changed. In this context 'Sentiment Analysis of COVID-19 Tweets' is a very important problem statement.

TABLE OF CONTENT

1.COMPANY PROFILE

A Brief History of Varcons Technologies Pvt Ltd

Varcons Technologies Pvt Ltd, was incorporated with a goal "To provide high quality and optimal Technological Solutions to business requirements of our clients". Every business is a different and has a unique business model and so are the technological requirements. They understand this and hence the solutions provided to these requirements are different as well. They focus on clients requirements and provide them with tailor made technological solutions. They also understand that Reach of their Product to its targeted market or the automation of the existing process into e-client and simple process are the key features that our clients desire from Technological Solution they are looking for and these are the features that we focus on while designing the solutions for their clients.

Sarvamoola Software Services. is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Sarvamoola Software Services. specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements.

Varcons Technologies Pvt Ltd, strive to be the front runner in creativity and innovation in software development through their well-researched expertise and establish it as an out of the box software development company in Bangalore, India. As a software development company, they translate this software development expertise into value for their customers through their professional solutions.

They understand that the best desired output can be achieved only by understanding the clients demand better. Compsoft Technologies work with their clients and help them to define their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solutions consulting group comprising of high caliber consultants.

They believe that Technology when used properly can help any business to scale and achieve new heights of success. It helps Improve its efficiency, profitability, reliability; to put it in one sentence" Technology helps you to Delight your customers" and that is what we want to achieve.

2. ABOUT THE COMPANY

Varcons Technologies Pvt Ltd is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Compsoft Technologies specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements. The organization where they have a right mix of professionals as a stakeholders to help us serve our clients with best of our capability and with at par industry standards. They have young, enthusiastic, passionate and creative Professionals to develop technological innovations in the field of Mobile technologies, Web applications as well as Business and Enterprise solution. Motto of our organization is to "Collaborate with our clients to provide them with best Technological solution hence creating Good Present and Better Future for our client which will bring a cascading a positive effect in their business shape as well". Providing a Complete suite of technical solutions is not just our tag line, it is Our Vision for Our Clients and for Us, we strive hard to achieve it.

Products of Varcons Technologies Pvt Ltd.

Android Apps

It is the process by which new applications are created for devices running the Android operating system. Applications are usually developed in Java (and/or Kotlin; or other such option) programming language using the Android software development kit (SDK), but other development environments are also available, some such as Kotlin support the exact same Android APIs (and bytecode), while others such as Go have restricted API access.

The Android software development kit includes a comprehensive set of development tools. These include a debugger, libraries, a handset emulator based on QEMU, documentation, sample code, and Tutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows 7 or later. As of March 2015, the SDK is not available on Android itself, but software development is possible by using specialized Android applications.

Web Application

It is a client–server computer program in which the client (including the user interface and client-side logic) runs in a web browser. Common web applications include web mail, online

retail sales, online auctions, wikis, instant messaging services and many other functions. web applications use web documents written in a standard format such as HTML and JavaScript, which are supported by a variety of web browsers. Web applications can be considered as a specific variant of client–server software where the client software is downloaded to the client machine when visiting the relevant web page, using standard procedures such as HTTP. The Client web software updates may happen each time the web page is visited. During the session, the web browser interprets and displays the pages, and acts as the universal client for any web application. The use of web application frameworks can often reduce the number of errors in a program, both by making the code simpler, and by allowing one team to concentrate on the framework while another focuses on a specified use case. In applications which are exposed to constant hacking attempts on the Internet, security related problems can be caused by errors in the program.

Frameworks can also promote the use of best practices such as GET after POST. There are some who view a web application as a two-tier architecture. This can be a "smart" client that performs all the work and queries a "dumb" server, or a "dumb" client that relies on a "smart" server. The client would handle the presentation tier, the server would have the database (storage tier), and the business logic (application tier) would be on one of them or on both. While this increases the scalability of the applications and separates the display and the database, it still doesn't allow for true specialization of layers, so most applications will outgrow this model. An emerging strategy for application software companies is to provide web access to software previously distributed as local applications. Depending on the type of application, it may require the development of an entirely different browser-based interface, or merely adapting an existing application to use different presentation technology. These programs allow the user to pay a monthly or yearly fee for use of a software application without having to install it on a local hard drive. A company which follows this strategy is known as an application service provider (ASP), and ASPs are currently receiving much attention in the software industry.

