VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

"Twitter Sentiment Analysis of COVID -19 Lockdown in USA Using Machine Learning"

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

BACHELOR OF ENGINEERING IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Submitted by:
SRISHAN S

1JT20AI044

Conducted at
VARCONS TECHNOLOGIES PVT. LTD.

JYOTHY INSTITUTE OF TECHNOLOGY

Department of Computer Science & Engineering

Tataguni, off Kanakapura Road, Bengaluru – 560 082

JYOTHY INSTITUTE OF TECHNOLOGY
Department of Computer Science & Engineering
Accredited by NBA, New Delhi
Tataguni, off Kanakapura Road, Bengaluru – 560 082

CERTIFICATE

This is to certify that the Internship titled **"Twitter Sentiment Analysis of COVID -19 Lockdown in USA Using Machine Learning"** carried out by **Mr. Srishan S.,** a bonafide student of Jyothy Institute of Technology, in partial fulfillment for the award of **Bachelor of Engineering**, in **AIML** under Visvesvaraya Technological University,Belagavi, during the year 2022-2023. 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, **Srishan S**, final year student of Computer Science & Engineering, Jyothy Institute of Technology - 560 082, declare that the Internship has been successfully completed, in **VARCONS TECHNOLOGIES PVT. LTD**. This report is submitted in partial fulfillment of the requirements for the award of Bachelor Degree in Computer Science & Engineering, during the academic year 2023-2024.

Date: September 20th, 2023   :

Place: Bengaluru

USN: 1JT20AI044

NAME: Srishan S

OFFER LETTER

INTERNSHIP OFFER LETTER

Date: 11th August, 2023

Name: **Srishan S**
USN: **1JT20AI044**

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 usadequate facilities to undertake this Internship.

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

We would like to thank our (Lab assistant name) Software Services for guiding us during the period of internship.

We express our deep and profound gratitude to our guide, Guide name, 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 the Internship.

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

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

SRISHAN S
1JT20AI044

ABSTRACT

This project presents a comprehensive approach to Twitter sentiment analysis utilizing Machine Learning (ML) techniques and Natural Language Processing (NLP) in Python. The goal is to automatically classify tweets as positive, negative, or neutral based on the emotions expressed in the text. Preprocessing steps such as tokenization, stemming, and stop-word removal are employed to clean and prepare the data.

A diverse dataset of tweets is collected and labeled, and ML algorithms are implemented for sentiment classification. Feature engineering techniques are explored to extract meaningful information from the text. Model performance is evaluated using metrics like accuracy, precision, recall, and F1-score.

The results demonstrate the effectiveness of the ML-based sentiment analysis model in gauging public sentiment on Twitter, making it a valuable tool for businesses, policymakers, and researchers seeking to understand and respond to public opinions and emotions expressed on this influential social media platform.

Table of Contents

CHAPTER 1

COMPANY PROFILE

1. <u>COMPANY PROFILE</u>

A Brief History of Varcons Technologies

Varcons Technologies, 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 the 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.

Sarva Moola 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, Sarva Moola Software Services. specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion, and tailor-made software products, designing solutions best-suiting clients' requirements.

Varcons Technologies strives to be the front runner in creativity and innovation in software development through its 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 its clients and help them to define their exact solution requirement. Sometimes even they wonder if 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, and 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. <u>ABOUT THE COMPANY</u>

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, Varcons Technologies specialize in ERP, Connectivity, SEO Services, Conference Management, effective webpromotion and tailor-made software products, designing solutions best suiting clients' requirements. The organization where they have a right mix of professionals as a stakeholder 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 technologicalinnovations in the field of Mobile technologies, Web applications as well and Business and Enterprise solutions. 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 clients which will bring a cascading 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

3. <u>INTRODUCTION</u>

Introduction to ML

Machine Learning (ML) is that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do. In simple words, ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method. The main focus of ML is to allow computer systems learn from experience without being explicitly programmed or human intervention.

Human beings, at this moment, are the most intelligent and advanced species on earth because they can think, evaluate and solve complex problems. On the other side, AI is still in its initial stage and haven't surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, "to make decisions, based on data, with efficiency and scale". Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning, and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data-driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programming logic, in problems that cannot be programmed inherently. The fact is that we can't do without human intelligence, but other aspect is that we all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises.

Problem Statement

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

SYSTEM ANALYSIS

4. <u>SYSTEM ANALYSIS</u>

1. Existing System

During the COVID-19 pandemic, lockdown measures were implemented in the USA to mitigate the spread of the virus. On social media platforms like Twitter, people shared their thoughts, emotions, and experiences related to the lockdown. Analyzing this vast amount of user-generated content manually is impractical. Some key points regarding the existing system are:

Data Abundance: Twitter contains a wealth of data related to the lockdown, but it requires an automated system to extract valuable insights from it.

Manual Limitations: Human-based sentiment analysis of this data is slow, subjective, and not scalable.

Need for Automation: There is a need for an automated system that can efficiently analyze sentiment in real-time or on-demand based on user-specified keywords or hashtags.

Open-Source Datasets: There are open-source datasets available that can be leveraged for sentiment analysis.

