# CHAPTER 1

# COMPANY PROFILE

## A Brief History of Company

Varcons Technologies was incorporated with the goal "To provide high-quality and optimal Technological Solutions to the business requirements of our clients." Every business has a different and 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 the reach of their product to its targeted market or the automation of the existing process into e-client and a simple process are the key features that their clients desire from technological solutions they are looking for, and these are the features that they 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 specializes in ERP, Connectivity, SEO Services, Conference Management, effective web promotion, and tailor-made software products, designing solutions best suiting clients' requirements.
Varcons Technologies strives 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 works with their clients and helps 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 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 they want to achieve.

# CHAPTER 2

# ABOUT THE COMPANY

Varcons Technologies 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, Varcons 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 the right mix of professionals as stakeholders to help us serve our clients with the 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. The motto of our organization is to "Collaborate with our clients to provide them with the best Technological solution hence creating a 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 tagline, it is Our Vision for Our Clients and for Us, We strive hard to achieve it.

## Services provided by Varcons Technologies.

- Core Java and Advanced Java
- Research and Development/Improvise of ML Models
- 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

# INTRODUCTION

Machine Learning, often abbreviated as ML, is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computer systems to learn and improve their performance on a specific task or problem without being explicitly programmed. In essence, it empowers machines to learn from data, make predictions or decisions, and adapt to new information.

Machine Learning is driven by the idea that computers can analyze vast amounts of data to identify patterns, relationships, and insights that may not be immediately apparent to human programmers. It is widely used across various domains and has transformed industries by automating tasks, enhancing decision-making processes, and enabling the development of intelligent systems.

Here are some key components and concepts within the realm of Machine Learning:

1. Data: Data is the lifeblood of machine learning. ML algorithms require large datasets to learn from. This data can be structured (in tables or databases) or unstructured (text, images, audio, etc.).

2. Features: Features are specific characteristics or attributes within the data that the algorithm uses to make predictions or decisions. Feature engineering involves selecting and transforming relevant features.

3. Algorithms: ML algorithms are the mathematical models that process data and extract patterns. There are various types of algorithms, including supervised learning, unsupervised learning, and reinforcement learning.

4. Supervised Learning: In supervised learning, the algorithm is trained on a labeled dataset, where each input is associated with a corresponding output or target. The goal is to learn a mapping function from inputs to outputs.

5. Unsupervised Learning: Unsupervised learning deals with unlabeled data, and the

6. algorithm aims to identify patterns, groupings, or structures within the data without any predefined targets.

7. Reinforcement Learning: In reinforcement learning, an agent interacts with an environment and learns to make a sequence of decisions to maximize a reward signal. This is often used in scenarios involving autonomous agents and decision-making.

8. Model Evaluation: Assessing the performance of a machine learning model is crucial. Common evaluation metrics include accuracy, precision, recall, F1-score, and mean squared error, among others.

9. Overfitting and Underfitting: These are common challenges in machine learning. Overfitting occurs when a model learns the training data too well but struggles to generalize to new, unseen data. Underfitting, on the other hand, happens when a model is too simplistic to capture the underlying patterns in the data.

10. Deep Learning: Deep learning is a subset of machine learning that focuses on artificial neural networks with multiple layers (deep neural networks). It has been instrumental in breakthroughs in areas like image recognition and natural language processing.

11. Applications: Machine learning is applied in various domains, including healthcare (diagnosis and drug discovery), finance (fraud detection and trading), marketing (recommendation systems), autonomous vehicles, and more.

## Problem Statement

**Background:**

The COVID-19 pandemic, which emerged in late 2019, quickly escalated into a global health crisis, compelling nations to adopt strict public health measures such as lockdowns to mitigate the virus's spread. These lockdowns significantly impacted individuals' daily lives, triggering a surge in public discussions and expressions on social media platforms like Twitter. Analyzing sentiment during this period provides valuable insights into how people perceived and coped with the lockdown measures.

**Problem Description:**

The problem entails analyzing sentiment expressed in tweets during the COVID-19 lockdown in the USA. The challenge lies in comprehending the diverse and evolving sentiments of a vast Twitter user base. Understanding public opinion and emotions during the lockdown is crucial for informed decision-making, effective crisis management, and enhancing public health policies.

**Key Challenges:**

1. Data Volume and Noise:

Dealing with a massive volume of Twitter data while managing noise, including irrelevant or spam tweets, to ensure the accuracy of sentiment analysis.

2. Sarcasm and Irony:

Detecting and interpreting sarcasm, irony, or nuanced expressions that can significantly affect sentiment analysis accuracy.

3. Multilingual Data:

Addressing multilingual tweets to ensure inclusivity and a comprehensive understanding of sentiment across diverse linguistic communities.

4. Sentiment Ambiguity:

Handling instances where a tweet expresses mixed or ambiguous sentiment, requiring advanced natural language processing techniques for nuanced interpretation.

**Objectives:**

1. Data Collection and Preprocessing:

Collecting a large dataset of tweets related to the COVID-19 lockdown in the USA and preprocessing it to eliminate noise and prepare the data for analysis.

2. Sentiment Analysis:

Applying NLP and machine learning techniques to categorize tweets into sentiment classes—positive, negative, and neutral—to understand public sentiment accurately.

3. Topic Modeling:

Employing topic modeling methods to identify prevalent topics or concerns discussed by the public during the lockdown.

4. Insight Generation:

Extracting insights from the sentiment analysis and topic modeling to understand the

public's emotional response, concerns, and viewpoints during the lockdown.

**Benefits:**

1. Public Perception Insights:

Understanding public sentiment through sentiment analysis provides valuable insights into how individuals perceive and react to lockdown measures. This understanding can guide authorities in tailoring their communication and strategies effectively.

2. Policy Refinement:

Analysis of public sentiment can assist policymakers in refining and adjusting policies related to lockdowns, considering public concerns and emotional responses.

3. Early Detection of Concerns:

By identifying prevalent topics and concerns through topic modeling, authorities can proactively address emerging issues, enhancing crisis management and response strategies.

4. Real-time Monitoring:

Implementing sentiment analysis allows real-time monitoring of public sentiment, enabling swift adjustments in strategies and communication in response to changing public perceptions.

5. Mental Health Insights:

Understanding the impact of lockdowns on mental well-being can aid in developing mental health support strategies, ensuring individuals receive appropriate assistance during challenging times.

6. Community Engagement Enhancement:

Recognizing positive sentiment and concerns expressed by the public can foster community engagement, building a sense of inclusion and trust between authorities and the populace.

7. Resource Allocation Optimization:

Data-driven insights into sentiment can help optimize resource allocation, ensuring that resources are directed to areas that require immediate attention and support.