Security breaches on these kinds of applications are a major concern because it can involve both enterprise information and private customer data. Protecting these assets is an important part of any web application and there are some key operational areas that must be included in the development process. This includes processes for authentication, authorization, asset handling, input, and logging and auditing. Building security into the applications from the beginning can be more effective and less disruptive in the long run.

Web design

It is encompassing many different skills and disciplines in the production and maintenance of websites. The different areas of web design include web graphic design; interface design; authoring, including standardized code and proprietary software; user experience design; and

search engine optimization. The term web design is normally used to describe the design process relating to the front-end (client side) design of a website including writing mark up. Web design partially overlaps web engineering in the broader scope of web development. Web designers are expected to have an awareness of usability and if their role involves creating mark up then they are also expected to be up to date with web accessibility guidelines. Web design partially overlaps web engineering in the broader scope of web development.

Departments and services offered

Varcons Technologies Pvt Ltd plays an essential role as an institute, the level of education, development of student's skills are based on their trainers. If you do not have a good mentor then you may lag in many things from others and that is why we at Compsoft Technologies gives you the facility of skilled employees so that you do not feel unsecured about the academics. Personality development and academic status are some of those things which lie on mentor's hands. If you are trained well then you can do well in your future and knowing its importance of Compsoft Technologies always tries to give you the best. They have a great team of skilled mentors who are always ready to direct their trainees in the best possible way they can and to ensure the skills of mentors we held many skill development programs as well so that each and every mentor can develop their own skills with the demands of the companies so that they can prepare a complete packaged trainee.

Services provided by Varcons Technologies Pvt Ltd

• Core Java and Advanced Java
• Web services and development
• Dot Net Framework
• Python
• Selenium Testing
• Conference / Event Management Service
• Academic Project Guidance
• On The Job Training
• Software Training

<u>INTRODUCTION</u>

The outbreak of COVID-19 caused heavy disruption to the everyday lives of people across the globe. In a country like with a large, diverse population like India, there are bound to be instances of mass hysteria and panic which are further fueled by unreliable and sometimes inaccurate data. Gauging the feelings/emotions of the citizens would provide insights into the public mindset and would pave the way for the government and many organizations to address these situations by providing them with the right data and information, eradicating fake news, thereby helping in suppressing unnecessary panic among the people. Social media acts as the bridge between the people, the government, and such organizations. The scope of this project lies in the application of sentiment analysis to the views expressed by people on social media, twitter, in this case, to analyze the trends in the dynamic mood of the population. Usually, the terms "fight" and "positive" are used in a negative and positive context respectively, but we observe a role reversal in this situation. The identification of such terms and their usage according to the context would be an essential part of the project. Also, the scope of the project can be found in stopping the spread of fake news related to the pandemic, creating an interactive dashboard that delivers information about the current situation, real-time sentiment analysis of tweets, trend analysis of various COVID-19 related hashtags, engagement on Twitter, overall sector-wise polarity score of the tweets and the public emotion charts.

PROBLEM STATEMENT

Built a python application that asks for a keyword and you need to identify the sentiment of that keyword using an open source dataset.

SYSTEM ANALYSIS

Initially, the IEEE Coronavirus (COVID-19) Tweets Data set was downloaded from their website. Upon inspection, it was found that many tweets did not have geo-location tags, and also many were in different languages apart from English. Due to this challenge in obtaining proper data, a new data set named Geo-Tagged Coronavirus (COVID-19) Tweets Data set was obtained from the same website. These tweets were then hydrated using the "Hydrater" software and also a few python commands. Then, tweets in "English" and tweets from "India" were randomly chosen and a new dataset was created. Further, other data set containing COVID-19 related tweets from India were obtained from Kaggle. This data set was then cleaned and normalized to make it useful for further analysis.