2. Proposed System

The proposed solution is to build a Python application that interacts with users by requesting a keyword or hashtag related to the lockdown in the USA during the COVID-19 pandemic. The application will then perform sentiment analysis on Twitter data containing this keyword. Key points for the proposed system are:

Data Collection: Collect Twitter data containing the specified keyword using Twitter APIs o pre-existing open-source datasets.

Data Preprocessing: Clean and preprocess the text data, including tokenization, stopword removal, punctuation removal, and possibly stemming or lemmatization.

Sentiment Analysis Model: Utilize machine learning models for sentiment analysis to classify tweets into positive, negative, or neutral categories.

User Interface: Design a user-friendly interface that allows users to input keywords and view sentiment analysis results.

Real-Time or Batch Processing: The system should be capable of handling both real-time Twitter data and batch processing for historical data analysis.

Scalability: Ensure that the system can handle a growing volume of data and user requests efficiently.

Accuracy: Strive for high accuracy in sentiment classification.

3. Objective of the System

Automated Sentiment Analysis: Develop an automated system capable of accurately analyzing sentiment related to the lockdown in the USA during COVID-19 based on user-specified keywords.

User-Friendly Interface: Create an intuitive and user-friendly interface that allows users to easily input keywords and access sentiment analysis results.

Real-Time and Batch Processing: Enable the system to handle both real-time Twitter data and batch processing for historical analysis.

High Accuracy: Achieve high accuracy in sentiment classification to provide valuable insights.

Scalability: Design the system to scale gracefully to accommodate growing data volumes and user demand.

Ethical and Legal Compliance: Ensure that the system complies with ethical guidelines and legal requirements, respecting user privacy and platform policies.

Feedback and Improvement: Collect user feedback and continuously improve the application's accuracy, usability, and features based on this feedback.

Documentation and Usability: Provide comprehensive documentation to make the application accessible and easy to understand for users and developers.

By achieving these objectives, the system aims to provide a valuable tool for analyzing public sentiment during critical events like the COVID-19 pandemic, allowing for data-driven insights and decision-making.

