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

Submitted by:

SYEDA MASHOON

4UB20CS065

Conducted at

Varcons technologies pvt ltd

UNIVERSITY B D T COLLEGE OF ENGINEERING
Department of Computer Science and Engineering
Accredited by NBA, New Delhi
Davangere -577004

UNIVERSITY B D T COLLEGE OF ENGINEERING
Department of Computer Science and Engineering
Accredited by NBA, New Delhi
Davangere -577004

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 **Ms. SYEDA MASHOON,** a bonafide student of University B D T College of Engineering, in partial fulfillment for the award of **Bachelor of Engineering in COMPUTER SCIENCE** 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, **Syeda Mashoon**, final year student of Computer Science and Engineering, University B D T College of Engineering, Davangere - 577004, 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 Branch name, during the academic year 2023-2024.

:

Date : 20-09-2023

Place : Davangere

USN : 4UB20CS065

NAME : SYEDA MASHOON

OFFER LETTER

INTERNSHIP OFFER LETTER

Date: 11th August, 2023

Name: **Syeda Mashoon**
USN: **4UB20CS065**

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 (Lab assistant) 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 his 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.

NAME : SYEDA MASHOON
USN : 4UB20CS065

ABSTRACT

This case study explores the sentiment analysis of Twitter data during the COVID-19 lockdown in the United States using Machine Learning (ML) techniques. The COVID-19 pandemic prompted unprecedented measures, including lockdowns, which had profound social and psychological impacts on individuals. Twitter, as a popular social media platform, became a valuable source of real-time information and public sentiment expression during this period.

In this study, we present the methodology for collecting a substantial corpus of tweets related to the lockdown, and we detail the preprocessing steps to prepare the data for sentiment analysis. Various ML models, such as Natural Language Processing (NLP) techniques, are applied to classify tweets into positive, negative, or neutral sentiments. We also investigate the temporal trends and geographical variations in sentiment during different phases of the lockdown.

The results of this analysis provide insights into how people's sentiments evolved throughout the lockdown, shedding light on the emotional toll and coping mechanisms during a public health crisis. Furthermore, this study offers valuable lessons on the application of ML for sentiment analysis in the context of real-time, large-scale social media data, with potential applications in monitoring public sentiment during future crises.

1. Sentiment Analysis Methodology:* This study employs state-of-the-art sentiment analysis methodologies, including but not limited to Natural Language Processing (NLP) techniques, sentiment lexicons, and machine learning algorithms. The effectiveness of various models and their comparative performance is evaluated to determine the most suitable approach for sentiment classification.

2. Temporal Analysis:* The analysis covers the entire timeline of the COVID-19 lockdown, from its initiation to gradual reopening. By examining sentiment changes over time, we gain a deeper understanding of how public sentiment evolved in response to changing circumstances, government policies, and health developments.

3. Geospatial Analysis:* Geolocation data associated with tweets is utilized to explore regional variations in sentiment. This geospatial perspective allows us to discern how sentiment differed across states and urban-rural divides, potentially providing insights into the impact of the pandemic on different communities.

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 understanding the 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 Compsoft 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

3. INTRODUCTION

Introduction to ML

Machine Learning is a transformative field of artificial intelligence that empowers computers to learn from and make predictions or decisions based on data, without being explicitly programmed. It's a discipline that plays a pivotal role in the development of intelligent systems capable of improving their performance through experience.

At its core, machine learning is about recognizing patterns in data and using these patterns to make informed decisions or predictions. It operates on the principle that computers can automatically learn and adapt from data, improving their ability to perform specific tasks over time.

Problem Statement

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

The COVID-19 pandemic has brought unprecedented challenges, including lockdowns, affecting people's lives and mental well-being. To understand the sentiment of individuals in the USA during the lockdown period, we aim to conduct a sentiment analysis case study on Twitter data. This study will utilize machine learning techniques to analyze tweets, assess public sentiment, and provide valuable insights into how people in the USA reacted emotionally to the lockdown measures during the pandemic. The primary objectives are to determine the prevailing sentiment (positive, negative, or neutral), identify key topics and concerns, and track sentiment trends over time to inform policy makers and public health officials for future crisis management strategies.