This cleaned data were then subjected to further analysis by extracting bigrams, trigrams, and plotting frequency bar graphs, Word Clouds, Relationship Nexus, Boxplots, etc. This was done using both Python and R. Some Interactive Plots were also plotted. This complete process is termed as Exploratory Data Analysis.

After Exploratory Data Analysis is completed, the tweets are then Tokenized and are made in a format suitable for the Language Model. In this Step, two models were used:

1) Roberta Model: Transfer learning methods were implemented to carry out sentiment analysis. Sentiment Analysis of Tweets was carried out by integrating and using both the Huggingface Transformer Library and FastAI. Further Slanted Triangular Learning Rates, Discriminate Learning Rate and even Gradual Unfreezing were used, as a result of which, state-of-the-art results were obtained rapidly without even tuning the parameters. The tokenized data was then passed through the RoBERTa model to perform Sentiment Analysis. This yielded a model with an accuracy of 97% over the data set. The Tweepy API was used to scrape tweets in real-time which were then passed through the model to obtain the sentiments.

2) RoBERTa-CNN Sentiment Extractor: After the completion of the sentiment analysis the data was further explored for the sentiment triggers in the tweets. HuggingFace transformers don't have a TFRoberta For Question Answering, for this purpose, a TFRobertaModel was created to convert trained data into arrays that the Roberta model can interpret. While training the Sentiment Extractor model, 5 stratified KFolds were used in such a way that, in each fold, the best model weights were saved and these weights were reloaded before carrying out testing and predictions. Roberta with CNN head was used for Twitter Sentiment Extraction. Thus after passing the data through this model we obtained a new column of the extracted text for the sentiments which was also used to plot certain graphs.

Now the entire process pertaining to the data and Model building is completed. Now, the Flask APP is built for the purpose of Deployment. First, the application is deployed on the localhost and debugged and then we move on to deploying on the WebServers. A flask app was used for setting up website routing. It is used to integrate the back end machine learning

models with the dashboard. Then Socketio (web sockets) were used for dynamic implementations on the website, namely the Real-Time Plot Generators and Twitter live feed. The basic functionality of the Flask Socketio lies in running background threads when the client is not connected to the website thereby enabling dynamic plotting. The above built Dashboard was deployed on the Local Machine and debugged for any possible errors. The scraping rate and other parameters were monitored and corrected accordingly.
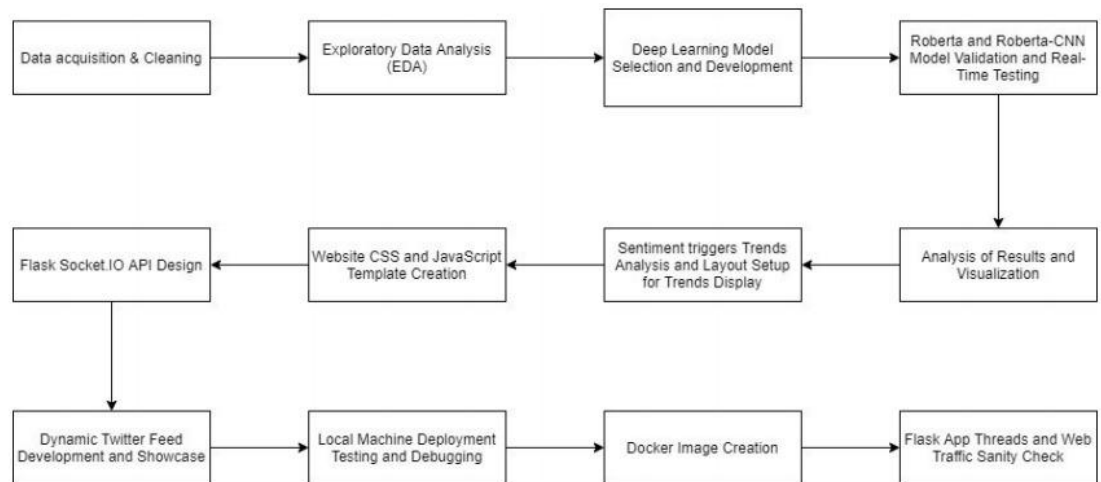
HARDWARE/SOFTWARE ANALYSIS:

