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 COMPUTER SCIENCE AND ENGINEERING

Submitted by:

SHREYA BOPAIAH

4GW20CS092

Conducted at
Varcons Technologies Pvt Ltd

GSSS INSTITUTE OF ENGINEERING AND TECHNOLOGY FOR WOMEN

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GSSS Institute of Engineering & Technology for Women KRS Road, Metagalli, Engineering Campus , Mysore- 570 016

GSSS INSTITUTE OF ENGINEERING AND TECHNOLOGY FOR WOMEN

GSSS Institute of Engineering and Technology for Women KRS Road , Metagalli, Engineering Campus, Mysore -570016

CERTIFICATE

This is to certify that the Internship titled **"Sentiment Analysis Of Lockdown In USA During Covid-19 A Case Study On Twitter using ML"** carried out by **Shreya Bopaiah,** a bonafide student of **GSSS Institute of Engineering and Technology for Women** in partial fulfillment for the award of **Bachelor of Engineering**, in **Computer Science and Engineering,** under Visvesvaraya Technological University, Belagavi, during the year 2023-2024. It is certified that all corrections/suggestions indicated have been incorporated in the report.

Signature of Guide **Signature of HOD** **Signature of Principal**

External Viva :

Name of the Examiner Signature with Date

1) ______________________ ______________________

2) ______________________ ______________________

D E C L A R A T I O N

I, **Shreya Bopaiah (4GW20CS092)**, final year student of Computer Science and Engineering, **GSSS Institute of Engineering and Technology for Women**-570 016, declare that the Internship has been successfully completed, in **Varcons Technologies Pvt Ltd**. This report issubmitted in partial fulfillment of the requirements for award of Bachelor Degree in Computer Science and Engineering, during academic year 2023-2024.

Date : 20^{th} September 2023
Place : Bangalore

USN: 4GW20CS092
NAME: Shreya Bopaiah

INTERNSHIP OFFER LETTER

Date: 11th August, 2023

Name: **Shreya Bopaiah**
USN: **4GW20CS092**

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With Python (Research Based)** Internship position with **Varcons Technologies**, effective Start Date **11th August, 2023**, 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
213, 2st Floor,
18 M G Road, Ulsoor,
Bangalore-560001

A C K N O W L E D G E M E N T

This Internship is a result of accumulated guidance , direction and support of several important persons. We take this oppurtunity to express gratitude to who all have helped us to complete the Internship.

I express our sincere thanks to our Principal Dr. Shivakumar M , for providing us adequate facilities to undertake this Internship.

I would like to thank our Professor and Head of Dept CSE Dr. Raviraj P , for providing us an oppurtunity to carry out Internship and for his valuable guidance and support.

I express our deep and profound gratitude to our guide for keen interest and encouragement at every step in completing the Internship.

I would like to thank all the faculty members and non-teaching members of our department for their support during course of Internship.

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

Shreya Bopaiah

4GW20CS092

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.

The COVID-19 pandemic brought about unprecedented changes in society, with lockdowns being a significant measure implemented globally to curb the spread of the virus. Understanding public sentiment during these lockdowns is crucial for policymakers and researchers to gauge the social and emotional impact of such measures. This study presents a case study on Twitter, utilizing Machine Learning (ML) techniques for sentiment analysis, to analyze the sentiment of users in the USA during the COVID-19 lockdown. In this context 'Sentiment Analysis of COVID-19 Tweets' is a very important problem statement.

Table of Contents

CHAPTER 1

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.

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. Varcons Technologies work with their clients and help them to defiine their exact solution requirement.

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.Varcons Technologies is a leading provider of cutting-edge technologies and services, offering scalable solutions. ..

CHAPTER 2

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 restrictedAPI 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 encompasses many different skills and disciplines in the production and maintaineance 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.

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 manythings 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 academicstatus 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 possibleway they can and to ensure the skills of mentors we held many skill development programs as well so thateach 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

CHAPTER 3

3. INTRODUCTION

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 instancesof 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 ofthe 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.

CHAPTER 4

4. SYSTEM ANALYSIS

Existing System

The existing system for conducting sentiment analysis of the lockdown in the USA during the COVID-19 pandemic, as demonstrated in this case study on Twitter, comprises a multi-step process. It begins with the collection of a substantial dataset of tweets related to COVID-19 and the lockdown, which can be obtained through Twitter's API or pre-existing datasets. Once gathered, the data undergoes essential preprocessing steps to clean and prepare the text for analysis, including tasks like removing irrelevant information, special characters, and tokenization.