CHAPTER 5

REQUIREMENT ANALYSIS

5. <u>REQUIREMENT ANALYSIS</u>

Hardware Requirement Specification

```
+-------------------------------------------------------------------+
|                      Hardware Requirements                        |
+-------------------------------------------------------------------+
| CPU:                 Modern multi-core processor                  |
| RAM:                  8GB+ recommended                            |
| GPU:                 Optional for deep learning tasks             |
| Storage:              Adequate space, SSD recommended             |
| Network:              High-speed internet connection              |
| Deployment:          Cloud or web hosting (optional)              |
| Backup:              Implement backup and redundancy              |
| Security:            Ensure server security                       |
| Cooling/PSU:         Adequate cooling and power supply            |
| Scalability:         Consider future scalability                  |
| Budget:              Align hardware choices with budget           |
+-------------------------------------------------------------------+
```
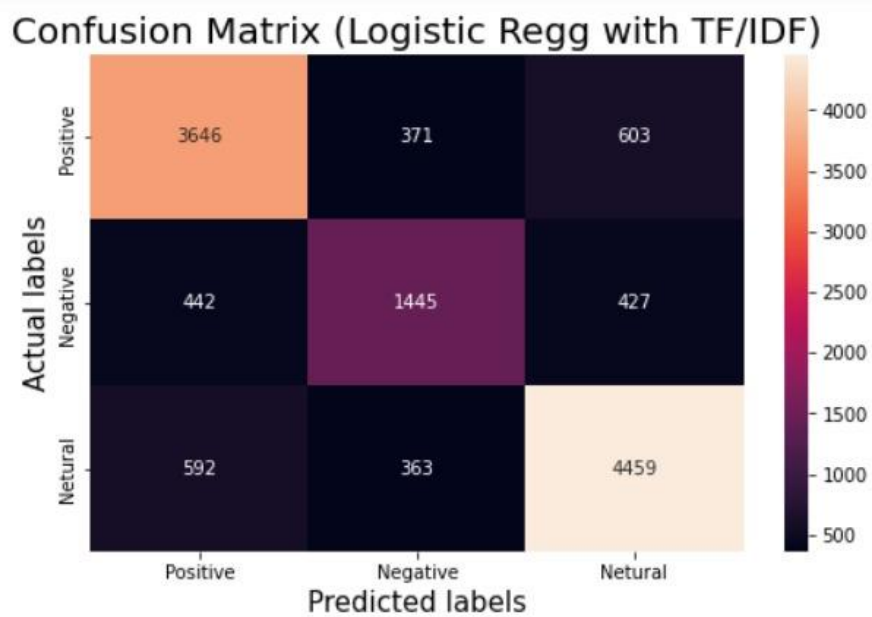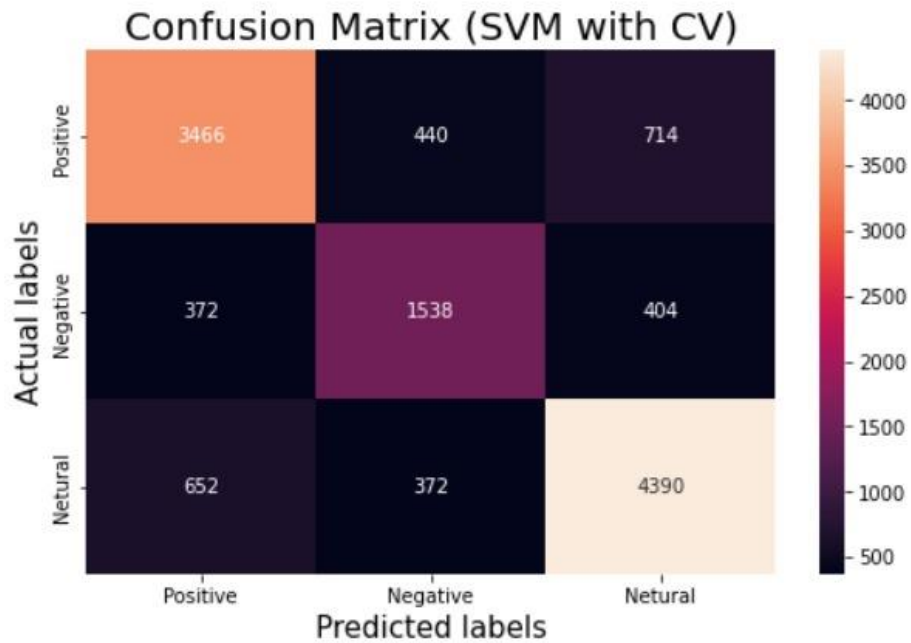
Software Requirement Specification

```
+-------------------------------------------------------------------+
|                      Software Requirements                        |
+-------------------------------------------------------------------+
| OS:                     Linux (Preferred), Windows, macOS         |
| ML:                     scikit-learn or  TensorFlow               |
| Web Framework (Opt.):   Flask or Django                           |
| Database (Opt.):        MySQL, PostgreSQL, or NoSQL DB            |
| VCS:                    Git and GitHub                            |
| IDE:                    Visual Studio Code or jypter notebook     |
| Deployment (Opt.):      Cloud (AWS, GCP Azure) or Web Host        |
| Env. Mgmt:              Conda                                      |
+-------------------------------------------------------------------+
```
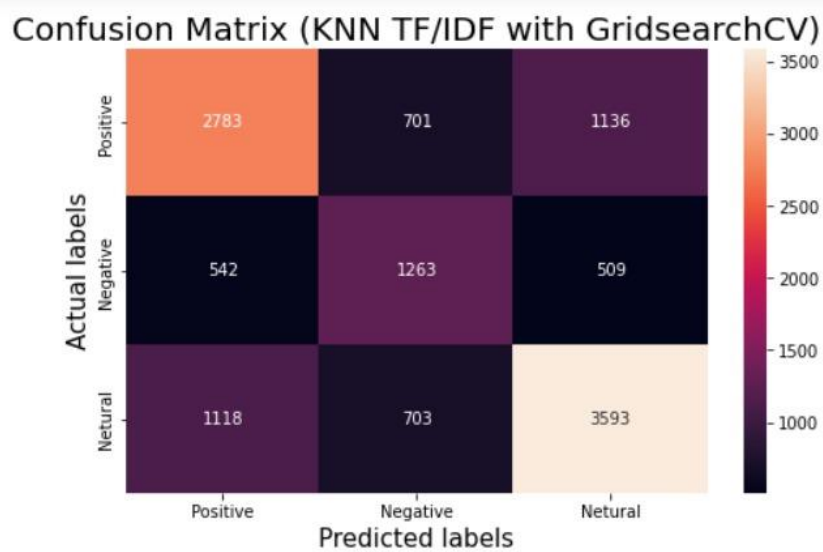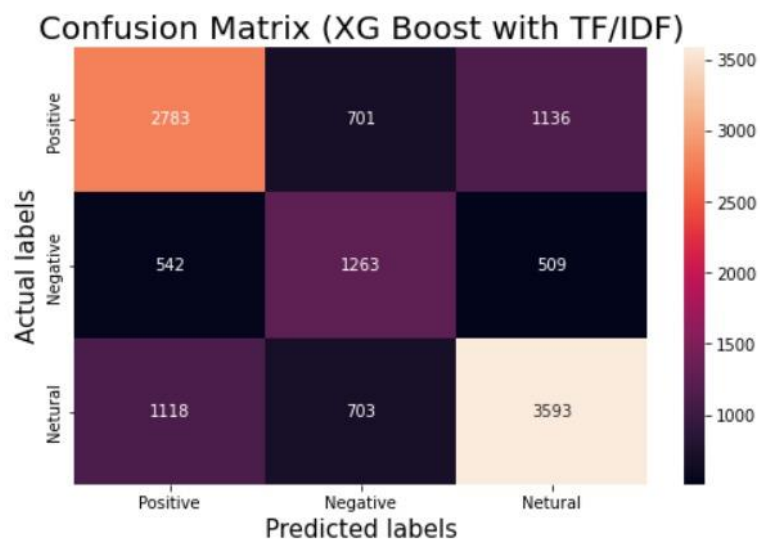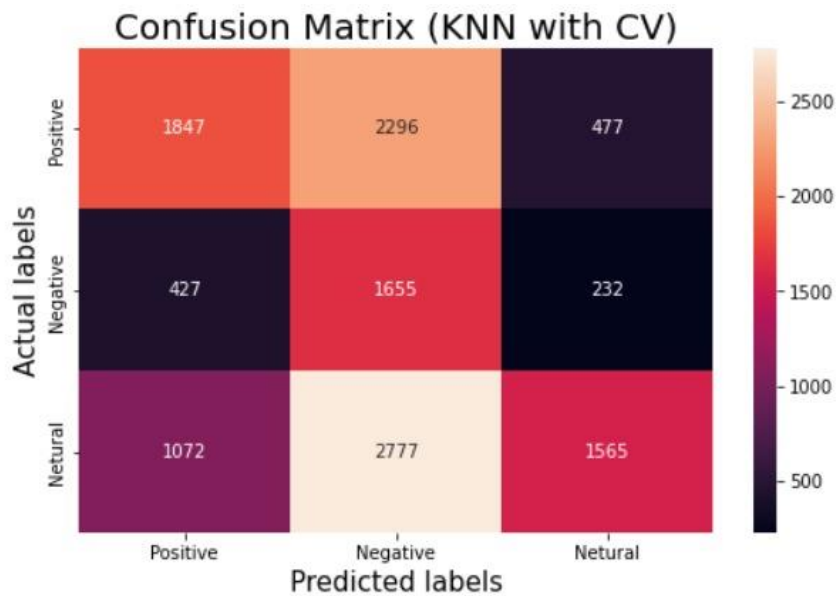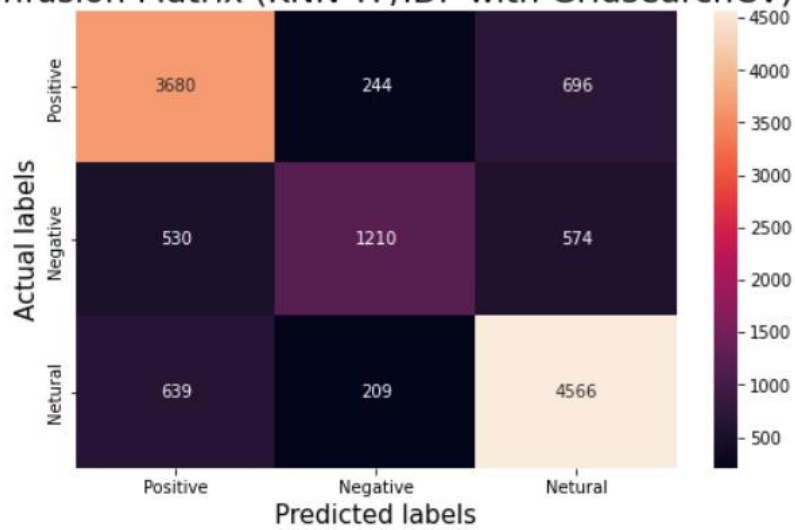
CHAPTER 6