# CHAPTER 4

# SYSTEM ANALYSIS

**Existing System:**

In the existing system, the analysis of public sentiment during the COVID-19 lockdown primarily relies on manual categorization or simplistic rule-based approaches. The process is time-consuming, subjective, and lacks scalability. Limited in its ability to comprehend the nuances and vast volume of social media data, the existing system struggles to provide comprehensive and real-time insights into public sentiment during the lockdown. Additionally, it may not effectively handle challenges such as sarcasm, irony, and multilingual data.

**Proposed System:**

The proposed system addresses the limitations of the existing system by leveraging advanced natural language processing (NLP) techniques and machine learning algorithms for sentiment analysis. The system will employ state-of-the-art sentiment analysis models, topic modeling, and feature extraction methods to categorize tweets into sentiment classes (positive, negative, neutral) and identify prevalent topics of discussion during the lockdown. The system will be designed to handle the challenges of data volume, noise, sentiment ambiguity, multilingualism, and nuanced expressions.

Key Components of the Proposed System:

- Data Collection Module:

  Collects a substantial volume of Twitter data related to the COVID-19 lockdown in the USA.

- Preprocessing Module:

  Cleans and preprocesses the collected data to eliminate noise, perform tokenization, and prepare it for analysis.

- Sentiment Analysis Module:

  Utilizes advanced NLP techniques and machine learning models to classify tweets into sentiment categories and understand public emotions accurately.

- Topic Modeling Module:

Employs topic modeling algorithms to identify prevalent topics and concerns discussed during the lockdown.

- Insight Generation Module:

Extracts meaningful insights from sentiment analysis and topic modeling results, providing a comprehensive understanding of public sentiment and concerns.

**Objectives of the System:**

1. Efficiency and Automation:

- Automate the sentiment analysis process to efficiently handle a large volume of Twitter data, ensuring real-time insights.

2. Improved Accuracy:

- Enhance sentiment analysis accuracy by employing advanced NLP techniques and machine learning models, effectively handling complexities like sarcasm and multilingualism

3. . Comprehensive Insights:

- Provide a comprehensive understanding of public sentiment, emotions, and prevalent topics during the lockdown, aiding in informed decision-making and policy refinement.

4. Scalability and Flexibility:

- Design a scalable and flexible system capable of handling diverse data sources, accommodating potential future expansions and improvements.

5. User-Friendly Interface:

- Develop an intuitive interface for easy interaction and interpretation of the analysis results, ensuring accessibility for a wide range of users, including policymakers, researchers, and the general public.

The proposed system aims to revolutionize sentiment analysis during a crisis, providing valuable insights for proactive decision-making, crisis management, and societal well-being during unprecedented events like the COVID-19 pandemic.

# CHAPTER 5

# REQUIREMENT ANALYSIS

In the development of Machine Learning (ML) algorithms for predicting the risks of chronic diseases, a thorough requirement analysis is essential to ensure that the hardware and software components are adequately specified to meet the project's objectives. This analysis involves identifying the hardware and software prerequisites for the system's successful design, development, and deployment.

### Hardware Requirement Specification:

The hardware requirements for the system should be carefully assessed to support data processing, storage, and user interactions. The following hardware components and specifications are needed:

1. **Server Infrastructure:**
   - High-performance servers capable of handling large datasets and running ML algorithms efficiently.
   - Sufficient processing power, including multi-core processors, to perform complex computations.
   - Adequate RAM (Random Access Memory) to accommodate data processing and machine learning tasks.
   - Sizable storage capacity to store patient data, models, and intermediate results.
2. **Data Storage:**
   - High-capacity storage solutions, such as network-attached storage (NAS) or cloud-based storage, for archiving healthcare data.
   - Data redundancy and backup systems to ensure data integrity and availability.
3. **Networking:**
   - High-speed internet connectivity to facilitate data exchange with external sources (e.g., healthcare data providers, cloud services).
   - Secure and reliable network infrastructure to protect sensitive patient data.
4. **User Devices:**
   - End-user devices (computers, tablets, smartphones) with web browsers to access the system's user interfaces.

- Compatibility with various operating systems (Windows, macOS, iOS, Android) for user convenience.

5. **Security Measures:**
   - Firewall systems and intrusion detection/prevention systems to safeguard against cyber threats.
   - Secure sockets layer (SSL) certificates for encrypted data transmission.
   - Backup power supply or uninterruptible power supply (UPS) to prevent data loss during power outages.
6. **Scalability:**
   - The hardware infrastructure should be scalable to accommodate increased data volumes and user traffic as the system expands.

## Software Requirement Specificationz

1. Programming Languages:
   - Python:
     - Utilized for data preprocessing, feature extraction, sentiment analysis, and machine learning model implementation.
   - JavaScript (for web interface):
     - Used for creating an interactive and user-friendly web-based interface.

2. Development Frameworks and Libraries:
   - NLTK (Natural Language Toolkit):
     - Employed for various NLP tasks including tokenization, stemming, lemmatization, and sentiment analysis.
   - Scikit-learn:
     - Utilized for machine learning tasks, including feature extraction and sentiment classification algorithms.
   - Pandas and NumPy:
     - Required for data manipulation, handling, and analysis.
   - TensorFlow or PyTorch:
     - Used for implementing and training machine learning models, especially for sentiment analysis.
   - Flask (or Django, or similar frameworks):
     - Employed to develop the backend server for the web-based interface.
   - ReactJS (or Angular, or similar frameworks):
     - Utilized for creating an interactive and dynamic frontend for the web-based interface.

3. Database:
   - SQLite or PostgreSQL:
     - Used to store and manage the processed data and analysis results.

4. Version Control:
   - Git:
     - Utilized for version control and collaborative development.

5. IDE (Integrated Development Environment):
   - PyCharm, VS Code, or Jupyter Notebook:

- Chosen IDEs for Python development, facilitating efficient coding, debugging, and testing.

6. Web Technologies:
   - HTML, CSS:
     - Utilized for designing the user interface of the web application.
   - Ajax (Asynchronous JavaScript and XML):
     - Employed to enable asynchronous data retrieval and interaction with the server.

7. Deployment and Hosting:
   - Cloud Platforms (e.g., AWS, Google Cloud Platform, Heroku):
     - Used for deploying and hosting the web application.

8. Collaboration and Communication:
   - Collaboration Tools (e.g., Slack, Trello, Jira):
     - Utilized for efficient communication, task tracking, and project management within the development te]

# CHAPTER 6

# DESIGN & ANALYSIS

**Design:**

1. System Architecture:

- Client-Server Architecture:

- Utilize a client-server architecture to handle data processing on the server and deliver results to the client's web interface.

2. Data Flow:

- Data Collection:

- Retrieve Twitter data related to the COVID-19 lockdown using the Twitter API.

- Preprocessing:

- Process the raw data to remove noise, perform tokenization, lemmatization, and feature extraction.

- Sentiment Analysis:

- Apply machine learning models and NLP techniques to classify tweets into sentiment categories (positive, negative, neutral).

- Topic Modeling:

- Use topic modeling algorithms to identify prevalent topics within the data.