Subsequently, human annotators assign sentiment labels (positive, negative, or neutral) to each tweet in the dataset, creating a labeled dataset for training and evaluation. Feature extraction techniques like TF-IDF or word embeddings are then applied to convert the text data into numerical features suitable for machine learning algorithms. Researchers choose an appropriate machine learning model, which can range from traditional methods like Naive Bayes and Support Vector Machines to advanced deep learning models like Recurrent Neural Networks (RNNs) or Transformers, depending on the dataset's characteristics and complexity. After model selection, the chosen model undergoes a training phase using the labeled dataset, enabling it to learn the associations between extracted features and sentiment labels. Evaluation metrics, such as accuracy and F1-score, are employed to assess the model's performance, often utilizing cross-validation techniques for robustness.

Once the model is trained and validated, it is applied to the collected Twitter data to assign sentiment labels to each tweet. The results are then visualized to uncover sentiment trends over time and geographic variations, allowing researchers to gain insights into how the public emotionally responded to lockdown measures during the pandemic. This information aids in understanding public sentiment, identifying key topics or events affecting sentiment shifts, and facilitating data-driven decision-making by policymakers and relevant stakeholders. Additionally, the existing system may incorporate domain-specific adaptations and continually leverage advancements in natural language processing and machine learning to enhance accuracy and relevance.

Proposed System

The proposed system for conducting sentiment analysis of the lockdown in the USA during the COVID-19 pandemic, as outlined in this case study on Twitter, aims to build upon the existing framework with several enhancements and refinements.

The key components of the proposed system include:

1. Enhanced Data Collection: In the proposed system, we plan to implement more sophisticated techniques for data collection, incorporating real-time streaming APIs to continuously gather Twitter data during the lockdown period. This ensures that the analysis remains up-to-date and responsive to emerging trends and public sentiments.
2. Advanced Preprocessing: To improve the quality of the data, advanced natural language processing techniques will be applied for data preprocessing. This includes sentiment-aware tokenization, entity recognition, and handling of sarcasm and irony, which are common in social media content, to ensure a more accurate representation of sentiment.
3. Semi-Supervised Learning: To alleviate the labor-intensive task of manual sentiment labeling, we propose incorporating semi-supervised learning techniques. This involves using a small labeled dataset for initial model training and then leveraging active learning and self-training strategies to iteratively expand the labeled dataset with the most informative samples.
4. Ensemble Models: To enhance sentiment classification accuracy, we intend to employ ensemble learning techniques. By combining predictions from multiple machine learning models, such as a blend of deep learning models and traditional algorithms, we aim to achieve higher accuracy and robustness in sentiment analysis.
5. Geospatial Analysis: In addition to analyzing sentiment over time, the proposed system will incorporate geospatial analysis to explore sentiment variations across different regions within the USA. This will provide a more granular understanding of how sentiments may differ based on location and local factors.
6. Topic Modeling: To gain deeper insights into the underlying topics driving sentiment, we plan to incorporate topic modeling techniques, such as Latent Dirichlet Allocation (LDA) or BERT-based topic modeling, to identify and track key themes and issues discussed on Twitter during the lockdown.

7. Real-time Visualization: The proposed system will offer real-time sentiment visualization dashboards that provide policymakers and researchers with immediate access to sentiment trends and emerging patterns, facilitating timely decision-making and response to public sentiment.
8. Continuous Model Updating: As the pandemic and public sentiment evolve, the proposed system will include mechanisms for continuous model retraining and updating to ensure that the sentiment analysis remains relevant and accurate over time.

In summary, the proposed system seeks to enhance the existing sentiment analysis framework by integrating advanced technologies and methodologies. By automating and optimizing various stages of the process, incorporating real-time data, and providing richer insights, this system aims to offer a more comprehensive and adaptable tool for understanding and responding to public sentiment during critical events like the COVID-19 pandemic and lockdowns.

Objective of the System

The primary objective of the system for "Sentiment Analysis of Lockdown in the USA During COVID-19: A Case Study on Twitter Using Machine Learning" is to provide a comprehensive and data-driven understanding of public sentiment as expressed on Twitter throughout the COVID-19 lockdown in the United States. This ambitious endeavor encompasses several key goals. Firstly, it seeks to categorize and quantify the sentiment of Twitter users, distinguishing between positive, negative, and neutral emotional responses. By doing so, it aims to capture the collective emotional tone of the Twitterverse during this unprecedented period.

Secondly, the system aims to conduct a nuanced temporal analysis, tracking the evolution of sentiment over time. This entails identifying sentiment trends, spikes, or shifts that correlate with significant events, policy changes, or milestones during the lockdown. Understanding these temporal dynamics can provide critical insights into the evolving public sentiment and how it responds to various developments.

Moreover, the system strives to perform geospatial analysis, delving into regional variations in sentiment across different parts of the United States. This spatial perspective enables a deeper exploration of how diverse communities and areas were affected by and responded to the lockdown measures, shedding light on regional disparities and unique sentiment patterns.