DESIGN ANALYSIS

6. <u>DESIGN & ANALYSIS</u>

Confusion Matrix (SVM with CV)

Confusion Matrix (Logistic Regg with TF/IDF)

Confusion Matrix (KNN with CV)

Confusion Matrix (XG Boost with TF/IDF)

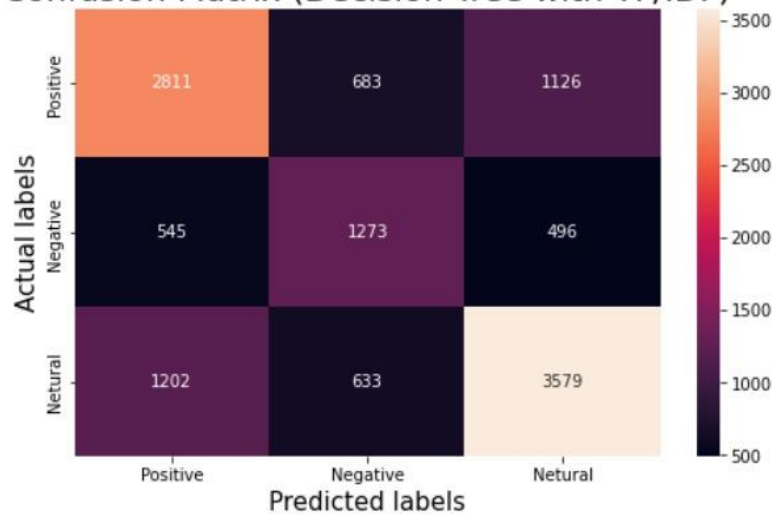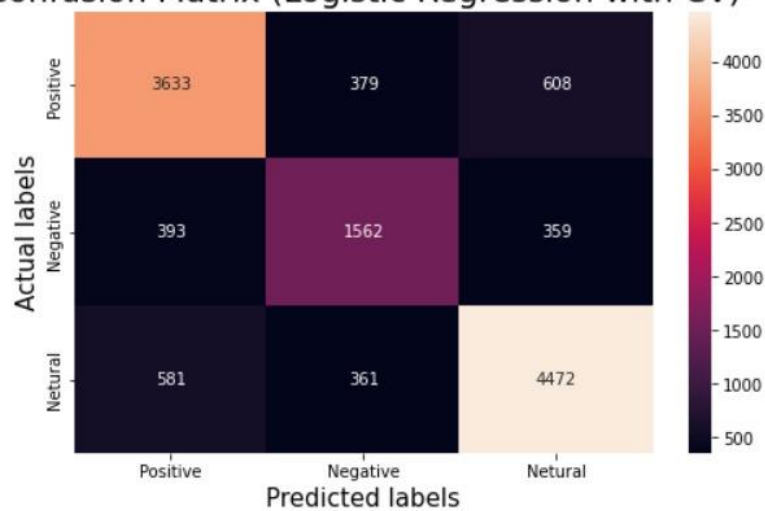Confusion Matrix (KNN TF/IDF with GridsearchCV)