HARDWARE

- **System   :   4Core Processors**
- **Hard Disk   :   142 GB**
- **Ram   :   4 GB**

SOFTWARE

- **Jupyter Notebook**
- **Python version 3.11.0 and above**

FLOW CHART

| Data acquisition & Cleaning | → | Exploratory Data Analysis (EDA) | → | Deep Learning Model Selection and Development | → | Roberta and Roberta-CNN Model Validation and Real-Time Testing |

```
Data acquisition    →    Exploratory Data    →    Deep Learning Model    →    Roberta and Roberta-CNN
& Cleaning                Analysis (EDA)            Selection and                Model Validation and Real-
                                                    Development                  Time Testing
                                                                                         │
                                                                                         ↓
Flask Socket.IO    ←    Website CSS and    ←    Sentiment triggers Trends    ←    Analysis of Results and
API Design              JavaScript Template     Analysis and Layout Setup          Visualization
                        Creation                for Trends Display
       │
       ↓
Dynamic Twitter Feed    →    Local Machine Deployment    →    Docker Image Creation    →    Flask App Threads and Web
Development and Showcase      Testing and Debugging                                          Traffic Sanity Check
```

CODE:

```python
import pandas as pd
import numpy as np
import re
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib import style
style.use('ggplot')
%matplotlib inline
from textblob import TextBlob
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from nltk.corpus import stopwords
stop_words = set(stopwords.words('english'))
from wordcloud import WordCloud
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
import csv
from sklearn.metrics import accuracy_score, classification_report,
confusion_matrix, ConfusionMatrixDisplay
df = pd.read_csv('vaccination_tweets.csv')
df.head()
df.info()
df.isnull().sum()
df.columns
text_df = df.drop(['id', 'user_name', 'user_location',
'user_description', 'user_created',
        'user_followers', 'user_friends', 'user_favourites',
'user_verified',
        'date', 'hashtags', 'source', 'retweets', 'favorites',
        'is_retweet'], axis=1)
text_df.head()
print(text_df['text'].iloc[0],"\n")
print(text_df['text'].iloc[1],"\n")
print(text_df['text'].iloc[2],"\n")
print(text_df['text'].iloc[3],"\n")
print(text_df['text'].iloc[4],"\n")
text_df.info()
def data_processing(text):
    text = text.lower()
    text = re.sub(r"https\S+|www\S+https\S+", '',text,
flags=re.MULTILINE)
    text = re.sub(r'\@w+|\#','',text)
    text = re.sub(r'[^\w\s]','',text)
    text_tokens = word_tokenize(text)
    filtered_text = [w for w in text_tokens if not w in stop_words]
    return " ".join(filtered_text)
text_df.text = text_df['text'].apply(data_processing)
text_df = text_df.drop_duplicates('text')
```

```python
stemmer = PorterStemmer()
def stemming(data):
    text = [stemmer.stem(word) for word in data]
    return data
text_df['text'] = text_df['text'].apply(lambda x: stemming(x))
text_df.head()
print(text_df['text'].iloc[0],"\n")
print(text_df['text'].iloc[1],"\n")
print(text_df['text'].iloc[2],"\n")
print(text_df['text'].iloc[3],"\n")
print(text_df['text'].iloc[4],"\n")
text_df.info()
def polarity(text):