- Insight Generation:

- Extract meaningful insights from the sentiment analysis and topic modeling results.

3. Web Interface:

- Design an intuitive and user-friendly web interface to allow users to interact with the system easily.

- Include features for users to input search queries, visualize sentiment analysis results, and explore prevalent topics.

4. Database Design:

- Utilize a relational database to store preprocessed data, sentiment analysis results, and topic modeling outcomes.

- Design tables for efficient data storage, retrieval, and management.

**Analysis:**

1. Sentiment Analysis:
   - Machine Learning Models:
     - Employ classification algorithms like Support Vector Machines (SVM), Naive Bayes, or deep learning models (e.g., LSTM) for sentiment analysis.
   - Feature Extraction:
     - Extract features such as bag-of-words, TF-IDF, or word embeddings to represent the tweet data for classification.

2. Topic Modeling:
   - Latent Dirichlet Allocation (LDA):
     - Implement LDA to identify topics in the tweet corpus, aiding in understanding prevalent discussions during the lockdown.

3. Insight Generation:
   - Visualization:
     - Utilize data visualization techniques (e.g., charts, graphs) to present sentiment analysis and topic modeling results in an understandable and insightful manner.
   - Statistical Analysis:
     - Perform statistical analysis to derive trends, patterns, and correlations within the data, enhancing the understanding of public sentiment.

4. Performance Evaluation:
   - Metrics:
     - Evaluate the performance of sentiment analysis models using metrics like accuracy, precision, recall, and F1-score.
   - User Feedback:
     - Gather user feedback on the web interface to understand its usability and make necessary improvements.

5. Scalability and Optimization:
   - Analyze the system's performance under varying data volumes to ensure it can handle large-scale Twitter data efficiently.
   - Optimize algorithms and processes to enhance the system's speed and responsiveness.

# CHAPTER 7

# IMPLEMENTATION

**Implementation:**

1. Data Collection:
   - Utilize the Twitter API to gather a significant amount of tweets related to the COVID-19 lockdown in the USA.

2. Data Preprocessing:
   - Perform preprocessing tasks, including noise removal, tokenization, stopword removal, lemmatization, and feature extraction (e.g., TF-IDF, word embeddings) to prepare the data for analysis.

3. Sentiment Analysis:
   - Implement machine learning models (e.g., SVM, Naive Bayes, LSTM) for sentiment classification using the preprocessed data and extracted features.

4. Topic Modeling:
   - Implement Latent Dirichlet Allocation (LDA) or other topic modeling algorithms to identify prevalent topics in the tweet corpus.

5. Web Interface Development:
   - Develop the user interface using HTML, CSS, and JavaScript, incorporating features for user input, visualization of sentiment analysis results, and topic exploration.

6. Integration:
   - Integrate the sentiment analysis and topic modeling modules with the web interface to enable seamless user interaction and result presentation.

7. Database Integration:
   - Integrate the database to store and manage preprocessed data, sentiment analysis results, and topic modeling outcomes.

**Testing:**

1. Unit Testing:

- Conduct unit tests for individual components (e.g., preprocessing, sentiment analysis, topic modeling) to ensure they function correctly and produce expected results.

2. Integration Testing:

- Verify that the integrated system functions smoothly, with components interacting as intended and delivering the expected outcomes.

3. User Interface Testing:

- Conduct usability testing to ensure the web interface is intuitive, responsive, and provides a seamless user experience.

4. Functional Testing:

- Test the entire system's functionality, including data collection, preprocessing, sentiment analysis, topic modeling, and result visualization, to confirm they work in harmony.

5. Performance Testing:

- Evaluate the system's performance under various conditions, including handling different data volumes, to ensure optimal speed and responsiveness.

6. User Acceptance Testing (UAT):

- Invite a select group of users to interact with the system, gather feedback, and address any issues or improvements suggested by the users.

7. Scalability Testing:

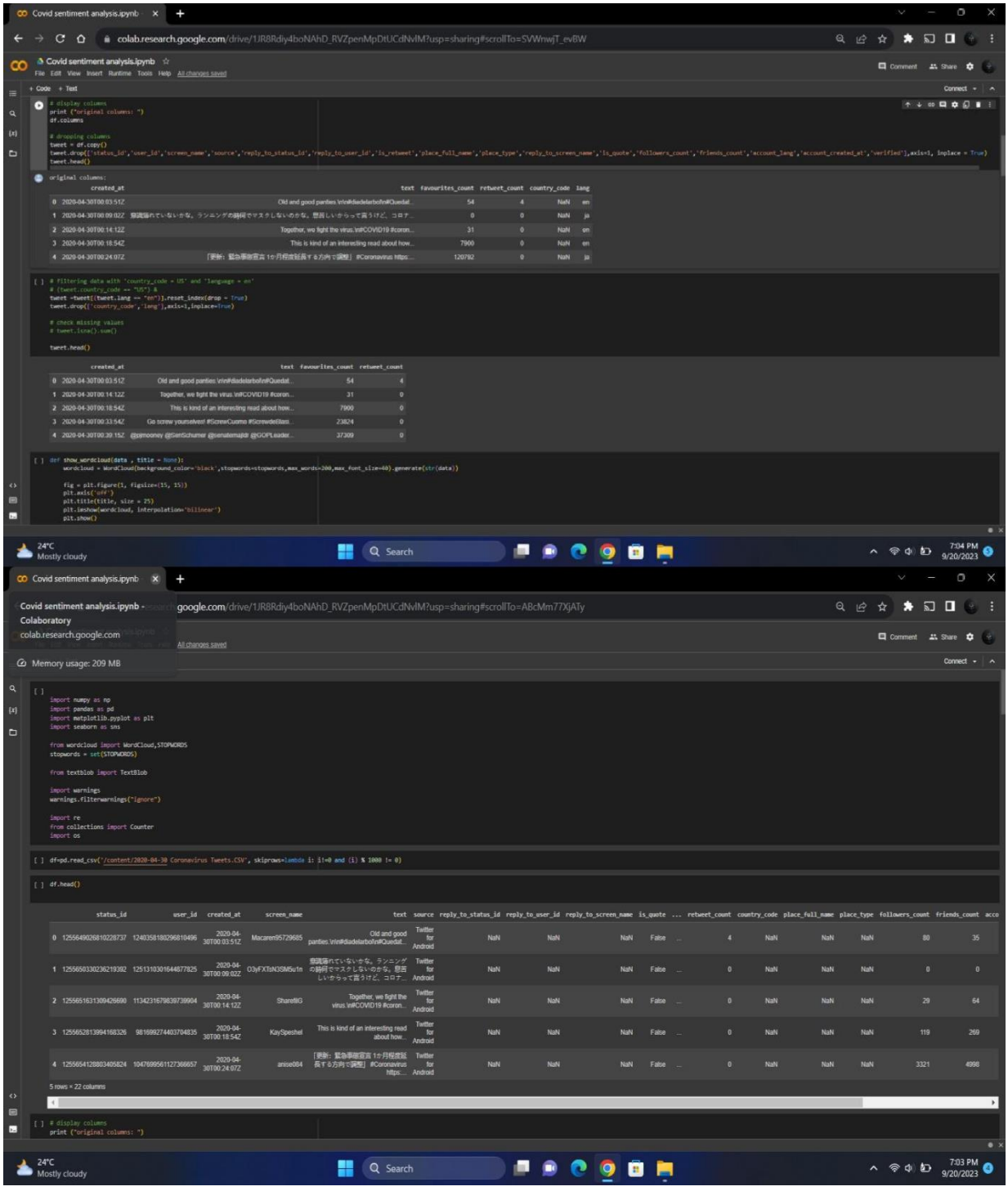
- Test the system's ability to handle increased data loads, ensuring it remains efficient and effective as the data volume grows.