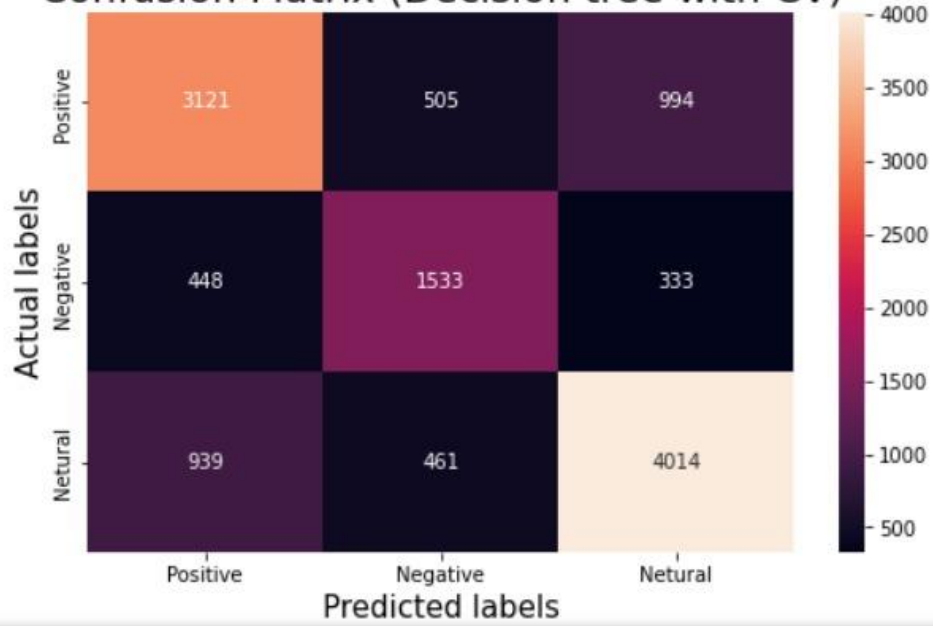## Confusion Matrix (KNN TF/IDF with GridsearchCV)

Confusion Matrix (Decision Tree with TF/IDF)

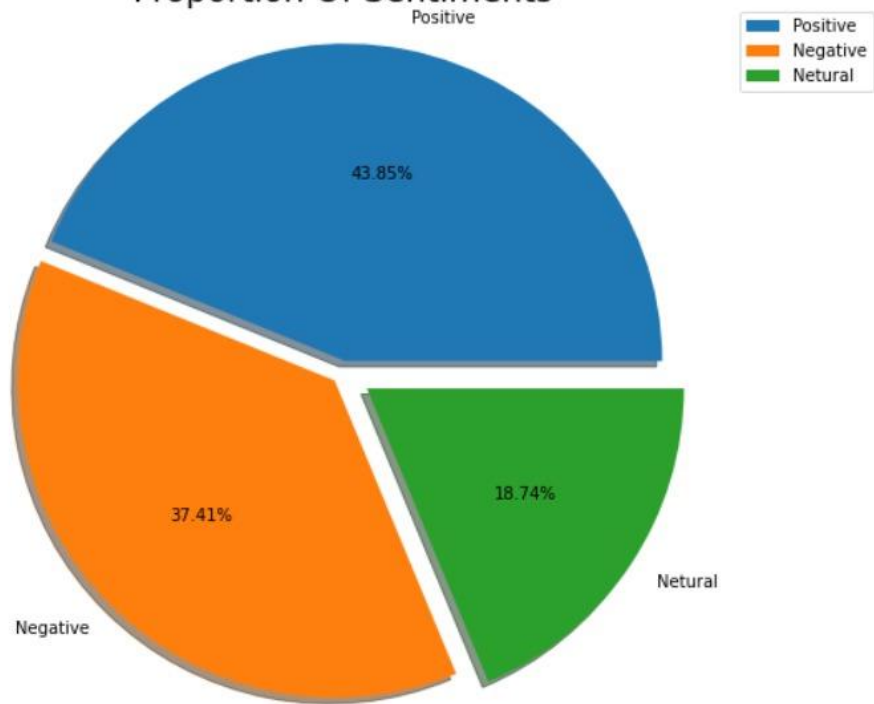## Confusion Matrix (Logistic Regression with CV)

Confusion Matrix (Decision tree with CV)