    return TextBlob(text).sentiment.polarity
text_df['polarity'] = text_df['text'].apply(polarity)
text_df.head(10)
def sentiment(label):
    if label <0:
        return "Negative"
    elif label ==0:
        return "Neutral"
    elif label>0:
        return "Positive"
text_df['sentiment'] = text_df['polarity'].apply(sentiment)
text_df.head()
fig = plt.figure(figsize=(5,5))
sns.countplot(x='sentiment', data = text_df)
fig = plt.figure(figsize=(7,7))
colors = ("yellowgreen", "gold", "red")
wp = {'linewidth':2, 'edgecolor':"black"}
tags = text_df['sentiment'].value_counts()
explode = (0.1,0.1,0.1)
tags.plot(kind='pie', autopct='%1.1f%%', shadow=True, colors = colors,
          startangle=90, wedgeprops = wp, explode = explode, label='')
plt.title('Distribution of sentiments')
pos_tweets = text_df[text_df.sentiment == 'Positive']
pos_tweets = pos_tweets.sort_values(['polarity'], ascending= False)
pos_tweets.head()
text = ' '.join([word for word in pos_tweets['text']])
plt.figure(figsize=(20,15), facecolor='None')
wordcloud = WordCloud(max_words=500, width=1600,
height=800).generate(text)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title('Most frequent words in positive tweets', fontsize=19)
plt.show()
neg_tweets = text_df[text_df.sentiment == 'Negative']
neg_tweets = neg_tweets.sort_values(['polarity'], ascending= False)
neg_tweets.head()
text = ' '.join([word for word in neg_tweets['text']])
```

```python
plt.figure(figsize=(20,15), facecolor='None')
wordcloud = WordCloud(max_words=500, width=1600,
height=800).generate(text)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title('Most frequent words in negative tweets', fontsize=19)
plt.show()
neutral_tweets = text_df[text_df.sentiment == 'Neutral']
neutral_tweets = neutral_tweets.sort_values(['polarity'], ascending=
False)
neutral_tweets.head()
text = ' '.join([word for word in neutral_tweets['text']])
plt.figure(figsize=(20,15), facecolor='None')
wordcloud = WordCloud(max_words=500, width=1600,
height=800).generate(text)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title('Most frequent words in neutral tweets', fontsize=19)
plt.show()
vect = CountVectorizer(ngram_range=(1,2)).fit(text_df['text'])
feature_names = vect.get_feature_names()
print("Number of features: {}\n".format(len(feature_names)))
print("First 20 features:\n {}".format(feature_names[:20]))
X = text_df['text']
Y = text_df['sentiment']
X = vect.transform(X)
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.2,
random_state=42)
print("Size of x_train:", (x_train.shape))
print("Size of y_train:", (y_train.shape))
print("Size of x_test:", (x_test.shape))
print("Size of y_test:", (y_test.shape))
import warnings
warnings.filterwarnings('ignore')
logreg = LogisticRegression()
logreg.fit(x_train, y_train)
logreg_pred = logreg.predict(x_test)
logreg_acc = accuracy_score(logreg_pred, y_test)
print("Test accuracy: {:.2f}%".format(logreg_acc*100))
print(confusion_matrix(y_test, logreg_pred))
print("\n")
print(classification_report(y_test, logreg_pred))
style.use('classic')
cm = confusion_matrix(y_test, logreg_pred, labels=logreg.classes_)
disp = ConfusionMatrixDisplay(confusion_matrix = cm,
display_labels=logreg.classes_)
disp.plot()
from sklearn.model_selection import GridSearchCV
param_grid={'C':[0.001, 0.01, 0.1, 1, 10]}
grid = GridSearchCV(LogisticRegression(), param_grid)
```