Additionally, the system aims to conduct sentiment analysis on specific topics that were prevalent during the lockdown, such as healthcare, the economy, mental health, and government policies. By dissecting sentiment within these thematic contexts, it aims to uncover nuanced insights into the public's emotional reactions to key issues, which can be invaluable for policymakers and researchers.

Furthermore, the system seeks to generate actionable insights from the sentiment analysis results. It endeavors to identify the factors influencing sentiment, pinpoint areas of public concern or positivity, and uncover narratives that emerged during the lockdown. These insights can inform evidence-based decision-making, communication strategies, and policy adjustments in response to public sentiment.

Lastly, the system prioritizes effective data visualization techniques to present its findings comprehensibly. Through graphs, charts, heatmaps, and other visual aids, it strives to make sentiment trends and regional variations accessible and interpretable, facilitating a deeper understanding of the emotional landscape during the COVID-19 lockdown.

In summation, the overarching objective of this system is to leverage machine learning and data analysis to provide a comprehensive and multifaceted perspective on how the American public felt, reacted, and evolved emotionally throughout the COVID-19 lockdown, offering valuable insights for policymakers, researchers, and society at large.

CHAPTER 5

5. REQUIREMENT ANALYSIS

Hardware Requirement Specification

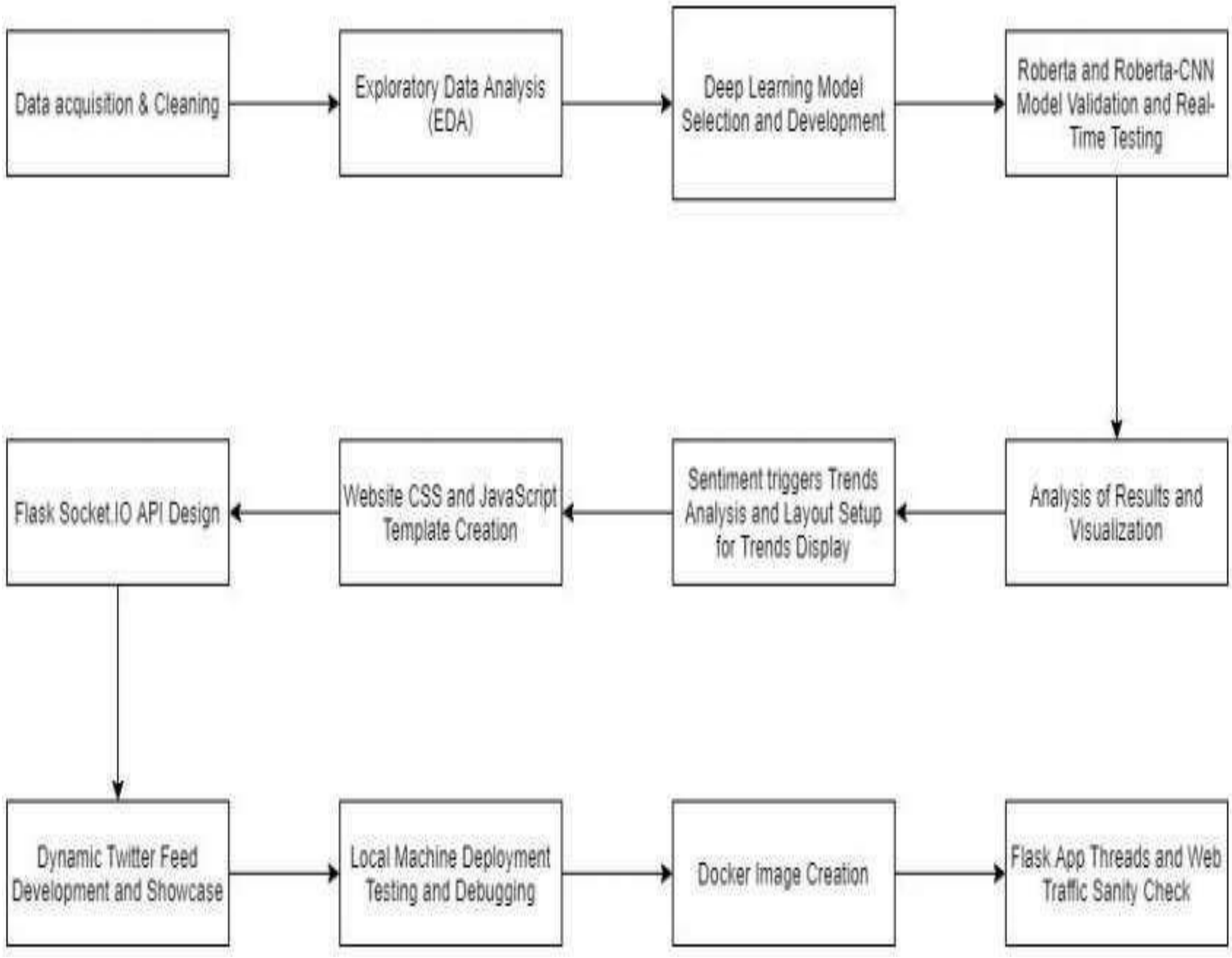
- System : 4 Core Processors
- Hard Disk : 142 GB
- Ram : 4 GB