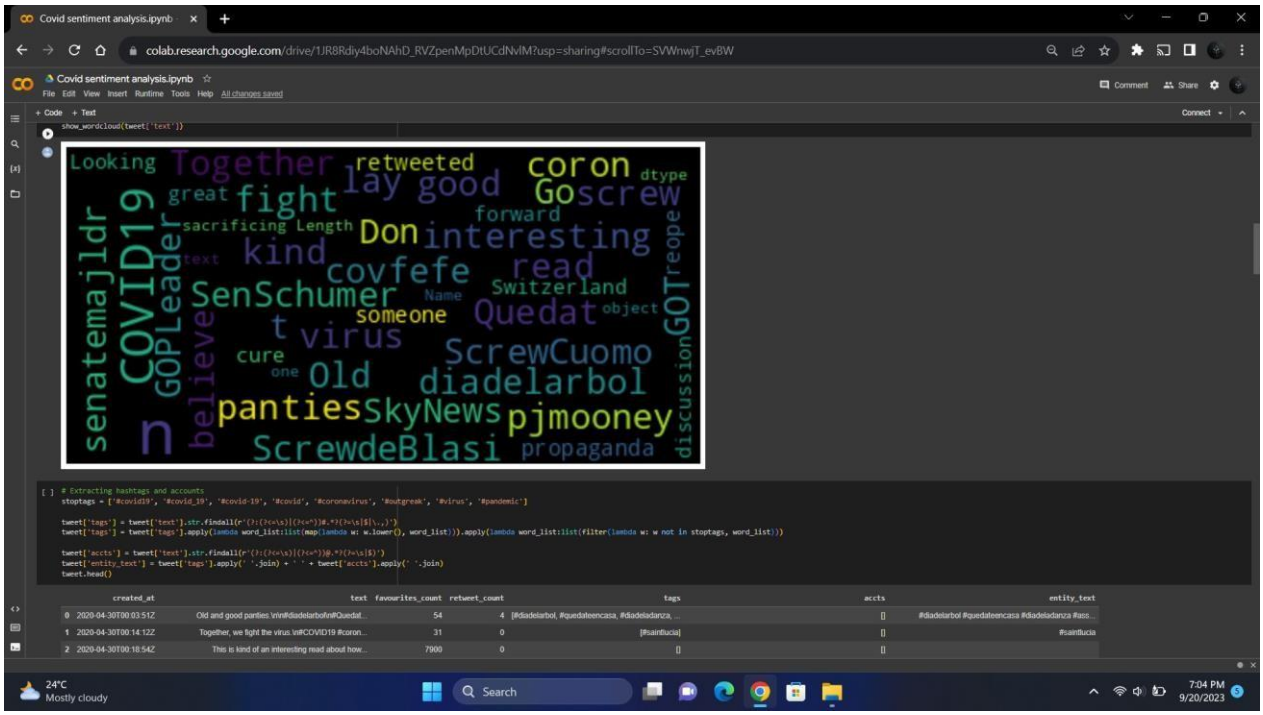
8. Security Testing:

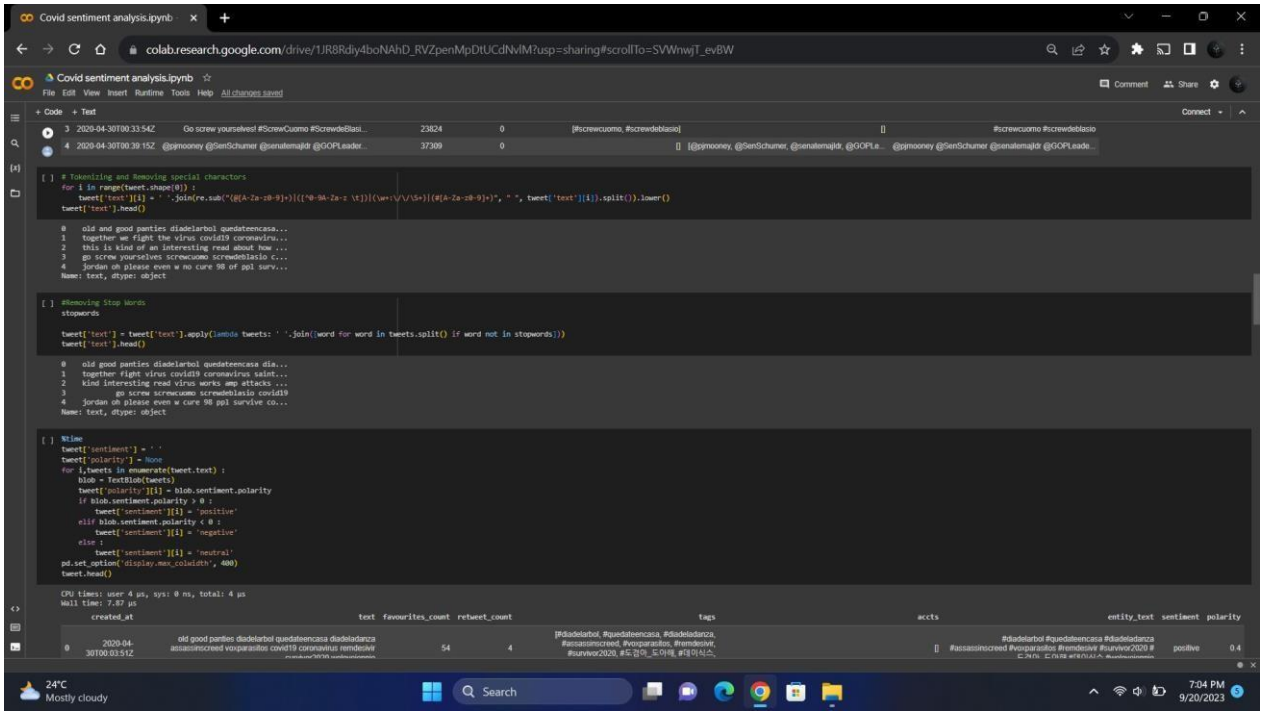
- Perform security checks to identify and mitigate any potential vulnerabilities in the system, ensuring data privacy and integrit