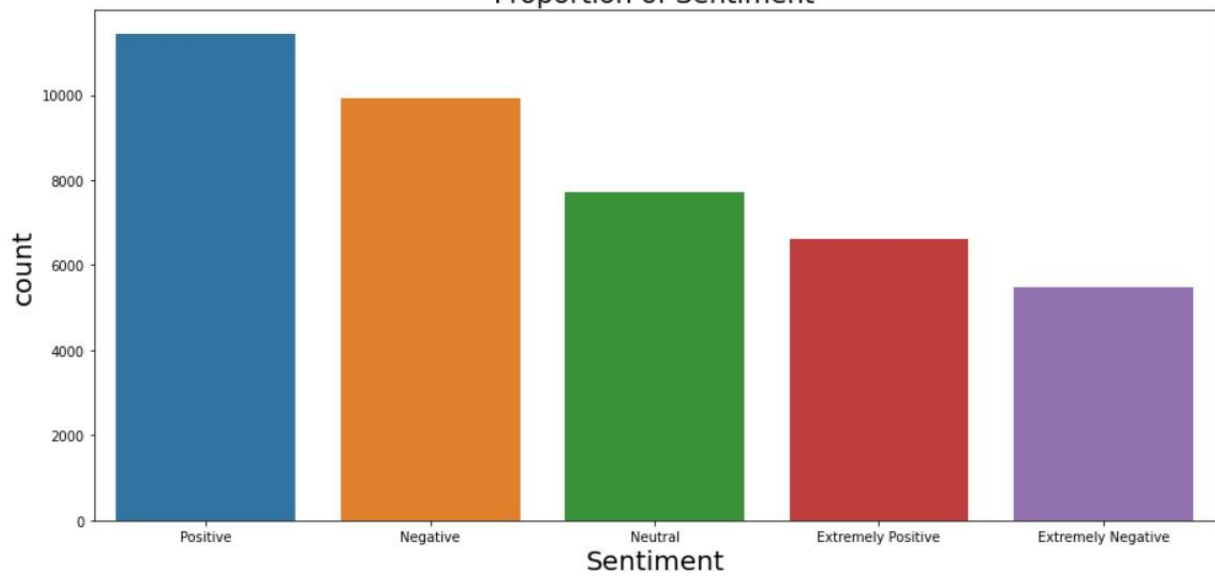|	Positive	Negative	Netural
Positive	3121	505	994
Negative	448	1533	333
Netural	939	461	4014

Actual labels / Predicted labels

Confusion Matrix (XG-Boost with CV)

	Positive	Negative	Netural
Positive	2769	1	1850
Negative	411	3	1900
Netural	748	0	4666

Actual labels / Predicted labels
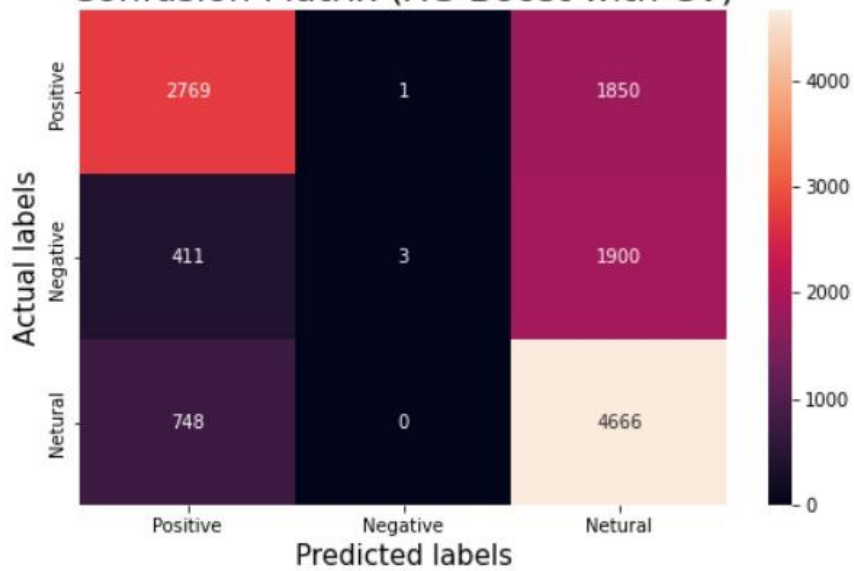
Proportion Of Sentiments

Proportion of Sentiment

CHAPTER 7

IMPLEMENTATION

7. <u>IMPLEMENTATION</u>

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.

Train-Test Split

```
In [43]: #Assigning dependent and independent features
         X= df['lemmed']
         y=df['Sentiment']
```

```
In [44]: # Applying Train test split
         X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.3,stratify=y,random_state=10)
```

```
In [45]: #checking shape of splitted data
         print(X_train.shape)
         y_test.shape
```

```
         (28809,)
```

```
Out[45]: (12348,)
```

```
In [46]: #checking splitted data
         print(X_train.head())
         y_train.head()
```

```
         23805    [target, pull, back, store, remodels, opening,...
         31828    [breaking, msf, call, patent, profiteering, dr...
         24504    [coronavirus, spread, world, leader, get, toug...
         17281    [store, shelf, empty, people, selling, certain...
         11488    [fresco, u, normal, night, final, reduction, n...
         Name: lemmed, dtype: object
```

```
Out[46]: 23805     Neutral
         31828    Negative
         24504    Negative
         17281    Positive
         11488     Neutral
         Name: Sentiment, dtype: object
```