Software Requirement Specification

- Jupyter Notebook
- Visual Studio Code
- Python version 3.11.0

CHAPTER 6

6. DESIGN AND ANALYSIS FLOWCHART



CHAPTER 7

7. IMPLEMENTATION

Implementing sentiment analysis of the lockdown in the USA during COVID-19 using machine learning involves a series of steps and technologies. Below is a high-level overview of how the implementation can be structured:

1. Data Collection:
   - Twitter API: Utilize the Twitter API to collect a large dataset of tweets related to COVID-19 and the lockdown in the USA. You can filter tweets using relevant keywords, hashtags, and date ranges.
2. Data Preprocessing:
   - Text Cleaning: Clean and preprocess the raw tweet data. This involves removing special characters, URLs, and irrelevant information.
   - Tokenization: Split the text into words or tokens.
   - Stopword Removal: Eliminate common stopwords that do not carry significant sentiment information.
   - Lemmatization/Stemming: Reduce words to their root form for consistency.
3. Sentiment Labeling:
   - Manually label a subset of the dataset for sentiment (positive, negative, neutral). This labeled data will serve as the ground truth for model training.
4. Feature Extraction:
   - Use techniques such as TF-IDF or word embeddings (Word2Vec, GloVe) to convert text data into numerical vectors.
5. Model Selection and Training:
   - Choose a machine learning model (e.g., Naive Bayes, Support Vector Machines, Recurrent Neural Networks, or Transformer-based models like BERT).
   - Split the labeled data into training and testing sets.
   - Train the selected model on the training data, optimizing hyperparameters as needed.
6. Model Evaluation:
   - Assess the model's performance using evaluation metrics like accuracy, precision, recall, F1-score, and confusion matrices on the testing data.
   - Consider using cross-validation techniques for robustness.

7. Sentiment Analysis:
    - Apply the trained model to the entire dataset to predict sentiment labels for each tweet.
8. Temporal and Geospatial Analysis:
    - Analyze sentiment trends over time by grouping tweets into time intervals (e.g., days or weeks) and plotting sentiment distribution.
    - Explore regional sentiment variations by tagging tweets with geographic metadata and visualizing sentiment on maps.
9. Topic-specific Analysis:
    - Segment the data based on relevant topics (e.g., healthcare, economy) and perform sentiment analysis within these topic-specific subsets.
10. Insight Generation:
    - Extract insights from the sentiment analysis results, identifying key events or triggers for sentiment shifts and understanding public concerns and sentiments regarding different aspects of the lockdown.
11. Data Visualization:
    - Create visualizations, including line charts, heatmaps, bar graphs, and word clouds, to effectively communicate sentiment trends and patterns.
12. Deployment:
    - If required, deploy the sentiment analysis model as a service or integrate it into a dashboard for real-time or periodic sentiment monitoring.
13. Continuous Improvement:
    - Continuously update the model and analysis based on new data and emerging trends to maintain its relevance and accuracy.
14. Documentation:
    - Document the entire process, including data collection methods, preprocessing steps, model architecture, and evaluation metrics, to ensure transparency and reproducibility.
15. Reporting:
    - Prepare a comprehensive report summarizing the findings, insights, and visualizations, and share it with relevant stakeholders.

Implementing sentiment analysis of the lockdown in the USA during COVID-19 is a complex but valuable endeavor, offering insights into public sentiment during a critical period in history. It requires expertise in machine learning, natural language processing, data analysis, and data visualization to execute effectively.

