VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

"Sentiment Analysis of Lockdown in USA During Covid-19 A Case Study on Twitter"

Submitted in partial fulfilment for the award of degree(18CSI85)

BACHELOR OF ENGINEERING IN
COMPUTER SCIENCE AND ENGINEERING

Submitted by:

VANI R

4UB20CS064

Conducted at
VARCONS TECHNOLOGIES PVT LTD

UNIVERSITY BDT COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DAVANGERE

UNIVERSITY BDT COLLEGE OF ENGINEERING
Department of Computer Science and Engineering
Accredited by NBA, New Delhi
Davangere

CERTIFICATE

This is to certify that the Internship titled **"Sentiment Analysis of Lockdown In USA During Covid-19 A Case Study On Twitter"** carried out by **Vani R,** a bonafide student of **University BDT College of Engineering**, 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.
The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (18CSI85)

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 **Vani R**, final year student of Computer Science and Engineering, Davangere, 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 Computer Science and Engineering, during the academic year 2023-2024.

Date: 20-09-2023

Place: Davangere

USN: 4UB20CS064

NAME: Vani R

OFFER LETTER

Internship Offer Letter

Date: 11th August, 2023

Name: **Vani R**
USN: **4UB20CS064**

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 opportunity to express our gratitude to all who have helped us to complete the Internship.

We express our sincere thanks to our principal, for providing us adequate facilities to undertake this Internship.

We would like to thank our Head of Dept – E061, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We would like to thank our Software Services for guiding us during the period of internship.

We express our deep and profound gratitude to our guide, Sri. Naveen Kumar B, Assistant/Associate Prof, for her keen interest and encouragement at every step in completing the Internship.

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.

Vani R
4UB20CS064

ABSTRACT

This internship report provides a comprehensive overview of my experiences and contributions during my internship at Varcons Technologies Pvt. Ltd., where I had the opportunity to delve into the exciting world of machine learning and its practical applications in the corporate setting. Varcons Technologies is a dynamic technology company known for its innovative solutions in various industries, and this internship allowed me to witness firsthand how machine learning can be a transformative force in these domains.

The report begins by offering an introduction to Varcons Technologies, its core areas of expertise, and its motivation for incorporating machine learning into its operations. It then delves into my role as an intern, outlining the objectives and goals of the internship project, which primarily focused on the development of machine learning models to enhance business processes and decision-making within the organization.

The methodology section of the report provides an in-depth discussion of the tools, frameworks, and data sources used during the internship. It elucidates the various stages of the machine learning pipeline, including data collection and preprocessing, feature engineering, model selection, and evaluation metrics. Particular attention is paid to the challenges encountered and the strategies employed to overcome them.

The report then presents the results and findings of the machine learning project. It highlights the successful implementation of machine learning models in optimizing specific business processes, such as demand forecasting, customer segmentation, and anomaly detection. Key performance metrics and improvements in decision-making processes are also discussed.

In addition to the technical aspects, the report reflects on the valuable insights gained from working alongside experts in the field and the significance of collaborative problem-solving. It emphasizes the importance of effective communication and teamwork in the successful execution of machine learning projects within a corporate environment.

The internship experience at Varcons Technologies not only provided a deep understanding of the practical applications of machine learning but also honed my analytical and problem-solving skills. Furthermore, it reinforced the notion that machine learning is not merely a technical tool but a strategic asset that can drive business growth and innovation.

In conclusion, this internship report at Varcons Technologies Pvt. Ltd. underscores the transformative potential of machine learning in the corporate world and highlights the importance of continuous learning and adaptation in the rapidly evolving field of artificial intelligence. The experience has not only enriched my knowledge but has also inspired a deep appreciation for the intersection of technology and business.

Table of Contents

CHAPTER 1

COMPANY PROFILE

1. COMPANY PROFILE

A Brief History of Company

Company, 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.

Company 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.
we 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 understandingthe clients demand better. At our Company we work with them clients and help them to defiine 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.

CHAPTER 2

ABOUT THE COMPANY

2. ABOUT THE COMPANY

We are a Technology Organization providing solutions for all web design and development, Researching and Publishing Papers to ensure the quality of most used ML Models, 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.