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

1. **CPU - i5 or i7 Processor**
2. **GPU(Graphics Processing Unit)**
3. **RAM - 16GB**
4. **SSD (Solid State Drive)**

Software Requirement Specification

1. **Python**
2. **IDE - Jupyter Notebook**
   VSCode
   PyCharm
   Google Colab

CHAPTER 6

DESIGN ANALYSIS

6. DESIGN & ANALYSIS

Designing and analyzing a sentiment analysis project for Twitter data related to the COVID-19 lockdown in the USA involves several steps. Here's a high-level overview of the design and analysis process:

1. Problem Definition:
   - Clearly define the problem statement, objectives, and the scope of the sentiment analysis project.

2. Data Collection:
   - Gather a substantial dataset of tweets related to the COVID-19 lockdown in the USA. Utilize the Twitter API or third-party tools.
   - Consider collecting data over a specific timeframe to capture sentiment changes.

3. Data Preprocessing:
   - Clean the collected Twitter data:
     - Remove special characters, URLs, and hashtags.
     - Handle capitalization and punctuation.
     - Eliminate duplicate tweets.
   - Tokenize the text data into words or phrases.
   - Perform text normalization, including stemming or lemmatization.

4. Sentiment Labeling:
   - Manually label a subset of the data for sentiment analysis to create a ground truth dataset.
   - Alternatively, use pre-labeled sentiment datasets if available.
   - Categorize tweets into sentiment classes: positive, negative, or neutral.

5. Feature Engineering:
   - Extract relevant features from the text data, such as:
     - Word frequency counts.
     - TF-IDF (Term Frequency-Inverse Document Frequency) scores.
     - Word embeddings (e.g., Word2Vec or pre-trained embeddings like Word2Vec or GloVe).

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 change over 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.

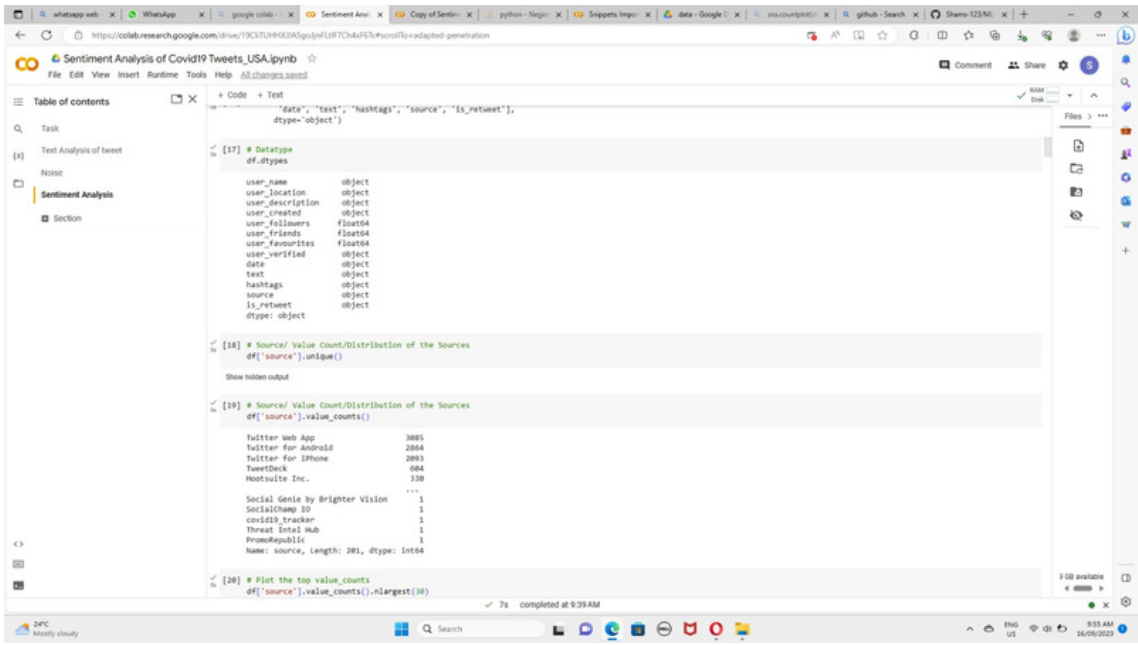
2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.