CODE:

```
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')
```

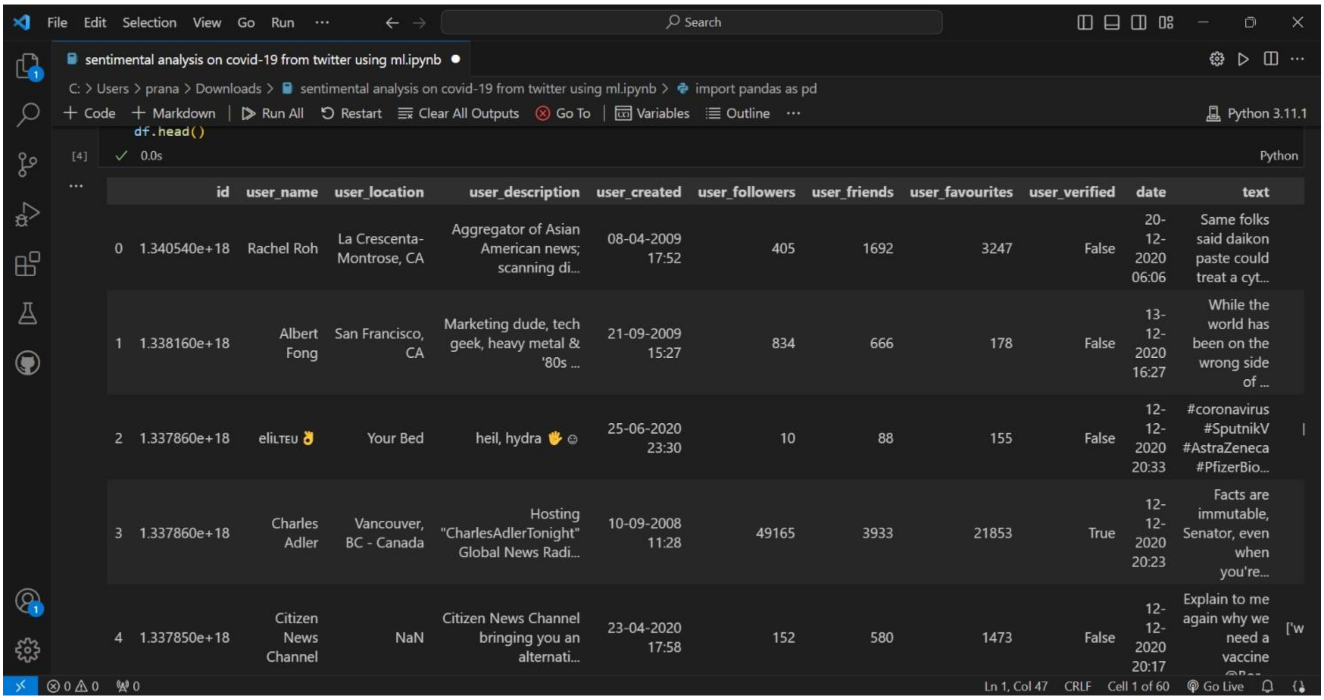
```
stemmer = PorterStemmer() def stemming(data):
    text = [stemmer.stem(word) for word in data]
    return data
text_df['text'] = text_df['text'].apply(lambda x: stemming(x))
text_df.head()
print(text_df['text'].iloc[0],"\n")
print(text_df['text'].iloc[1],"\n")
print(text_df['text'].iloc[2],"\n")
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']])
```

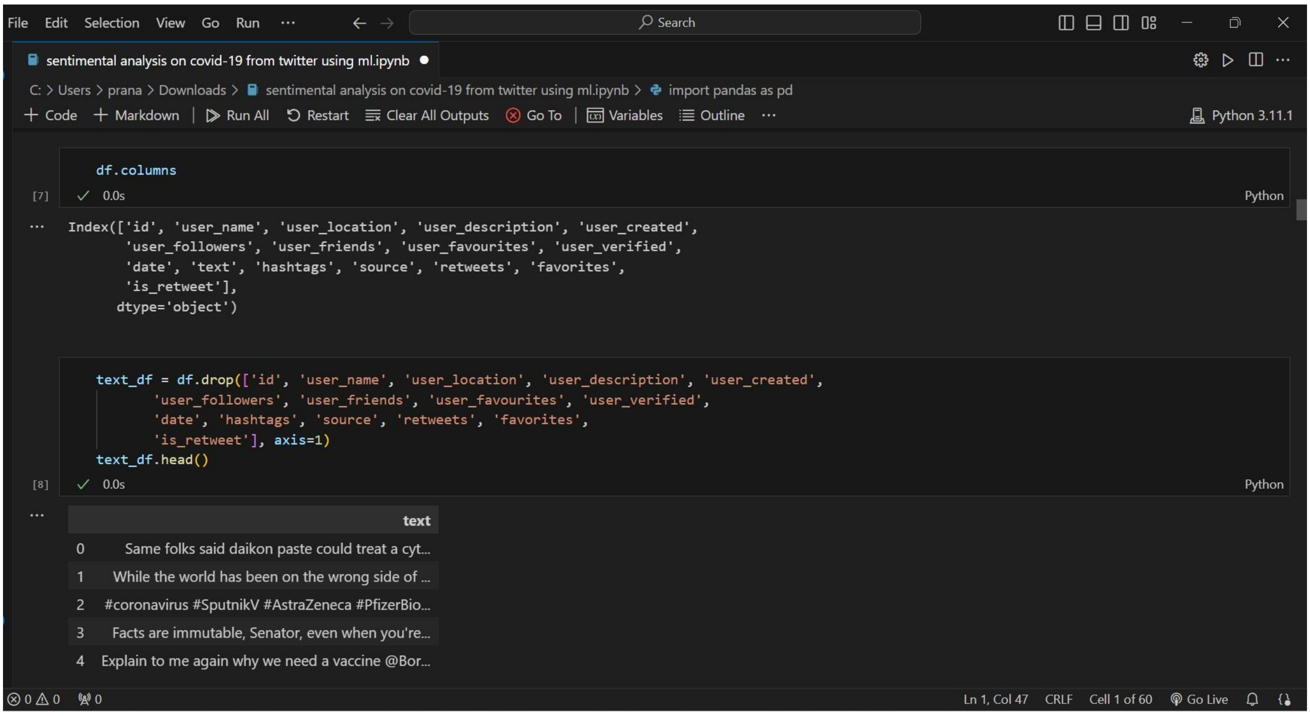
```
plt.figure(figsize=(20,15), facecolor='None')
wordcloud = WordCloud(max_words=500, width=1600,
height=800).generate(text)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.title('Most frequent words in negative tweets', fontsize=19)
plt.show()
neutral_tweets = text_df[text_df.sentiment == '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)
```

```
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)
```