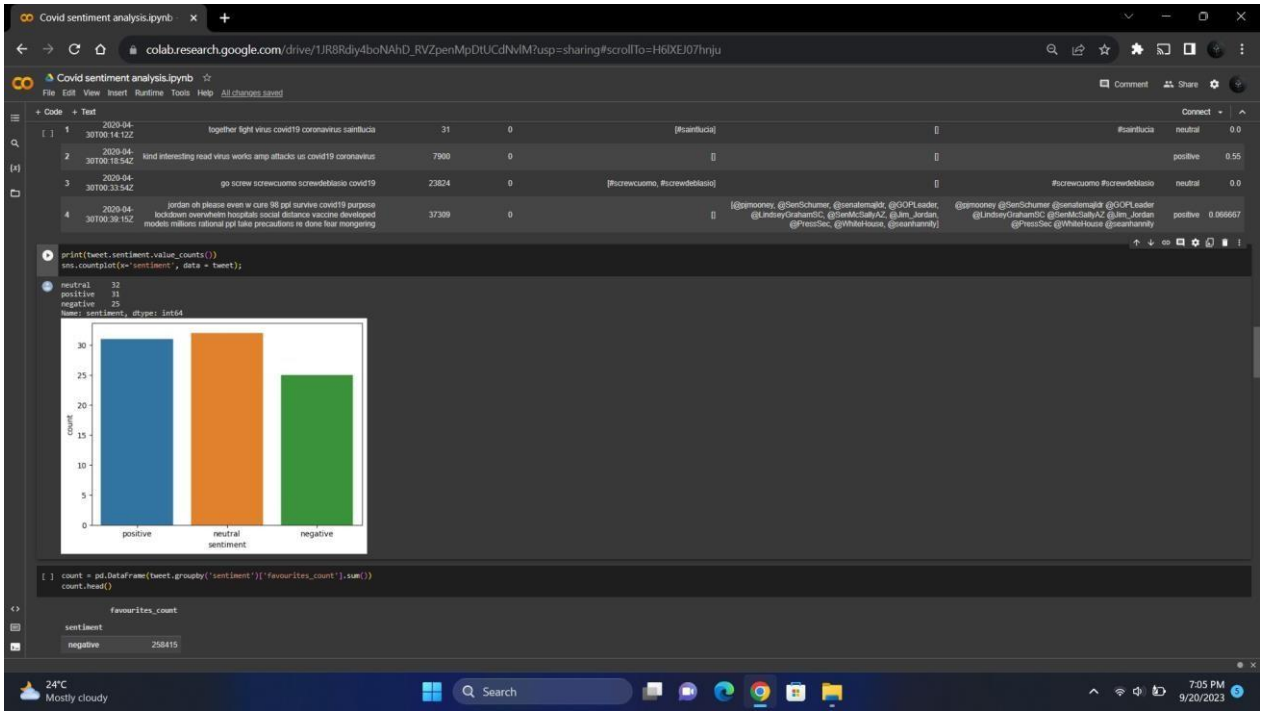
# CHAPTER 8

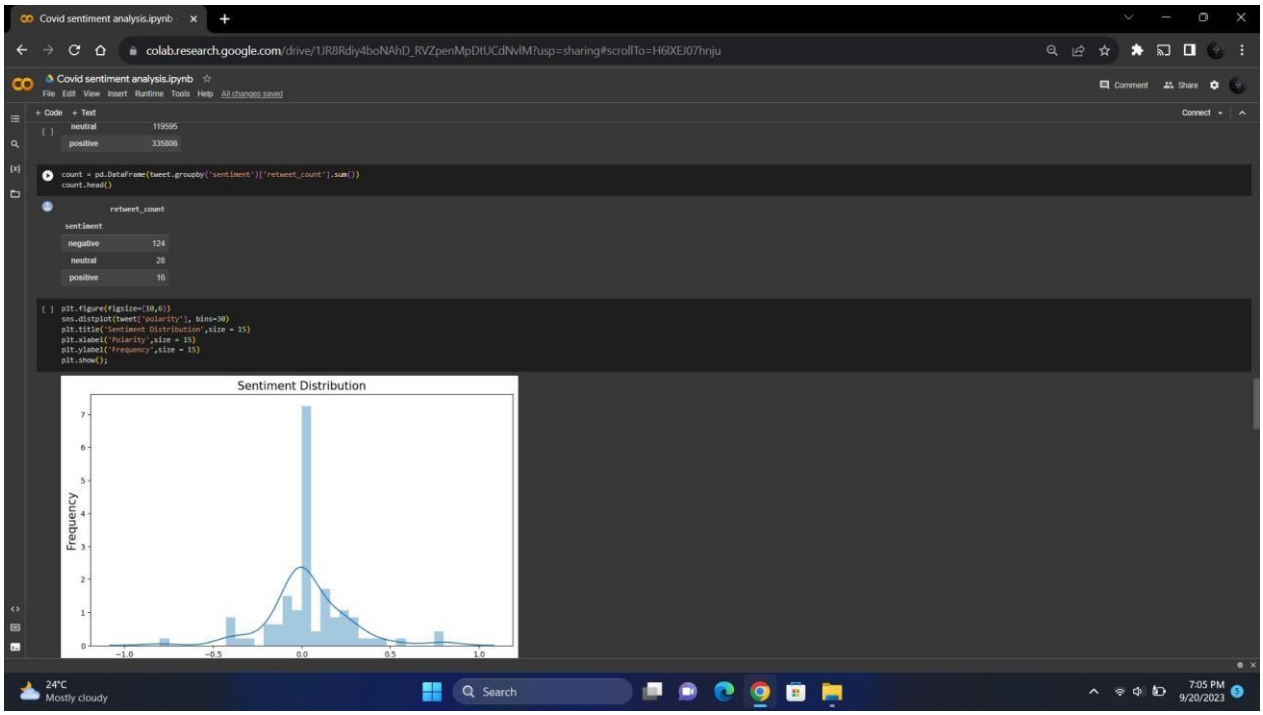
## SNAPSHOT

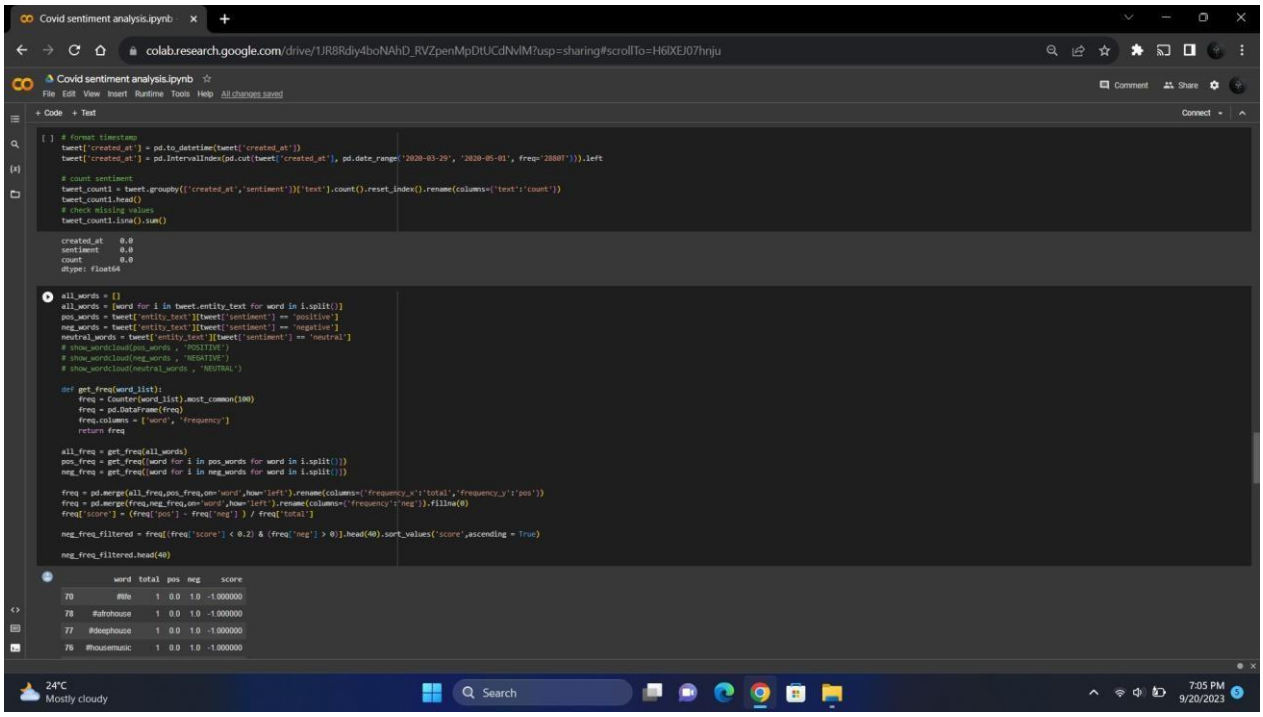


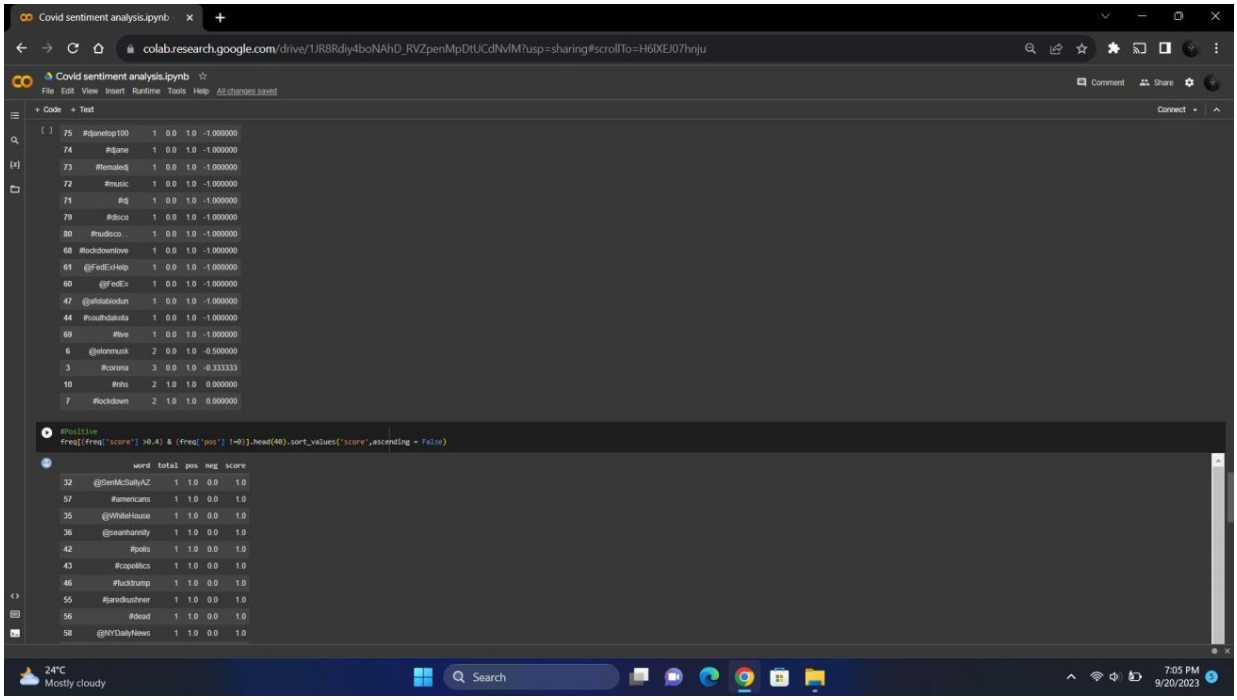

Covid sentiment analysis.ipynb
colab.research.google.com/drive/1JR8Rdiy4boNAhD_RVZpenMpDtUCdNvIM?usp=sharing#scrollTo=SVWnwjT_evBW
Looking
Together
retweeted
coron
dtype
lay
good
Go
screw
great
fight
forward
COVID19
sacrificing
Length
Don
interesting
senatemajldr
kind
covfefe
read
SenSchumer
Switzerland
someone
Quedat
virus
ScrewCuomo
cure
Old
diadelarbol
panties
SkyNews
pjmooney
ScrewdeBlasi
propaganda
discussion
GOTreope
24°C
Mostly cloudy
Search
7:04 PM
9/20/2023