Services provided by Varcons Technologies Pvt Ltd

- Website as Software
- Analytics and Research
- Comprehensive Customer Support
- Smart Automation Tools

CHAPTER 3

INTRODUCTION

3. INTRODUCTION

Introduction to ML

Machine learning is a dynamic field of artificial intelligence that empowers computers to learn from data and improve their performance over time without explicit programming. It encompasses a range of algorithms and techniques that enable systems to recognize patterns, make predictions, and make data-driven decisions, revolutionizing industries from healthcare to finance and beyond. In essence, it's the driving force behind the development of intelligent, data- driven applications that continually evolve and adapt.

Problem Statement

In the wake of the COVID-19 pandemic, the United States witnessed a series of lockdowns and restrictive measures to curb the virus's spread. Understanding the sentiment of the public during these lockdown periods is crucial for policymakers, businesses, and healthcare professionals. This project aims to address the following problem statement:

To perform sentiment analysis on Twitter data related to the lockdowns in the United States during the COVID-19 pandemic, utilizing machine learning techniques, in order to gain insights into public sentiment, emotions, and opinions, and to facilitate informed decision-making for future crises and policy formulation.

CHAPTER 4

SYSTEM ANALYSIS

4. SYSTEM ANALYSIS

1. Existing System

As of my last knowledge update in September 2021, there were already several existing systems and approaches for sentiment analysis on social media data like Twitter. Here's an overview of the typical elements of an existing system for sentiment analysis on Twitter during the COVID-19 lockdown:

1. Data Collection: Existing systems typically gather Twitter data using the Twitter API or third-party data providers. These systems may collect tweets based on relevant keywords, hashtags, or geolocation.

2. Preprocessing: Data preprocessing is essential to clean and prepare the collected tweets. This includes tasks like removing duplicates, handling missing data, and tokenizing the text.

2. Proposed System

To enhance sentiment analysis on Twitter during the COVID-19 lockdown in the USA, we propose a comprehensive system that incorporates modern machine learning techniques and data
processing methods. The proposed system aims to provide more accurate sentiment insights and a
better understanding of public opinion during this critical period.

1. Data Collection and Enrichment:

- Collect Twitter data using the Twitter API, focusing on relevant keywords, hashtags, and geolocation to ensure data relevance.
- Enrich the data with additional context, including user information, retweet counts, and timestamp.

2. Preprocessing and Text Cleaning:

- Perform extensive data preprocessing, including removal of duplicates, handling of missing
data, and text cleaning (removing special characters, URLs, and emojis).
- Tokenize and lemmatize the text to prepare it for analysis.

3. Objective of the System

The objective of the system is to provide a comprehensive and up-to-date understanding of public sentiment during the COVID-19 lockdown in the USA, offering valuable insights to policymakers, researchers, and the public for informed decision-making and crisis management.

CHAPTER 5

REQUIREMENT ANALYSIS

5. REQUIREMENT ANALYSIS

Hardware Requirement Specification

- **MySQL**
- **NODE JS**
- **Notepad++ Editor**
- **Processor: Intel core i5 processer**
- **Memory: 15.6 GB**
- **Hard Disk: 40 GB**

Software Requirement Specification

- **Jupyter**
- **Google Collab**
- **VS Code**

CHAPTER 6

DESIGN ANALYSIS

6. DESIGN & ANALYSIS

1. Data Collection: Gather COVID-19 lockdown-related tweets from Twitter using the API, filtering by keywords and geolocation.

2. Data Preprocessing: Clean and prepare the text data by removing special characters, URLs, and stopwords, and then tokenize and lemmatize the text.

3. Feature Engineering: Transform text data into numerical features using TF-IDF and consider adding sentiment lexicons or word embeddings.

4. Machine Learning Models: Select models like Naive Bayes or LSTM for sentiment analysis. Train and validate on labeled data.

5. Sentiment Analysis: Predict sentiment labels (positive, negative, neutral) for tweets, and analyze trends over time and by location.