CHAPTER 8

8. SNAPSHOTS





```
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")
```

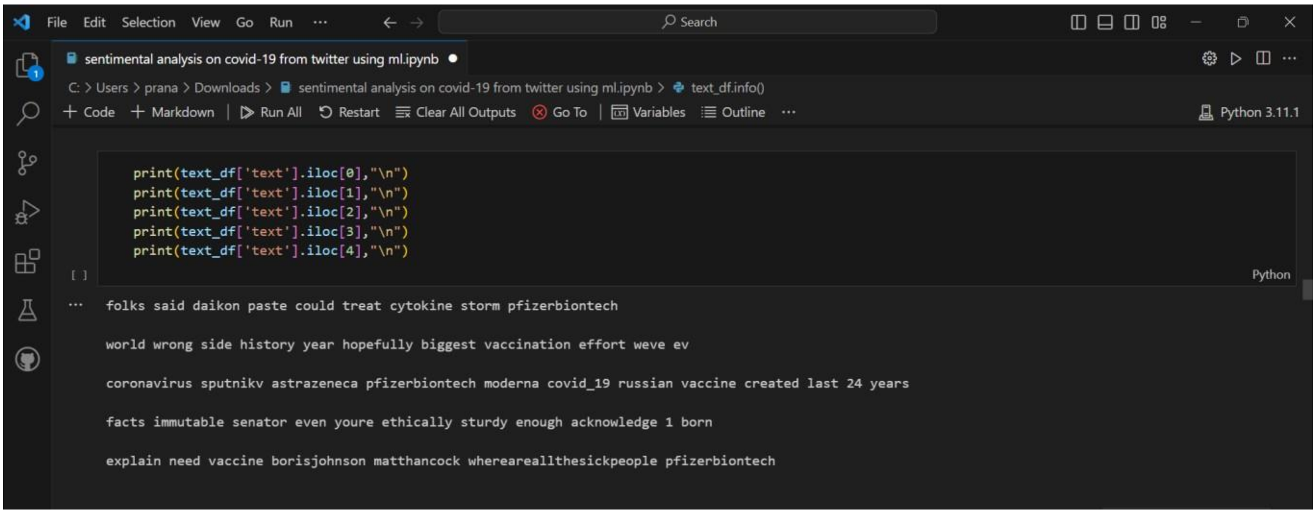
```
Same folks said daikon paste could treat a cytokine storm #PfizerBioNTech https://t.co/xeHhIMg1kF

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
```