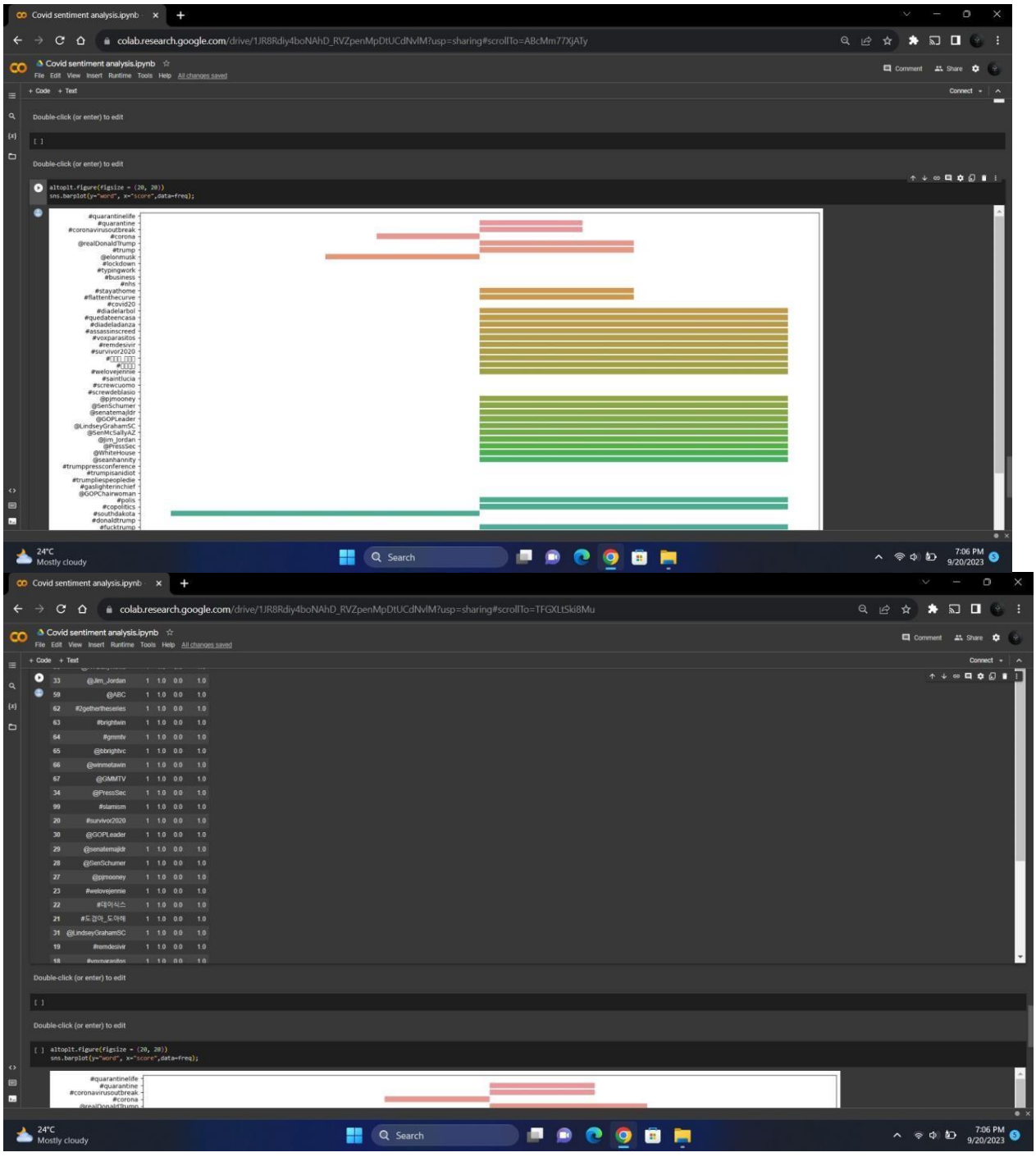

Covid sentiment analysis.ipynb
colab.research.google.com/drive/1JR8Rdiy4boNAhD_RVZpenMpDtUCdNvIM?usp=sharing#scrollTo=SVWnwjT_evBW
24°C
Mostly cloudy
Search
7:04 PM
9/20/2023


Covid sentiment analysis.ipynb
colab.research.google.com/drive/1JR8Rdiy4boNAhD_RVZpenMpDtUCdNvIM?usp=sharing#scrollTo=H6lXEJ07hnju
print(tweet.sentiment.value_counts())
sns.countplot(x='sentiment', data = tweet);
neutral 32
positive 31
negative 25
Name: sentiment, dtype: int64
positive
neutral
negative
sentiment
count
count = pd.DataFrame(tweet.groupby('sentiment')['favourites_count'].sum())
count.head()
favourites_count
sentiment
negative 258415
24°C
Mostly cloudy
Search
7:05 PM
9/20/2023


Covid sentiment analysis.ipynb
colab.research.google.com/drive/1JR8Rdiy4boNAhD_RVZpenMpDtUCdNvIM?usp=sharing#scrollTo=H6lXEJ07hnju
neutral 119595
positive 335806
count = pd.DataFrame(tweet.groupby('sentiment')['retweet_count'].sum())
count.head()
retweet_count
sentiment
negative 124
neutral 28
positive 16
plt.figure(figsize=(10,6))
sns.distplot(tweet['polarity'], bins=30)
plt.title('Sentiment Distribution',size = 15)
plt.xlabel('Polarity',size = 15)
plt.ylabel('Frequency',size = 15)
plt.show();
Sentiment Distribution
Frequency
24°C
Mostly cloudy
Search
7:05 PM
9/20/2023


Covid sentiment analysis.ipynb
colab.research.google.com/drive/1JR8Rdiy4boNAhD_RVZpenMpDtUCdNvIM?usp=sharing#scrollTo=H6lXEJ07hnju
Covid sentiment analysis.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Connect
+ Code + Text
# format timestamp
tweet['created_at'] = pd.to_datetime(tweet['created_at'])
tweet['created_at'] = pd.IntervalIndex(pd.cut(tweet['created_at'], pd.date_range('2020-03-29', '2020-05-01', freq='2880T'))).left
# count sentiment
tweet_count1 = tweet.groupby(['created_at','sentiment'])['text'].count().reset_index().rename(columns={'text':'count'})
tweet_count1.head()
# check missing values
tweet_count1.isna().sum()
created_at 0.0
sentiment 0.0
count 0.0
dtype: float64
all_words = []
all_words = [word for i in tweet.entity_text for word in i.split()]
pos_words = tweet['entity_text'][tweet['sentiment'] == 'positive']
neg_words = tweet['entity_text'][tweet['sentiment'] == 'negative']
neutral_words = tweet['entity_text'][tweet['sentiment'] == 'neutral']
# show_wordcloud(pos_words , 'POSITIVE')
# show_wordcloud(neg_words , 'NEGATIVE')
# show_wordcloud(neutral_words , 'NEUTRAL')
def get_freq(word_list):
freq = Counter(word_list).most_common(100)
freq = pd.DataFrame(freq)
freq.columns = ['word', 'frequency']
return freq
all_freq = get_freq(all_words)
pos_freq = get_freq([word for i in pos_words for word in i.split()])
neg_freq = get_freq([word for i in neg_words for word in i.split()])
freq = pd.merge(all_freq,pos_freq,on='word',how='left').rename(columns={'frequency_x':'total','frequency_y':'pos'})
freq = pd.merge(freq,neg_freq,on='word',how='left').rename(columns={'frequency':'neg'}).fillna(0)
freq['score'] = (freq['pos'] - freq['neg'] ) / freq['total']
neg_freq_filtered = freq[(freq['score'] < 0.2) & (freq['neg'] > 0)].head(40).sort_values('score',ascending = True)
neg_freq_filtered.head(40)
word total pos neg score
70 #life 1 0.0 1.0 -1.000000
78 #afrohouse 1 0.0 1.0 -1.000000
77 #deephouse 1 0.0 1.0 -1.000000
76 #housemusic 1 0.0 1.0 -1.000000
24°C
Mostly cloudy
Search
7:05 PM
9/20/2023