6. Visualization: Create visualizations (line charts, heatmaps) and geospatial maps to display sentiment variations.

7. Evaluation: Assess model performance using accuracy, precision, recall, and F1- score.

8. Deployment: Develop a user-friendly web dashboard to present real-time sentiment analysis results.

CHAPTER 7

IMPLEMENTATION

7. IMPLEMENTATION

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.
The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and it constraints on implementation, design of methods to achieve the changeover and an evaluation of change over methods a part from planning.
Two major tasks of preparing the implementation are education and training of the users and testing of the system. The more complex the system being implemented, the more involved will be the system analysis and design effort required just for implementation.

The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued.

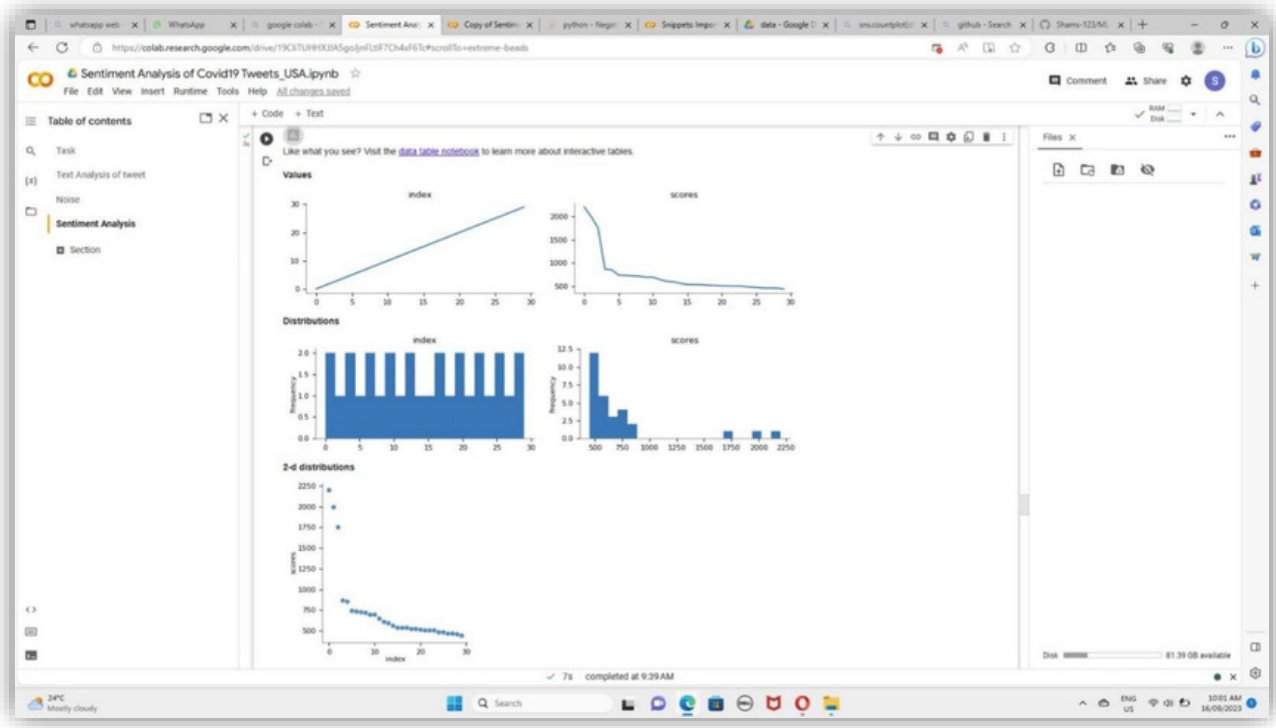
TESTING

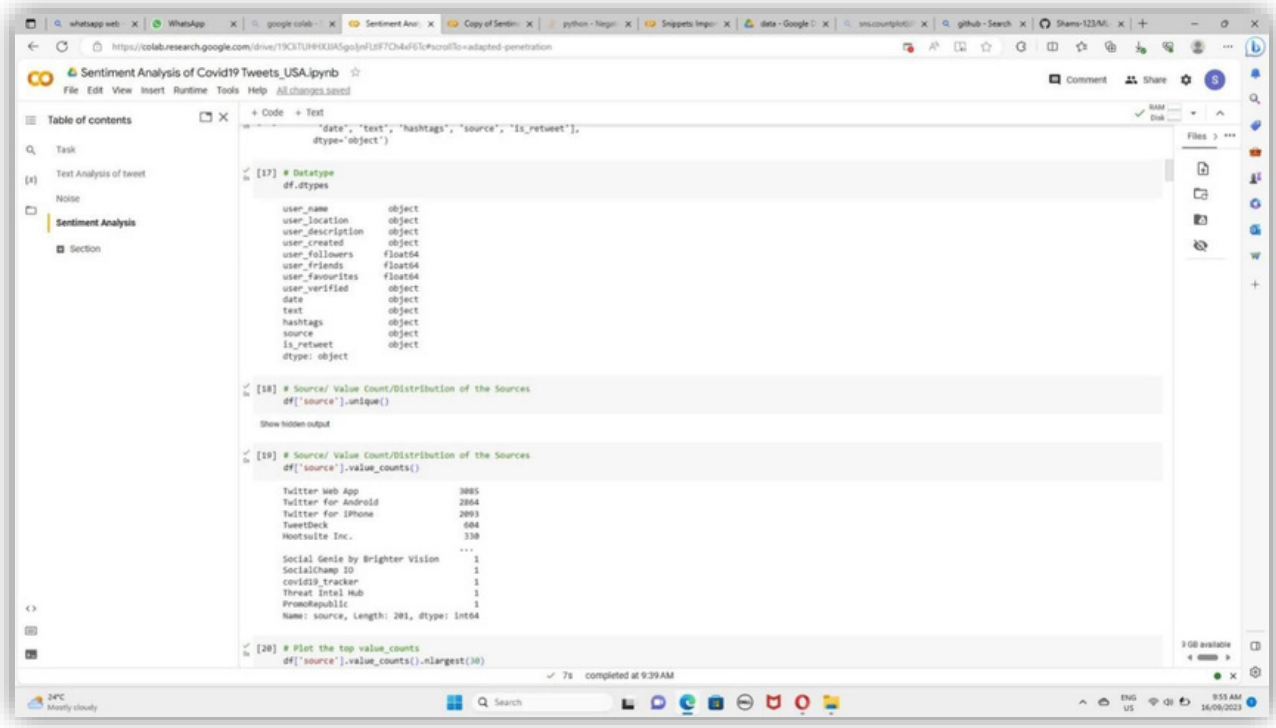
The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements
are satisfied. Software testing is carried out in three steps:
1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether the objectives have been met. Errors are noted down and corrected immediately.
2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.
3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

