VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Belagavi – 590018,
Karnataka

INTERNSHIP REPORT ON

"Sentiment Analysis of COVID 19 Tweets (INDIA)"

Submitted in the partial fulfillment of the requirement for 7th semester
BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING

Submitted by:

SHREYAS CR
[1VJ20CS046]

Conducted at

COMPSOFT TECHNOLOGIES

VIJAYA VITTALA INSTITUTE OF TECHNOLOGY
Department of COMPUTER SCIENCE
BENGALURU –560 077

VIJAYA VITTALA INSTITUTE OF TECHNOLOGY

BENGALURU –560 077

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the Internship titled **" Sentiment Analysis of COVID 19 Tweets (INDIA)"** carried out by **SHREYAS.C.R,** a bonafide student of **VIJAYA VITTALA INSTITUTE OF TECHNOLOGY**, 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.

DECLARATION

I,**SHREYAS.C.R [1VJ20CS046]** , final year student of Computer Science, **VIJAYA VITTALA INSTITUTE OF TECHNOLOGY** - 560 077, declare that the Internship has been successfully completed, in **COMPSOFT TECHNOLOGIES** 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:
Place:BANGALORE
USN:1VJ20CS046
NAME:SHREYAS.C.R

OFFER LETTER

COMPSOFT TECHNOLOGIES AUTHORIZED SIGNATORY

INTERNSHIP OFFER LETTER

Date: 14th August, 2023

Name: **Shreyas CR**
USN: **1VJ20CS046**
Placement ID: **1509ML019**

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning with Python (Research Based)** Internship position with **Compsoft Technologies**, effective Start Date **14th 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,

Nithin K. S
Project Manager
COMPSOFT TECHNOLOGIES
No. 363, 19th main road,
1st Block Rajajinagar
Bangalore - 560010

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, **Dr. Rajendra S** for providing us adequate facilities to undertake this Internship.

We would like to thank our **Mrs.Mangala Patel** Professor and Head of Department, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We express our deep and profound gratitude to our guide **Mr.William Deepan**,Assistant 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.

NAME:SHREYAS.CR
USN:1VJ20CS046

ABSTRACT

The outbreak of the COVID-19 pandemic in early 2020 had a profound impact on societies worldwide. In this context, social media platforms emerged as powerful channels for individuals to share their thoughts, experiences, and concerns. These platforms have become valuable sources of data for understanding public sentiment during significant events such as the pandemic.

Sentiment Analysis of COVID-19 Tweets in India project is dedicated to investigating the sentiments expressed in tweets originating from India, providing a comprehensive analysis of the evolving public sentiment.This project aims to gain insights into the evolving sentiments expressed by the Indian population on social media during the COVID-19 pandemic. With the proliferation of social media platforms as primary sources of information and expression, analyzing public sentiment becomes crucial for understanding the impact of the pandemic on society. This study focuses on tweets originating from India, offering a comprehensive analysis of the sentiments, trends, and factors influencing public opinions.

Methodology: The analysis utilizes a combination of natural language processing (NLP) techniques and machine learning models. A large dataset of COVID-19 related tweets is collected and preprocessed to prepare the text data. Sentiment analysis is performed using a state-of-the-art model, and the sentiments expressed in tweets are categorized as positive, negative, or neutral.

Results: The findings reveal a nuanced and evolving sentiment landscape. The sentiment analysis identifies prevalent emotions, tracks changes in sentiment over time, and explores regional variations in sentiment within India. The study discusses the factors contributing to sentiment shifts, including government policies, vaccination campaigns, and disease trends.

Implications: Understanding public sentiment during a pandemic is invaluable for policymakers, healthcare authorities, and the public. The analysis provides insights into the challenges faced, the level of public compliance with guidelines, and potential areas for improvement in communication strategies during health crises.

Challenges and Future Directions: The project faces challenges related to data quality and class imbalance in sentiment labels. Future research may explore more sophisticated sentiment models, consider additional data sources, and investigate the impact of misinformation on public sentiment.

This sentiment analysis of COVID-19 tweets in India contributes to the broader understanding of societal responses to health crises and informs strategies for effective communication and public engagement. The project underscores the significance of leveraging social media data for real-time sentiment monitoring and public health decision-making.

Table of Contents

CHAPTER 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 todefiine their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstormingsession, 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 itin one sentence ” Technology helps you to Delight your Customers” and that is what we want to achieve.

CHAPTER 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

The COVID-19 pandemic, which unfolded in early 2020, has been one of the most significant global crises in recent memory. Beyond its profound impact on public health, the pandemic has also triggered a surge in information sharing and discourse on social media platforms. India, as one of the world's most populous countries and a rapidly digitizing nation, has witnessed a substantial volume of COVID-19-related discussions on platforms like Twitter. These discussions encapsulate a wide array of emotions, opinions, and concerns expressed by individuals from diverse backgrounds.