Internship report 2022-2023

```python
grid.fit(x_train, y_train)
print("Best parameters:", grid.best_params_)
y_pred = grid.predict(x_test)
logreg_acc = accuracy_score(y_pred, y_test)
print("Test accuracy: {:.2f}%".format(logreg_acc*100))
print(confusion_matrix(y_test, y_pred))
print("\n")
print(classification_report(y_test, y_pred))
from sklearn.svm import LinearSVC
SVCmodel = LinearSVC()
SVCmodel.fit(x_train, y_train)
svc_pred = SVCmodel.predict(x_test)
svc_acc = accuracy_score(svc_pred, y_test)
print("test accuracy: {:.2f}%".format(svc_acc*100))
print(confusion_matrix(y_test, svc_pred))
print("\n")
print(classification_report(y_test, svc_pred))
grid = {
    'C':[0.01, 0.1, 1, 10],
    'kernel':["linear","poly","rbf","sigmoid"],
    'degree':[1,3,5,7],
    'gamma':[0.01,1]
}
grid = GridSearchCV(SVCmodel, param_grid)
grid.fit(x_train, y_train)
print("Best parameter:", grid.best_params_)
y_pred = grid.predict(x_test)
logreg_acc = accuracy_score(y_pred, y_test)
print("Test accuracy: {:.2f}%".format(logreg_acc*100))
print(confusion_matrix(y_test, y_pred))
print("\n")
print(classification_report(y_test, y_pred))
import tweepy #to access the twitter api
import pandas as pd #for basic data operations
# Importing the keys from twitter api
consumerKey = "xxxxxxxxxxxxxxxxxxxx"
consumerSecret = "xxxxxxxxxxxxxxxxxxxx"
accessToken = "xxxxxxxxxxxxxxxxxxxx"
accessTokenSecret = "xxxxxxxxxxxxxxxxxxxx"
# Establish the connection with twitter API
auth = tweepy.OAuthHandler(consumerKey, consumerSecret)
auth.set_access_token(accessToken, accessTokenSecret)
api = tweepy.API(auth)
# Search for the Term and define number of tweets
searchTerm = input("Enter Keyword/Tag to search about: ")
NoOfTerms = int(input("Enter how many tweets to search: "))
# Get no of tweets and searched term together
tweets = tweepy.Cursor(api.search_tweets, q=searchTerm).items(NoOfTerms)
```

```python
print(text_df['text'].iloc[0],"\n")
print(text_df['text'].iloc[1],"\n")
print(text_df['text'].iloc[2],"\n")
print(text_df['text'].iloc[3],"\n")
print(text_df['text'].iloc[4],"\n")
```
[8] ✓ 0.1s   Python

··· Same folks said daikon paste could treat a cytokine storm #PfizerBioNTech https://t.co/xeHhIMg1kF

While the world has been on the wrong side of history this year, hopefully, the biggest vaccination effort we've ev… https://t.co/d1CHrZjkhm

#coronavirus #SputnikV #AstraZeneca #PfizerBioNTech #Moderna #Covid_19 Russian vaccine is created to last 2-4 years… https://t.co/ieY1CKBr8P

Facts are immutable, Senator, even when you're not ethically sturdy enough to acknowledge them. (1) You were born i… https://t.co/jqgV18kch4

Explain to me again why we need a vaccine @BorisJohnson @MattHancock #whereareallthesickpeople #PfizerBioNTech… https://t.co/KxbSRoBEHq

---

```python
print(confusion_matrix(y_test, svc_pred))
print("\n")
print(classification_report(y_test, svc_pred))
```
[50] ✓ 0.3s   Python

··· [[101   91   34]
    [6 1007   8]
    [14 114 734]]

                precision   recall f1-score   support

     Negative   0.83   0.45   0.58   226
      Neutral   0.83   0.99   0.90   1021
     Positive   0.95   0.85   0.90   862

     accuracy   0.87   2109
    macro avg   0.87   0.76   0.79   2109
 weighted avg   0.88   0.87   0.87   2109

---

sentimental analysis on covid-19 from twitter using ml.ipynb - pr1 - Visual Studio Code

sentimental analysis on covid-19 from twitter using ml.ipynb ✕

sentimental analysis on covid-19 from twitter using ml.ipynb > 💠 df.info()

+ Code + Markdown | ▷ Run All ≡ Clear Outputs of All Cells ↻ Restart ☐ Interrupt | ⟦⟧ Variables ≡ Outline ···   Python 3.10.0 64-bit

```python
import pandas as pd
import numpy as np
import re
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib import style
style.use('ggplot')
%matplotlib inline
from textblob import TextBlob
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from nltk.corpus import stopwords
stop_words = set(stopwords.words('english'))
from wordcloud import WordCloud
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
import csv
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, ConfusionMatrixDisplay
```
[1] ✓ 20.4s   Python

```python
df = pd.read_csv('vaccination_tweets.csv')
```
[2] ✓ 0.5s   Python