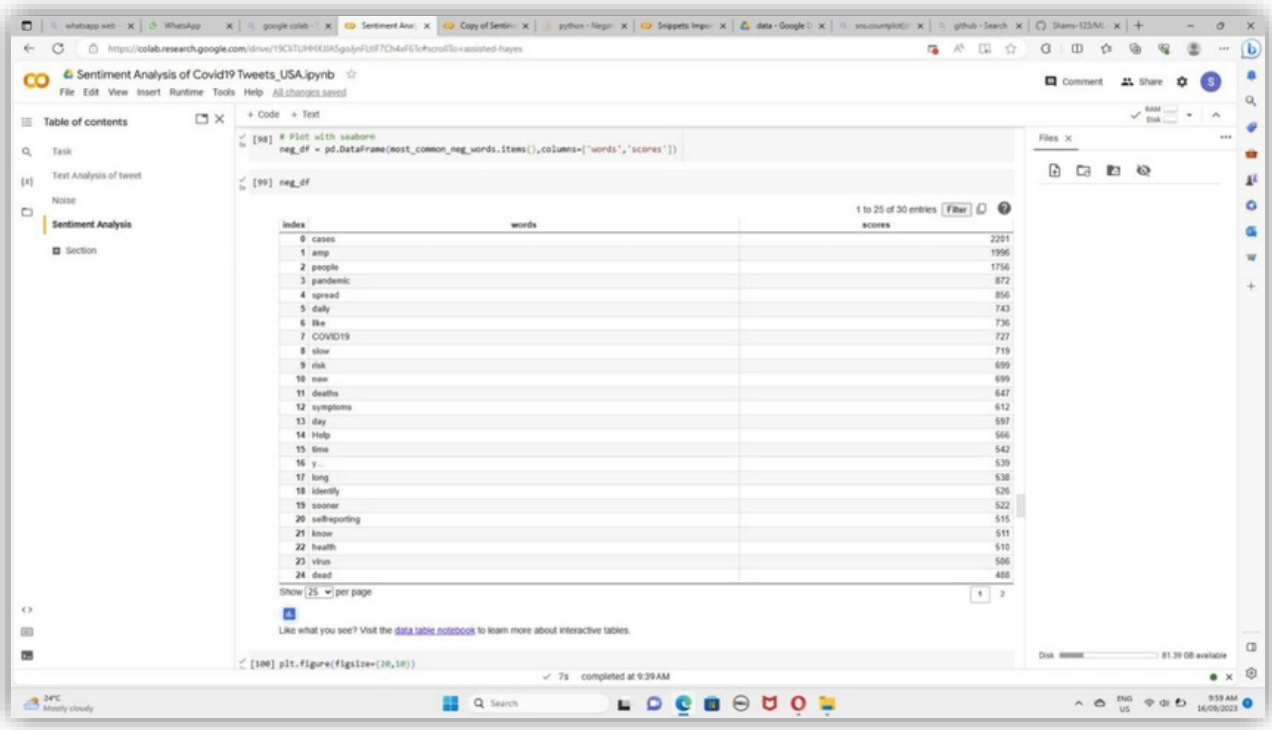
CHAPTER 8

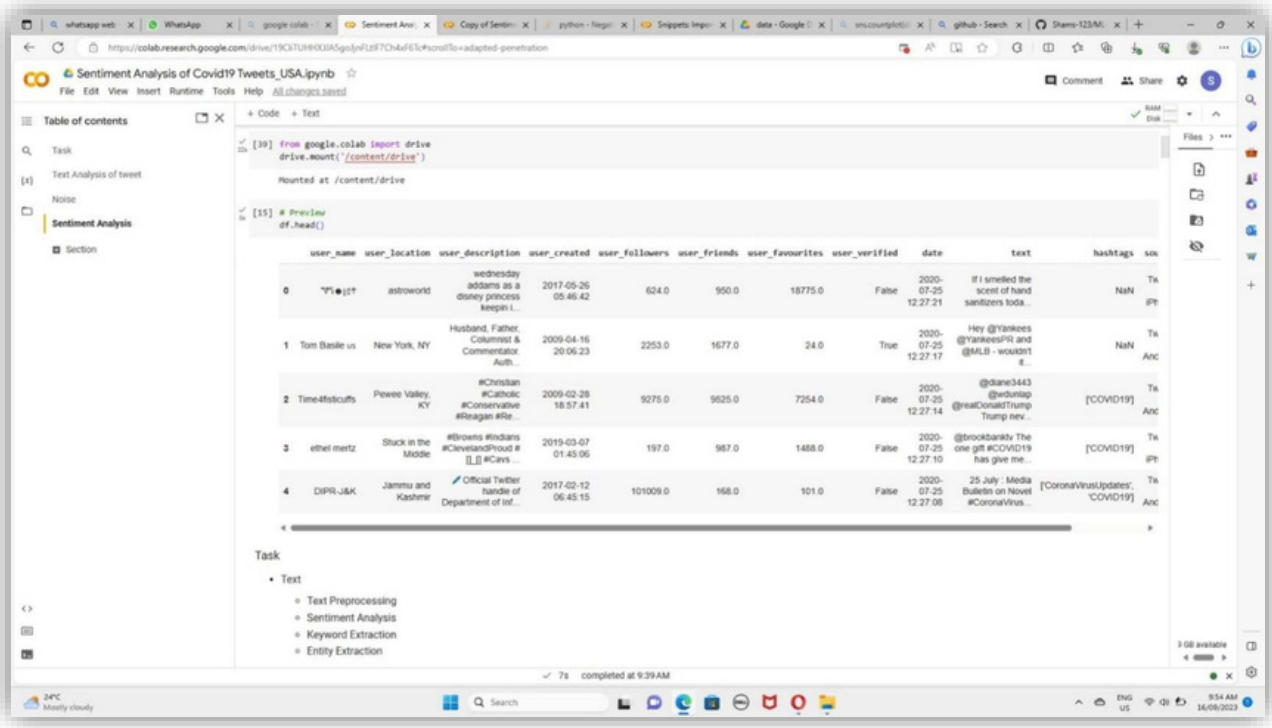
SNAPSHOTS

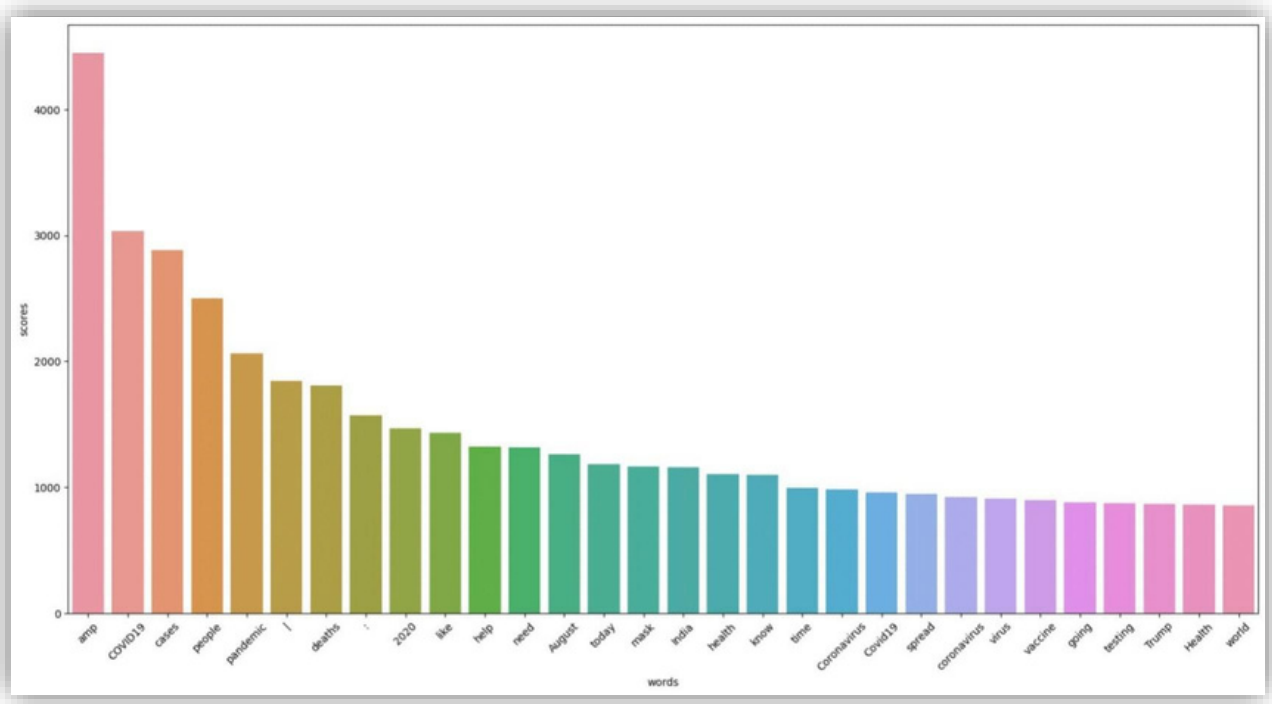
8. SNAPSHOTS

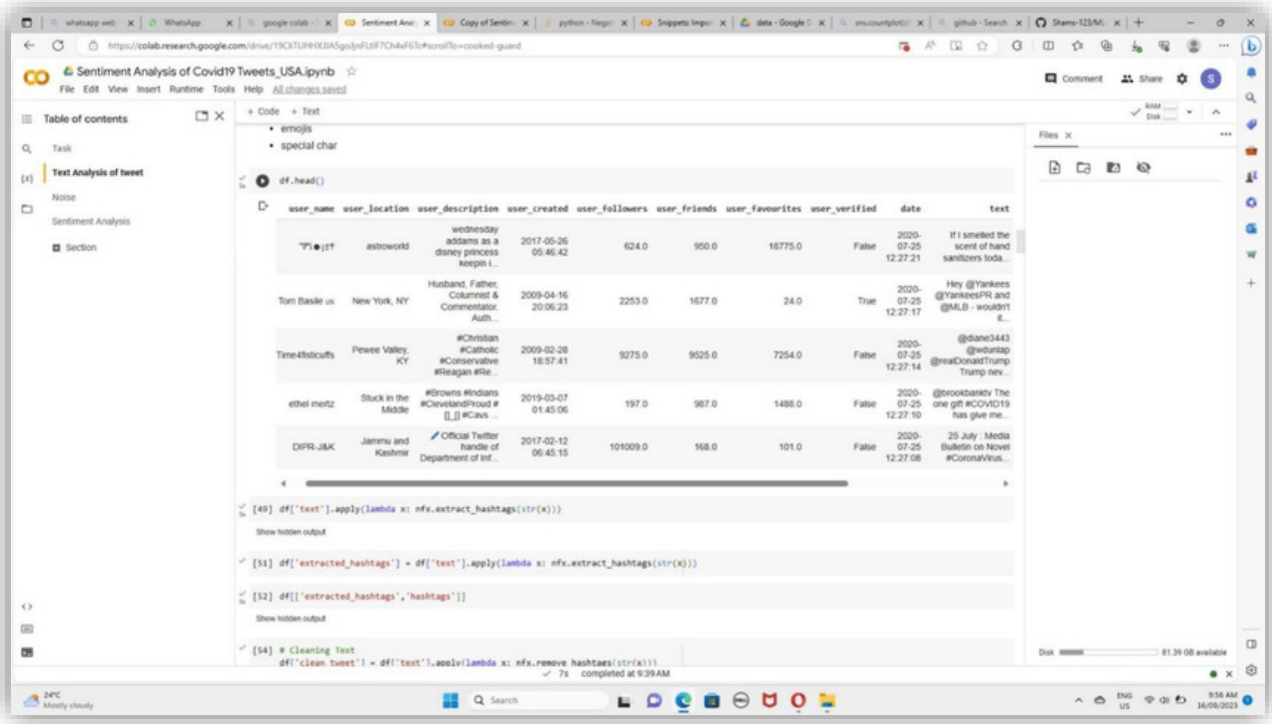

Sentiment Analysis of Covid19 Tweets_USA.ipynb
Table of contents
Like what you see? Visit the data table notebook to learn more about interactive tables.
Values
index
scores
Distributions
2-d distributions


Sentiment Analysis of Covid19 Tweets_USA.ipynb
Table of contents
[17] # Datatype
df.dtypes
[18] # Source/ Value Count/Distribution of the Sources
df['source'].unique()
[19] # Source/ Value Count/Distribution of the Sources
df['source'].value_counts()
[20] # Plot the top value_counts
df['source'].value_counts().nlargest(30)


Sentiment Analysis of Covid19 Tweets_USA.ipynb
Table of contents
Task
Text Analysis of tweet
Noise
Sentiment Analysis
Section
[98] # Plot with seaborn
neg_df = pd.DataFrame(most_common_neg_words.items(),columns=['words','scores'])
[99] neg_df
index words scores
0 cases 2201
1 amp 1996
2 people 1756
3 pandemic 872
4 spread 856
5 daily 743
6 like 736
7 COVID19 727
8 slow 719
9 risk 699
10 new 699
11 deaths 647
12 symptoms 612
13 day 597
14 Help 566
15 time 542
16 y... 539
17 long 538
18 identify 526
19 sooner 522
20 selfreporting 515
21 know 511
22 health 510
23 virus 506
24 dead 488
Like what you see? Visit the data table notebook to learn more about interactive tables.
[100] plt.figure(figsize=(20,10))
completed at 9:39 AM