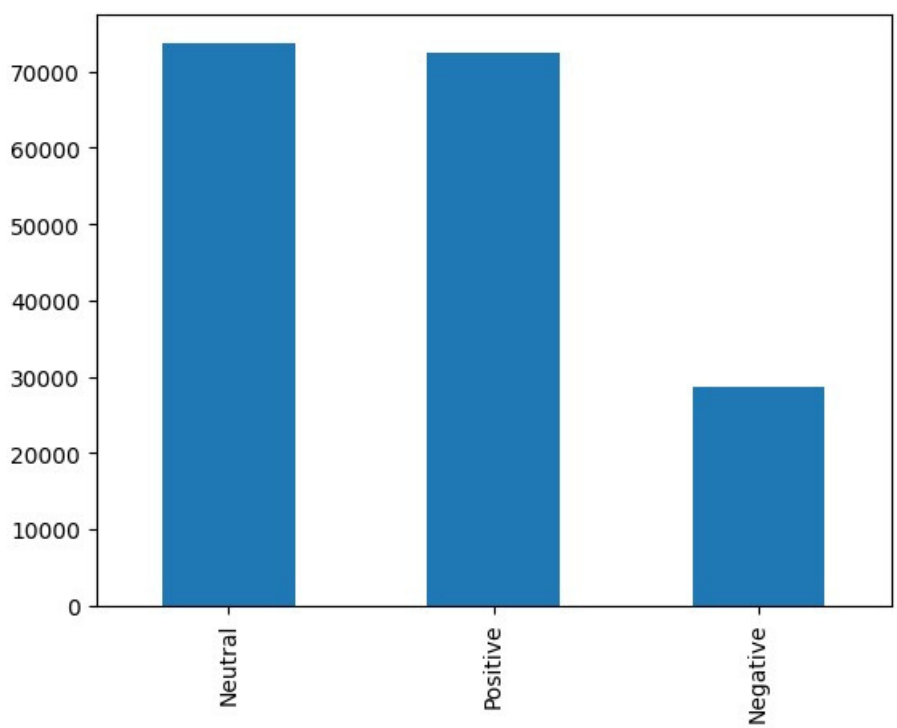
3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