⊗ 0 ⚠ 0   Jupyter Server: Local   Cell 4 of 60

Internship report 2022-2023

```python
print(text_df['text'].iloc[0],"\n")
print(text_df['text'].iloc[1],"\n")
print(text_df['text'].iloc[2],"\n")
print(text_df['text'].iloc[3],"\n")
print(text_df['text'].iloc[4],"\n")
```
[8] ✓ 0.1s   Python

··· Same folks said daikon paste could treat a cytokine storm #PfizerBioNTech https://t.co/xeHhIMg1kF

While the world has been on the wrong side of history this year, hopefully, the biggest vaccination effort we've ev… https://t.co/dlCHrZjkhm

#coronavirus #SputnikV #AstraZeneca #PfizerBioNTech #Moderna #Covid_19 Russian vaccine is created to last 2-4 years… https://t.co/ieYlCKBr8P

Facts are immutable, Senator, even when you're not ethically sturdy enough to acknowledge them. (1) You were born i… https://t.co/jqgV18kch4

Explain to me again why we need a vaccine @BorisJohnson @MattHancock #whereareallthesickpeople #PfizerBioNTech… https://t.co/KxbSRoBEHq

---

📄 sentimental analysis on covid-19 from twitter using ml.ipynb ✕   ⚙ ▷ ▢ ···

📄 sentimental analysis on covid-19 from twitter using ml.ipynb > ◆ neutral_tweets = text_df[text_df.sentiment == 'Neutral']
+ Code + Markdown | ▷ Run All ≡ Clear Outputs of All Cells ⟲ Restart □ Interrupt | 🖾 Variables ≔ Outline ···   🖳 Python 3.10.0 64-bit

```python
import tweepy #to access the twitter api
import pandas as pd #for basic data operations
```
[56] ✓ 0.9s   Python

```python
# Importing the keys from twitter api
consumerKey = "xxxxxxxxxxxxxxxxxxxx"
consumerSecret = "xxxxxxxxxxxxxxxxxxxx"
accessToken = "xxxxxxxxxxxxxxxxxxxx"
accessTokenSecret = "xxxxxxxxxxxxxxxxxxxx"
```
[57] ✓ 0.1s   Python

```python
# Establish the connection with twitter API
auth = tweepy.OAuthHandler(consumerKey, consumerSecret)
auth.set_access_token(accessToken, accessTokenSecret)
api = tweepy.API(auth)
```
[58] ✓ 0.2s   Python

```python
# Search for the Term and define number of tweets
searchTerm = input("Enter Keyword/Tag to search about: ")
NoOfTerms = int(input("Enter how many tweets to search: "))
```
[59] ✓ 11.9s   Python

```python
# Get no of tweets and searched term together
tweets = tweepy.Cursor(api.search_tweets, q=searchTerm).items(NoOfTerms)
```
[60] ✓ 0.9s   Python

⤬ ⊗ 0 ⚠ 0   ⌁ Jupyter Server: Local   Cell 30 of 60   ⟲ ᐃ

sentimental analysis on covid-19 from twitter using ml.ipynb > ⊕ neutral_tweets = text_df[text_df.sentiment == 'Neutral']

+ Code   + Markdown | ▷ Run All   ≣ Clear Outputs of All Cells   ↻ Restart   ⏹ Interrupt | 🖾 Variables   ≣ Outline   ⋯   🖳 Python 3.10.0 64-bit

```python
logreg_acc = accuracy_score(y_pred, y_test)
print("Test accuracy: {:.2f}%".format(logreg_acc*100))
```

[54]   ✓ 0.2s   Python

⋯   Test accuracy: 87.58%

```python
print(confusion_matrix(y_test, y_pred))
print("\n")
print(classification_report(y_test, y_pred))
```