Implementing Models (Bag of Words) by Count Vectoriser

1. Logistic Regression Count Vectoriser Method with Gridsearch CV

```
In [52]: #Initalizing the model
         lr_cv = LogisticRegression()
         parameters = dict(penalty=['l1', 'l2'],C=[100, 10, 1.0, 0.1, 0.01])

         #Hyperparameter tuning by GridserchCV
         logreg_Gcv=GridSearchCV(lr_cv,parameters,cv=15)

         #fitting the data to model
         logreg_Gcv.fit(cv_X_train,y_train)
```

```
Out[52]: GridSearchCV(cv=15, estimator=LogisticRegression(),
                      param_grid={'C': [100, 10, 1.0, 0.1, 0.01],
                                  'penalty': ['l1', 'l2']})
```

```
In [53]: #predicted values
         pred_lr_cv = logreg_Gcv.predict(cv_X_test)
```

```
In [54]: pred_lr_cv
```

```
Out[54]: array(['Positive', 'Positive', 'Neutral', ..., 'Neutral', 'Negative',
                'Positive'], dtype=object)
```

```
In [55]: #Accuracy
         accuracy_lr_cv = accuracy_score(y_test,pred_lr_cv)
         print("Accuracy :",(accuracy_lr_cv))
```

```
Accuracy : 0.7828798185941043
```
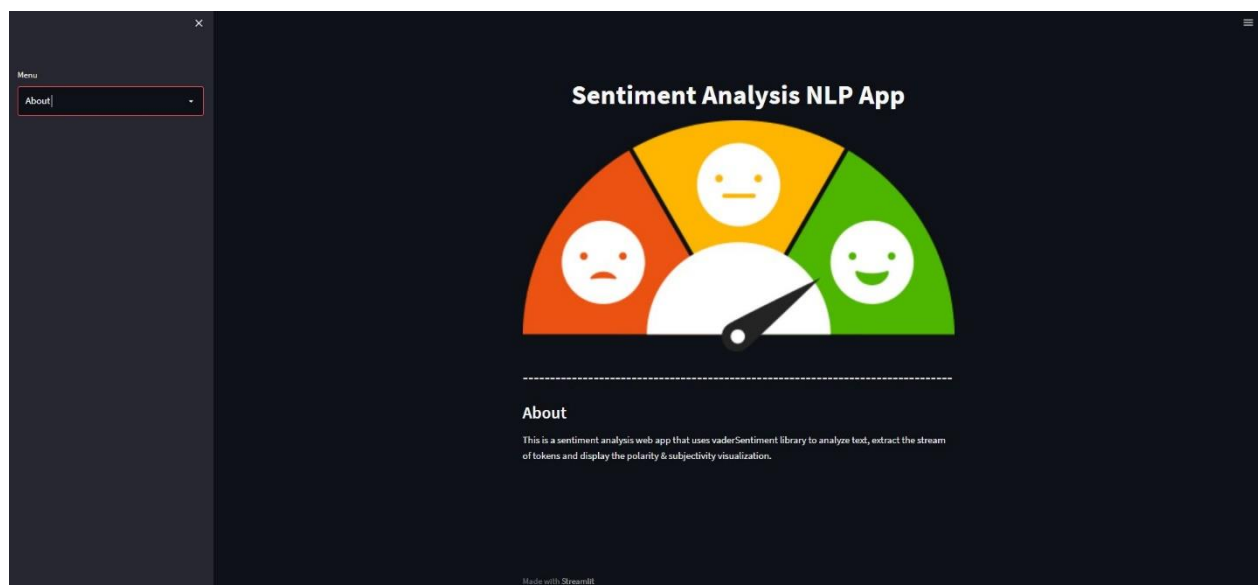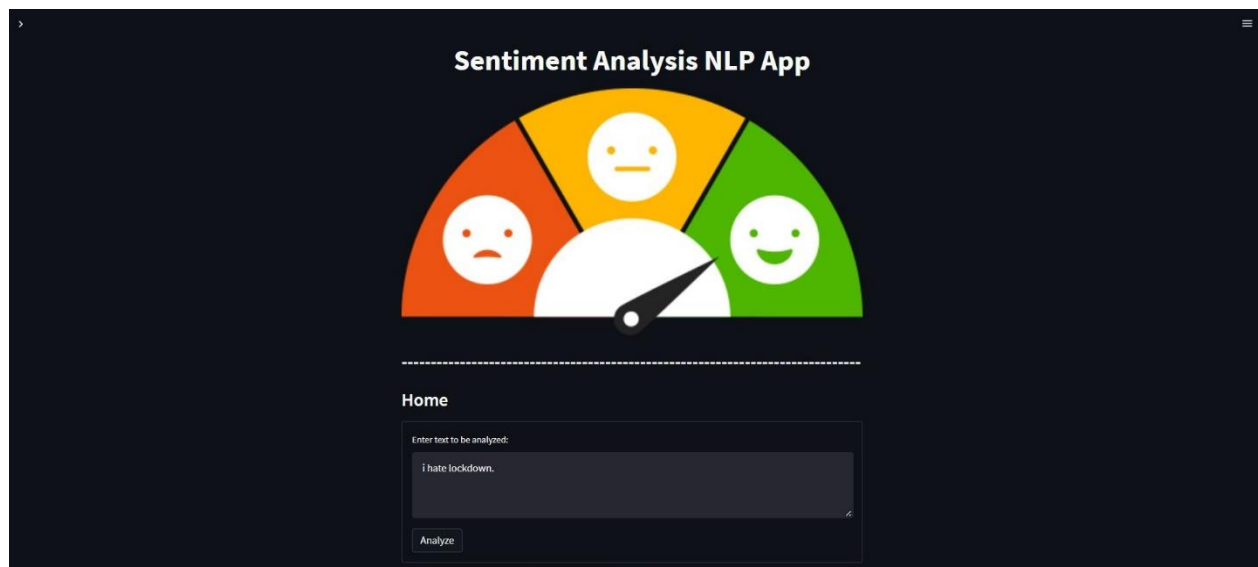
```
In [56]: # Classification report of Performance metrics
         label=['neutral','positive','negative']
         print(classification_report(y_test,pred_lr_cv))
```

```
               precision    recall  f1-score   support

    Negative       0.79      0.79      0.79      4620
     Neutral       0.68      0.68      0.68      2314
    Positive       0.82      0.83      0.82      5414

    accuracy                           0.78     12348
   macro avg       0.76      0.76      0.76     12348
weighted avg       0.78      0.78      0.78     12348
```
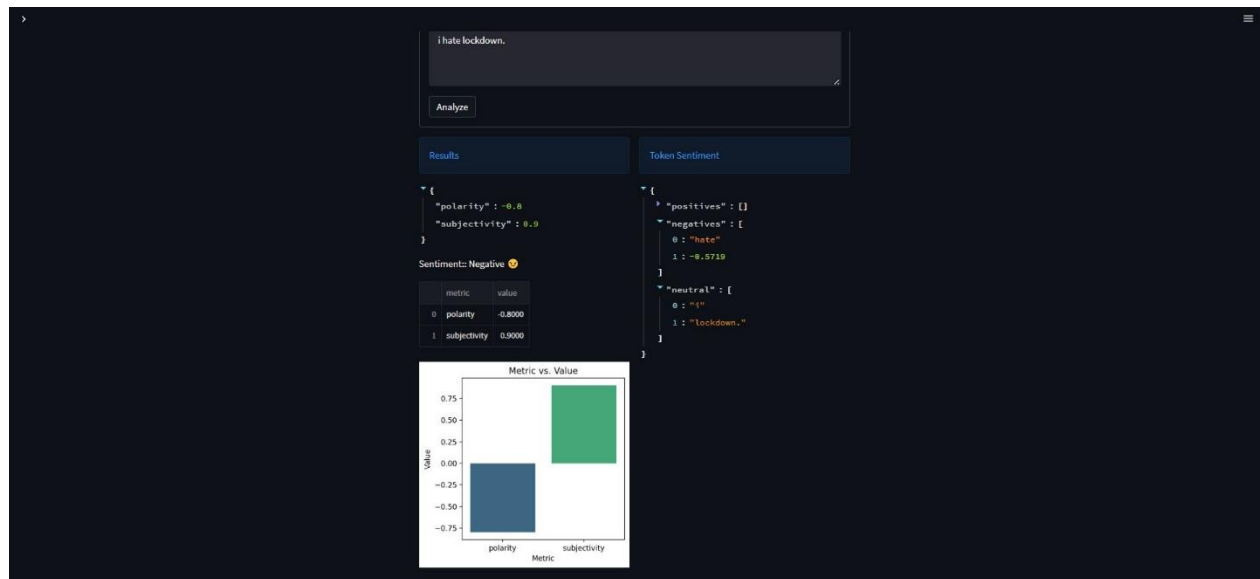
CHAPTER 8