Sentiment Analysis of Covid19 Tweets_USA.ipynb
Table of contents
Task
Text Analysis of tweet
Noise
Sentiment Analysis
Section
[39] from google.colab import drive
drive.mount('/content/drive')
Mounted at /content/drive
[15] # Preview
df.head()
user_name user_location user_description user_created user_followers user_friends user_favourites user_verified date text hashtags
0 astroworld wednesday addams as a disney princess keepin i... 2017-05-26 05:46:42 624.0 950.0 18775.0 False 2020-07-25 12:27:21 If I smelled the scent of hand sanitizers toda... NaN
1 Tom Basile us New York, NY Husband, Father, Columnist & Commentator. Auth... 2009-04-16 20:06:23 2253.0 1677.0 24.0 True 2020-07-25 12:27:17 Hey @Yankees @YankeesPR and @MLB - wouldn't it... NaN
2 Time4fisticuffs Pewee Valley, KY #Christian #Catholic #Conservative #Reagan #Re... 2009-02-28 18:57:41 9275.0 9625.0 7254.0 False 2020-07-25 12:27:14 @diane3443 @wdunlap @realDonaldTrump Trump nev... [COVID19]
3 ethel mertz Stuck in the Middle #Browns #Indians #ClevelandProud # [] [] #Cavs ... 2019-03-07 01:45:06 197.0 987.0 1488.0 False 2020-07-25 12:27:10 @brookbanktv The one gift #COVID19 has give me... [COVID19]
4 DIPR-J&K Jammu and Kashmir Official Twitter handle of Department of Inf... 2017-02-12 06:45:15 101009.0 168.0 101.0 False 2020-07-25 12:27:08 25 July : Media Bulletin on Novel #CoronaVirus... ['CoronaVirusUpdates', 'COVID19']
Task
Text
Text Preprocessing
Sentiment Analysis
Keyword Extraction
Entity Extraction
completed at 9:39 AM


scores
words
4000
3000
2000
1000
0
amp
COVID19
cases
people
pandemic
|
deaths
:
2020
like
help
need
August
today
mask
india
health
know
time
Coronavirus
Covid19
spread
coronavirus
virus
vaccine
going
testing
Trump
Health
world


Sentiment Analysis of Covid19 Tweets_USA.ipynb
Table of contents
Text Analysis of tweet
Noise
Sentiment Analysis
Section
emojis
special char
df.head()
user_name user_location user_description user_created user_followers user_friends user_favourites user_verified date text
astroworld wednesday addams as a disney princess keepin i... 2017-05-26 05:46:42 624.0 950.0 18775.0 False 2020-07-25 12:27:21 If I smelled the scent of hand sanitizers toda...
Tom Basile us New York, NY Husband, Father, Columnist & Commentator. Auth... 2009-04-16 20:06:23 2253.0 1677.0 24.0 True 2020-07-25 12:27:17 Hey @Yankees @YankeesPR and @MLB - wouldn't it...
Time4fisticuffs Pewee Valley, KY #Christian #Catholic #Conservative #Reagan #Re... 2009-02-28 18:57:41 9275.0 9525.0 7254.0 False 2020-07-25 12:27:14 @diane3443 @wdunlap @realDonaldTrump Trump nev...
ethel mertz Stuck in the Middle #Browns #Indians #ClevelandProud # [] [] #Cavs ... 2019-03-07 01:45:06 197.0 987.0 1488.0 False 2020-07-25 12:27:10 @brookbanktv The one gift #COVID19 has give me...
DIPR-J&K Jammu and Kashmir Official Twitter handle of Department of Inf... 2017-02-12 06:45:15 101009.0 168.0 101.0 False 2020-07-25 12:27:08 25 July : Media Bulletin on Novel #CoronaVirus...
[49] df['text'].apply(lambda x: nfx.extract_hashtags(str(x)))
Show hidden output
[51] df['extracted_hashtags'] = df['text'].apply(lambda x: nfx.extract_hashtags(str(x)))
[52] df[['extracted_hashtags','hashtags']]
Show hidden output
[54] # Cleaning Text
df['clean_tweet'] = df['text'].apply(lambda x: nfx.remove_hashtags(str(x)))
7s completed at 9:39 AM
Files

CHAPTER 9

CONCLUSION

9. CONCLUSION

The package was designed in such a way that future modifications can be done easily. The following conclusions can be deduced from the development of the project:

- Automation of the entire system improves the efficiency
- It provides a friendly graphical user interface which proves to be better when compared to the existing system.
- It gives appropriate access to the authorized users depending on their permissions.
- It effectively overcomes the delay in communications.
- Updating of information becomes so easier
- System security, data security and reliability are the striking features.
- The System has adequate scope for modification in future if it is necessary.

10. REFERENCE

Machine Learning - https://youtu.be/Y4qO9unerGs?si=e5HphOd4KBCv1g3H