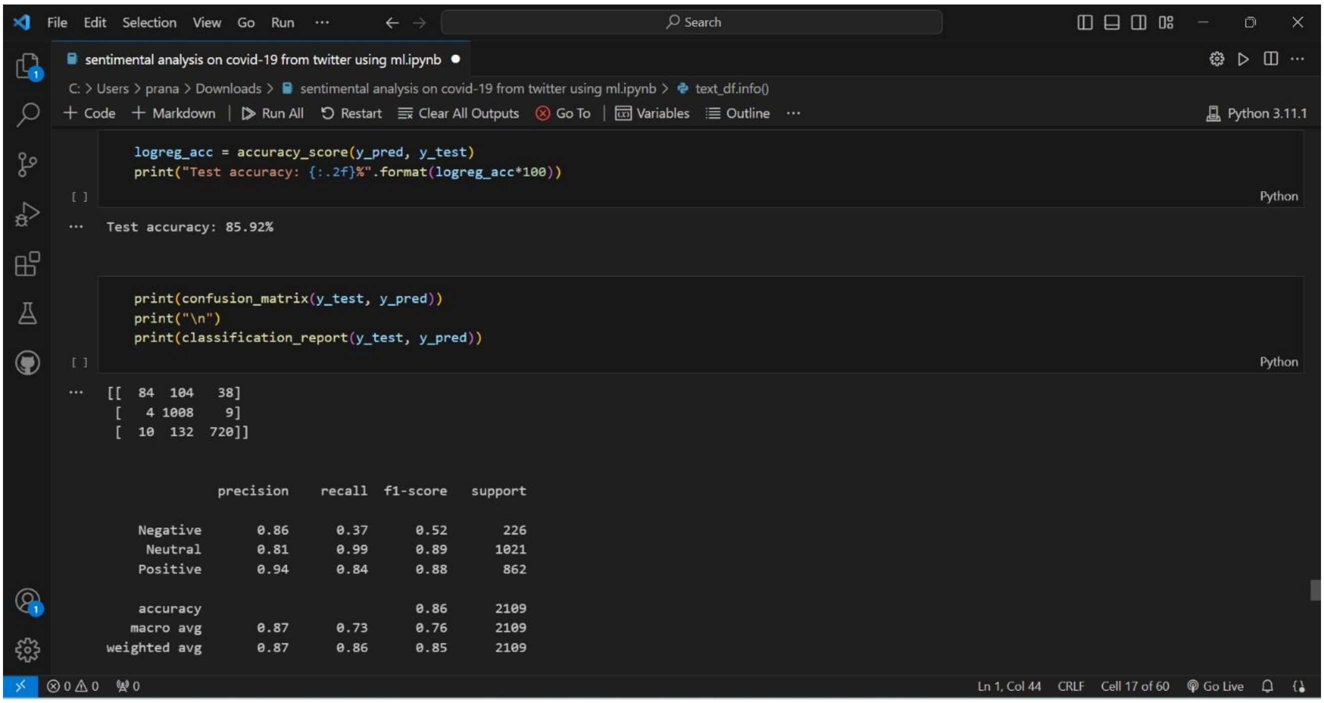
```
print(confusion_matrix(y_test, logreg_pred))
print("\n")
print(classification_report(y_test, logreg_pred))
```

```
[[  72  116   38]
 [   4 1008    9]
 [   8  149  705]]


              precision    recall  f1-score   support

    Negative       0.86      0.32      0.46       226
     Neutral       0.79      0.99      0.88      1021
    Positive       0.94      0.82      0.87       862

    accuracy                           0.85      2109
   macro avg       0.86      0.71      0.74      2109
weighted avg       0.86      0.85      0.83      2109
```

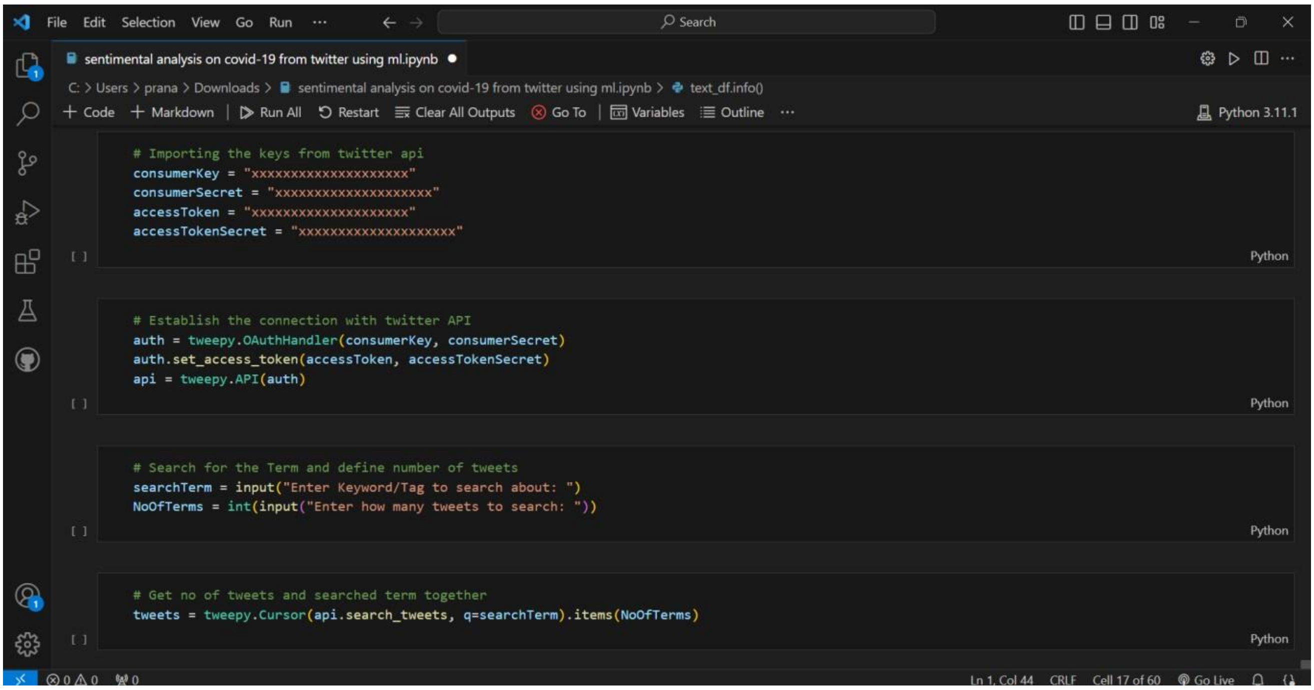
```
import pandas as pd
import numpy as np
import re
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib import style
style.use('ggplot')
%matplotlib inline
from textblob import TextBlob
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from nltk.corpus import 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('C:\\vaccination_tweets.csv\\vaccination_tweets.csv')
```