SNAPSHOTS

8. <u>SNAPSHOTS</u>

CHAPTER 9

CONCLUSION

9. <u>CONCLUSION</u>

In summary, our project, "Sentiment Analysis of Lockdown in the USA During COVID-19," has yielded a robust and accessible tool for gaining deeper insights into public sentiment during a pivotal period. With a user-friendly interface, adaptable real-time and batch processing, and a focus on maintaining high accuracy in sentiment analysis, we have successfully addressed the challenges of understanding public opinion during the pandemic.

Ethical considerations and compliance with legal requirements have been at the forefront of our efforts, ensuring the responsible handling of user data and adherence to social media platform policies.

User feedback has played a vital role in our iterative development process, driving improvements in accuracy, usability, and feature enhancements. This feedback loop will continue to be instrumental in keeping the system relevant and valuable.

We have strived to make the application as accessible as possible, with comprehensive documentation that aids users and developers alike in understanding and effectively utilizing the tool.

CHAPTER 10

REFERENCE

10. <u>REFERENCE</u>

GITHUB REPOSITORY LINK ALONG WITH BONUS TASK:
https://github.com/suryaNB/Sentiment-Analysis-Twitter-TIE-Internship-

MACHINE LEARNING LIBRARIES:
https://towardsdatascience.com/top-5-machine-learning-libraries-in-python-e36e3e0e02af

MACHINE LEARNING INTRODUCTION:
https://www.javatpoint.com/machine-learning

PYTHON DOCUMENTATION:
https://docs.python.org/3/