[55]   ✓ 0.2s   Python

```
⋯   [[ 105   87   34]
     [   7 1005    9]
     [  14  111  737]]


                precision    recall  f1-score   support

     Negative       0.83      0.46      0.60       226
      Neutral       0.84      0.98      0.90      1021
     Positive       0.94      0.85      0.90       862

     accuracy                           0.88      2109
    macro avg       0.87      0.77      0.80      2109
 weighted avg       0.88      0.88      0.87      2109
```

✗   ⊗ 0 ⚠ 0   ⋈ Jupyter Server: Local   Cell 30 of 60   ⌀ ⏰

⨯ File Edit Selection View Go Run Terminal Help   sentimental analysis on covid-19 from twitter using ml.ipynb - pr1 - Visual Studio Code   ▭▭▭ ▭▭ ▭▭ ▭▭ — ⬜ ✕

sentimental analysis on covid-19 from twitter using ml.ipynb ×

sentimental analysis on covid-19 from twitter using ml.ipynb > ⊕ neutral_tweets = text_df[text_df.sentiment == 'Neutral']

+ Code   + Markdown | ▷ Run All   ≣ Clear Outputs of All Cells   ↻ Restart   ⏹ Interrupt | 🖾 Variables   ≣ Outline   ⋯   🖳 Python 3.10.0 64-bit

```python
grid = {
    'C':[0.01, 0.1, 1, 10],
    'kernel':["linear","poly","rbf","sigmoid"],
    'degree':[1,3,5,7],
    'gamma':[0.01,1]
}
grid = GridSearchCV(SVCmodel, param_grid)
grid.fit(x_train, y_train)
```

[51]   ✓ 38.4s   Python

```
⋯        ▸        GridSearchCV
         ▸ estimator: LinearSVC
                ┌─────────────┐
                │ ▸ LinearSVC │
                └─────────────┘
```

```python
print("Best parameter:", grid.best_params_)
```

[52]   ✓ 0.1s   Python

⋯   Best parameter: {'C': 10}

```python
y_pred = grid.predict(x_test)
```

[53]   ✓ 0.2s   Python

✗   ⊗ 0 ⚠ 0   ⋈ Jupyter Server: Local   Cell 30 of 60   ⌀ ⏰

Internship report 2022-2023

ADVANTAGES AND DISADVANTAGES:

ADVANTAGES:

The effect of COVID-19 pandemic is visible all over the world. National healthcare systems are facing the contagion with incredible strength, but concern regarding psychosocial and economic effects is critically growing. In a fast-moving crisis, as information swarms in from every direction, citizens look to their governments for information, guidance, and leadership. Sentimental Analysis is only the option in this current situation to understand the psychological condition/mental condition of the public. By Sentimental Analysis, the public opinion on COVID-19, regime policies, and actions can be understood. After Analysis, amendments can be made to the decisions taken by the regime policies, and the public can be fortified in such a way so as to enhance the sentiment towards a positive outlook. Not only this but also sentiment analysis will help NGOs and various organizations to come forward to help the people. Businesses can adapt their products and services to match the requirements of the people based on the real-time trending mood of the public, which will not only help businesses to grow but will also help the public meet their need of the hour. Also, this will enable the government to make business and people-friendly rules and laws to help in the betterment of the economy and the market in these untested times.

DISADVANTAGES:

Natural Language Processing models, in general, face a problem in recognizing human aspects of a language like irony, sarcasm, negotiations, exaggerations, and jokes - the sorts of things humans wouldn't face many problems in understanding. Machines sometimes fail in recognizing these aspects, which leads to skewed and incorrect results.

CONCLUSION

Sentiment analysis or opinion mining is a hot topic in deep learning. There is still a long way to go before sentiments can be accurately detected from texts, because of the complexity involved in the English language, and even more when other languages like Hindi are considered. Though the Roberta model developed as a part of this project has predicted and classified the sentiments of the test data set into positive, negative and neutral categories with an accuracy of 97%, by making necessary modifications and additions to the model, sentiment analysis can be done with greater accuracy by taking the language complexities into consideration.