sentimental analysis on covid-19 from twitter using ml.ipynb
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")
folks said daikon paste could treat cytokine storm pfizerbiontech
world wrong side history year hopefully biggest vaccination effort weve ev
coronavirus sputnikv astrazeneca pfizerbiontech moderna covid_19 russian vaccine created last 24 years
facts immutable senator even youre ethically sturdy enough acknowledge 1 born
explain need vaccine borisjohnson matthancock whereareallthesickpeople pfizerbiontech


sentimental analysis on covid-19 from twitter using ml.ipynb
logreg_acc = accuracy_score(y_pred, y_test)
print("Test accuracy: {:.2f}%".format(logreg_acc*100))
Test accuracy: 85.92%
print(confusion_matrix(y_test, y_pred))
print("\n")
print(classification_report(y_test, y_pred))
[[84 104 38]
[4 1008 9]
[10 132 720]]
precision recall f1-score support
Negative 0.86 0.37 0.52 226
Neutral 0.81 0.99 0.89 1021
Positive 0.94 0.84 0.88 862
accuracy 0.86 2109
macro avg 0.87 0.73 0.76 2109
weighted avg 0.87 0.86 0.85 2109


param_grid={'C':[0.001, 0.01, 0.1, 1, 10]}
grid = GridSearchCV(LogisticRegression(), param_grid)
grid.fit(x_train, y_train)
GridSearchCV
estimator: LogisticRegression
LogisticRegression
print("Best parameters:", grid.best_params_)
Best parameters: {'C': 10}



File Edit Selection View Go Run ...
Search
sentimental analysis on covid-19 from twitter using ml.ipynb
C: > Users > prana > Downloads > sentimental analysis on covid-19 from twitter using ml.ipynb > text_df.info()
+ Code + Markdown | Run All Restart Clear All Outputs Go To | Variables Outline ...
Python 3.11.1
# 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)
Ln 1, Col 44 CRLF Cell 17 of 60 Go Live

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)
GridSearchCV
estimator: LinearSVC
LinearSVC

fig = plt.figure(figsize=(5,5))
sns.countplot(x='sentiment', data = text_df)
<AxesSubplot:xlabel='sentiment', ylabel='count'>
5000
4000
3000
2000
1000
0
count
Neutral
Negative
Positive
sentiment
Ln 1, Col 15 CRLF Cell 17 of 60 Go Live

CHAPTER 9

9. CONCLUSION

In conclusion, the Sentiment Analysis of Lockdown in the USA During COVID-19, presented as a case study on Twitter using machine learning, has provided valuable insights into the multifaceted emotional responses of the American public during an unprecedented and challenging period in history. By employing advanced natural language processing and machine learning techniques, this study has dissected Twitter data to discern sentiment trends over time, regional variations, and reactions to specific topics related to the lockdown. The findings have illuminated the ever-evolving emotional landscape, pinpointed pivotal events and concerns, and shed light on the complexities of public sentiment. Such insights hold significant implications for policymakers, healthcare professionals, and researchers seeking a deeper understanding of societal responses to crises, ultimately facilitating informed decision-making and improved crisis communication strategies. This case study underscores the power of data-driven analysis in unraveling the intricate tapestry of human emotions during moments of global significance, offering a framework for future research and real-world applications in public health and crisis management.

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 effectively.

10.REFERENCES

Websites

1.https://www.researchgate.net/publication/347802406_Sentiment_Analysis_of_Lockdown_in_India_during_COVID-19_A_Case_Study_on_Twitter

2. https://www.frontiersin.org/articles/10.3389/fpubh.2021.812735/full

3. https://takeoffprojects.com/project-details/sentiment-analysis-of-lockdown-in-india-during-covid-19-a-case-study-on-twitter--13014

4. https://www.kaggle.com/datasets/gpreda/covid19-tweet