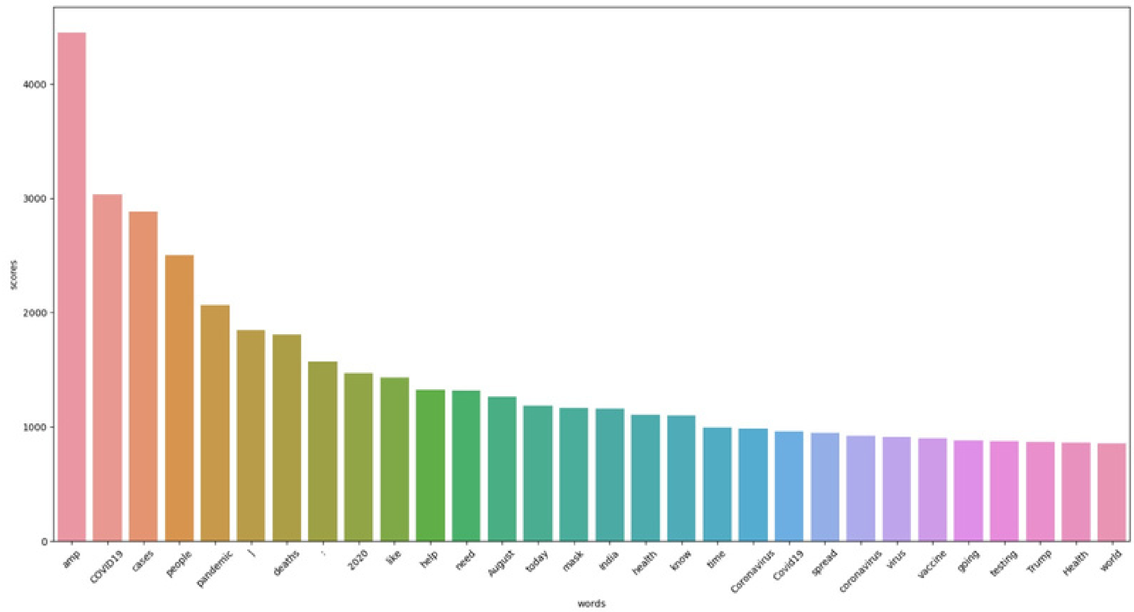
CHAPTER 8

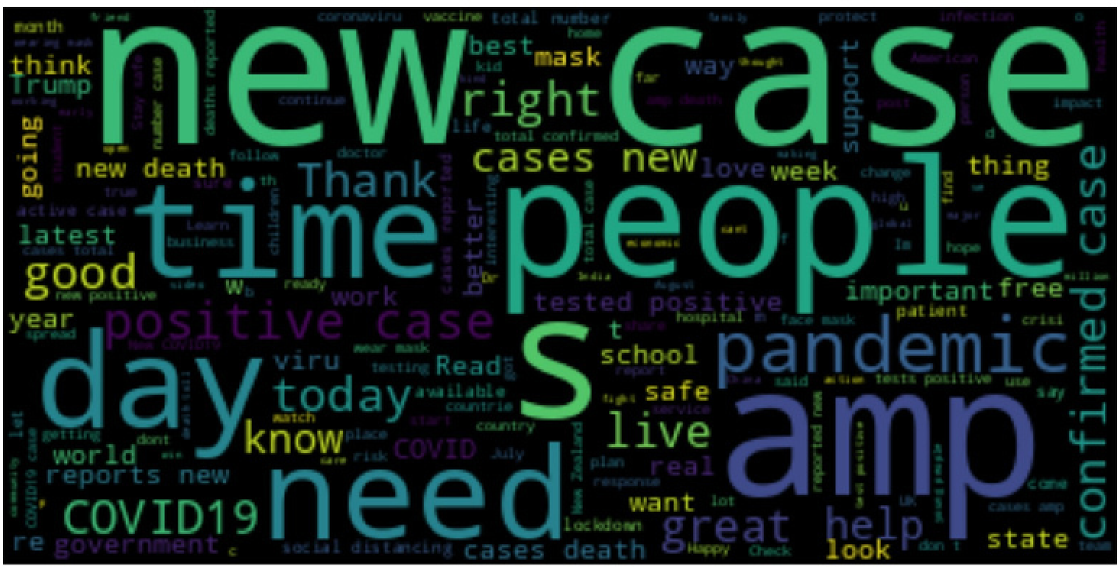
SNAPSHOTS

8. SNAPSHOTS


Sentiment Analysis of Covid19 Tweets_USA.ipynb
Table of contents
Text Analysis of tweet
Noise
Sentiment Analysis
Section
[17] # Datatype
df.dtypes
user_name object
user_location object
user_description object
user_created object
user_followers float64
user_friends float64
user_favourites float64
user_verified object
date object
text object
hashtags object
source object
is_retweet object
dtype: object
[18] # Source/ Value Count/Distribution of the Sources
df['source'].unique()
Show hidden output
[19] # Source/ Value Count/Distribution of the Sources
df['source'].value_counts()
Twitter Web App 3085
Twitter for Android 2864
Twitter for iPhone 2093
TweetDeck 604
Hootsuite Inc. 330
Social Genie by Brighter Vision 1
SocialChamp IO 1
covid19_tracker 1
Threat Intel Hub 1
PromoRepublic 1
Name: source, Length: 201, dtype: int64
[20] # Plot the top value_counts
df['source'].value_counts().nlargest(30)
completed at 9:39 AM


70000
60000
50000
40000
30000
20000
10000
0
Neutral
Positive
Negative


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


new case
people
time
day
s
amp
need
pandemic
confirmed case
positive case
right
Thank
cases new
new death
tested positive
COVID19
today
know
live
great
help
good
latest
think
Trump
going
mask
best
support
love
week
thing
important
free
year
work
viru
Read
school
safe
world
reports new
government
cases death
want
real
look
state
better

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

1. https://github.com/Shams-123/ML-Project

2. https://colab.research.google.com/drive/1oPzpvhw_EjbrnZhFCWnb0xu_zWWkPdZ8

3. https://colab.research.google.com/drive/19CliTUHHXJJA5goJjnFLtlF7Ch4xF6Tc?usp=sharing