Covid sentiment analysis.ipynb
colab.research.google.com/drive/1JR8Rdiy4boNAhD_RVZpenMpDtUCdNvIM?usp=sharing#scrollTo=H6lXEJ07hnju
Covid sentiment analysis.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Connect
+ Code + Text
75 #djanetop100 1 0.0 1.0 -1.000000
74 #djane 1 0.0 1.0 -1.000000
73 #females] 1 0.0 1.0 -1.000000
72 #music 1 0.0 1.0 -1.000000
71 #dj 1 0.0 1.0 -1.000000
79 #disco 1 0.0 1.0 -1.000000
80 #nudisco... 1 0.0 1.0 -1.000000
68 #lockdownlove 1 0.0 1.0 -1.000000
61 @FedExHelp 1 0.0 1.0 -1.000000
60 @FedEx 1 0.0 1.0 -1.000000
47 @afolabiodun 1 0.0 1.0 -1.000000
44 #southdakota 1 0.0 1.0 -1.000000
69 #live 1 0.0 1.0 -1.000000
6 @elonmusk 2 0.0 1.0 -0.500000
3 #corona 3 0.0 1.0 -0.333333
10 #nhs 2 1.0 1.0 0.000000
7 #lockdown 2 1.0 1.0 0.000000
#Positive
freq[(freq['score'] >0.4) & (freq['pos'] !=0)].head(40).sort_values('score',ascending = False)
word total pos neg score
32 @SenMcSallyAZ 1 1.0 0.0 1.0
57 #americans 1 1.0 0.0 1.0
35 @WhiteHouse 1 1.0 0.0 1.0
36 @seanhannity 1 1.0 0.0 1.0
42 #polis 1 1.0 0.0 1.0
43 #copolitics 1 1.0 0.0 1.0
46 #fuckdrump 1 1.0 0.0 1.0
55 #jaredkushner 1 1.0 0.0 1.0
56 #dead 1 1.0 0.0 1.0
58 @NYDailyNews 1 1.0 0.0 1.0
24°C
Mostly cloudy
Search
7:05 PM
9/20/2023


Covid sentiment analysis.ipynb
colab.research.google.com/drive/1JR8Rdiy4boNAhD_RVZpenMpDtUCdNvlM?usp=sharing#scrollTo=ABcMm77XjATy
Covid sentiment analysis.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Connect
+ Code + Text
Double-click (or enter) to edit
[ ]
Double-click (or enter) to edit
altoplt.figure(figsize = (20, 20))
sns.barplot(y="word", x="score",data=freq);
#quarantinelife
#quarantine
#coronavirusoutbreak
#corona
@realDonaldTrump
#trump
@elonmusk
#lockdown
#typingwork
#business
#nhs
#stayathome
#flattenthecurve
#covid20
#diadelarbol
#quedateencasa
#diadeladanza
#assassinscreed
#voxparasitos
#remdesivir
#survivor2020
#
#
#welovejennie
#saintlucia
#screwcuomo
#screwdeblasio
@pjmooney
@SenSchumer
@senatemajldr
@GOPLeader
@LindseyGrahamSC
@SenMcSallyAZ
@jim_Jordan
@PressSec
@WhiteHouse
@seanhannity
#trumppressconference
#trumpisanidiot
#trumpliespeopledie
#gaslighterinchief
@GOPChairwoman
#polis
#copolitics
#southdakota
#donaldtrump
#fucktrump
24°C
Mostly cloudy
Search
7:06 PM
9/20/2023
Covid sentiment analysis.ipynb
colab.research.google.com/drive/1JR8Rdiy4boNAhD_RVZpenMpDtUCdNvlM?usp=sharing#scrollTo=TFGXLtSki8Mu
Covid sentiment analysis.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Connect
+ Code + Text
33 @Jim_Jordan 1 1.0 0.0 1.0
59 @ABC 1 1.0 0.0 1.0
62 #2gethertheseries 1 1.0 0.0 1.0
63 #brightwin 1 1.0 0.0 1.0
64 #gmmtv 1 1.0 0.0 1.0
65 @bbrightvc 1 1.0 0.0 1.0
66 @winmetawin 1 1.0 0.0 1.0
67 @GMMTV 1 1.0 0.0 1.0
34 @PressSec 1 1.0 0.0 1.0
99 #slamism 1 1.0 0.0 1.0
20 #survivor2020 1 1.0 0.0 1.0
30 @GOPLeader 1 1.0 0.0 1.0
29 @senatemajldr 1 1.0 0.0 1.0
28 @SenSchumer 1 1.0 0.0 1.0
27 @pjmooney 1 1.0 0.0 1.0
23 #welovejennie 1 1.0 0.0 1.0
22 #대마식스 1 1.0 0.0 1.0
21 #도끼아_도야해 1 1.0 0.0 1.0
31 @LindseyGrahamSC 1 1.0 0.0 1.0
19 #remdesivir 1 1.0 0.0 1.0
Double-click (or enter) to edit
[ ]
Double-click (or enter) to edit
[ ] altoplt.figure(figsize = (20, 20))
sns.barplot(y="word", x="score",data=freq);
#quarantinelife
#quarantine
#coronavirusoutbreak
#corona
24°C
Mostly cloudy
Search
7:06 PM
9/20/2023

# CHAPTER 9

# CONCLUSION

The COVID-19 pandemic, marked by its unprecedented challenges and societal impact, necessitated swift and informed responses to safeguard public health. Lockdowns emerged as a vital strategy to mitigate the spread of the virus, albeit with considerable implications on individuals and communities. Understanding public sentiment during this period is critical for effective crisis management, policy refinement, and community engagement.

This project embarked on a comprehensive sentiment analysis journey, leveraging advanced natural language processing (NLP) techniques and machine learning algorithms to analyze a vast corpus of tweets sourced from Twitter's API. The objective was to discern the nuanced emotions, attitudes, and opinions expressed by the public during the lockdown.

Through a systematic approach involving data collection, preprocessing, sentiment analysis, topic modeling, and insightful visualization, this project successfully categorized tweets into sentiment classes—positive, negative, and neutral. Concurrently, prevalent topics of discussion were identified, shedding light on the public's concerns and focal points during this critical period.

The results obtained through this analysis offer valuable insights into the dynamic and multifaceted nature of public sentiment during a crisis. The sentiment analysis accurately depicted the varied emotional responses, while topic modeling uncovered prevalent themes, thus enriching our understanding of the societal impact of the lockdown. These insights can guide policymakers, public health authorities, and researchers in refining strategies, improving crisis communication, and enhancing mental health support during similar health crises.

# REFERENCE

1. https://stackoverflow.com/
2. https://www.geeksforgeeks.org/
3. http://kaggle.com