This project is aimed at harnessing the power of machine learning and natural language processing (NLP) techniques to perform sentiment analysis on COVID-19-related tweets originating in India. Sentiment analysis, also known as opinion mining, is a sub field of NLP that focuses on determining the sentiment or emotional tone expressed within a piece of text. In the context of the pandemic, understanding public sentiment in India can provide valuable insights into how individuals are perceiving and reacting to various aspects of the crisis.

Introduction to Machine Learning

Machine learning, a subset of artificial intelligence, equips computers with the ability to learn and improve their performance on a specific task through data-driven experiences. In the realm of sentiment analysis, machine learning models are trained on large datasets of text to recognize patterns, associations, and linguistic features that allow them to classify text as expressing positive, negative, or neutral sentiment.

Machine learning algorithms, such as decision trees, random forests, support vector machines, and deep learning models like neural networks, have demonstrated exceptional capabilities in understanding and interpreting textual data. These algorithms can be applied to large volumes of tweets to classify their sentiment accurately and efficiently.

Problem Statement

The problem at hand involves the analysis of COVID-19-related tweets originating in India with the goal of determining the sentiment expressed within these tweets. Specifically, this project seeks to address the following:

Sentiment Classification: Develop a sentiment analysis model that classifies each tweet into one of three sentiment categories: positive, negative, or neutral.

Sentiment Trends: Analyze temporal trends in sentiment to understand how public sentiment has evolved over the course of the COVID-19 pandemic. Identify key inflection points, shifts, or spikes in sentiment and associate them with significant events or developments.

Regional Variations: Explore regional variations in sentiment within India. Determine whether sentiments differ among states, cities, or specific geographical regions, and investigate potential factors contributing to these variations.

Influencing Factors: Investigate external factors that influence sentiment, including government policies, public health campaigns, media coverage, and the spread of misinformation. Identify the drivers of sentiment changes and their impact on public discourse.

The successful execution of this project will not only provide a comprehensive understanding of public sentiment during the pandemic in India but also offer insights that can inform policymaking, public health communication, and the development of strategies to address public concerns effectively.

CHAPTER 4 SYSTEM ANALYSIS

1. Existing System:

Background:

The existing system refers to the state of sentiment analysis in the context of COVID-19 tweets in India before the initiation of this project. At the outset, the sentiment analysis landscape was characterized by limited regional specificity, broad sentiment labels, and relatively simple NLP techniques. Researchers and data analysts typically conducted manual sentiment labeling or employed basic sentiment lexicons to analyze the overall sentiment of tweets.

Challenges:

The existing system faced several challenges :

- Lack of Regional Specificity: Sentiment analysis models often overlooked regional variations in sentiment within India, providing generalized insights.
- Complexity of Emotional Expression: COVID-19 tweets often expressed complex emotions and opinions that were not fully captured by rudimentary sentiment lexicons.
- Limited Insights: The existing system offered limited insights into the evolving public sentiment and failed to account for temporal variations during different pandemic phases.

2. Proposed System:

Introduction:

The proposed system represents a significant enhancement in the sentiment analysis of COVID-19 tweets in India. It leverages advanced natural language processing (NLP) techniques, machine learning models, and a comprehensive dataset to provide detailed and real-time sentiment analysis.

Key Features:

The proposed system includes the following features:

- Regional Sensitivity: The system is designed to identify and account for regional variations in sentiment within India, allowing for a more nuanced understanding of localized responses to the pandemic.

- Temporal Analysis: It conducts temporal sentiment analysis, tracking sentiment changes over time. This enables the identification of shifts in public sentiment during different phases of the pandemic.

- Emotion Recognition: The system employs advanced NLP techniques to recognize complex emotions and opinions expressed in tweets, going beyond simple positive, negative, or neutral classifications.

- Data Integration: In addition to tweets, the system integrates data from other sources such as news articles and official reports to provide a more comprehensive sentiment analysis.

3. Objectives of the System:

The primary objectives of the proposed system for Sentiment Analysis of COVID-19 Tweets in India are:

- Regional Sensitivity: To develop a sentiment analysis model that can discern and analyze regional variations in sentiment within India, providing localized insights.

- Temporal Analysis: To implement a temporal sentiment analysis component, allowing for the identification of shifts in public sentiment over different time periods.

- Emotion Recognition: To employ advanced NLP techniques for the recognition of complex emotions and opinions expressed in tweets, resulting in more detailed sentiment labels.

- Data Integration: To integrate data from diverse sources, including tweets, news articles, and official reports, to enhance the comprehensiveness of sentiment analysis.

- Real-time Insights: To provide real-time insights into the evolving public sentiment during the ongoing COVID-19 pandemic, assisting policymakers, healthcare authorities, and the public in making informed decisions.

- Improved Communication: To contribute to improved communication strategies by addressing public misconceptions, concerns, and emotional needs.

By achieving these objectives, the proposed system aims to advance the field of sentiment analysis and enhance our understanding of how the public responds to health crises, ultimately contributing to more effective public health interventions and crisis management.

CHAPTER 5

REQUIREMENT ANALYSIS

Hardware Requirement Specification:

1. Computing Infrastructure:

- High-performance computing systems capable of handling large datasets and running machine learning models efficiently.

2. Storage:

- Sufficient storage capacity for storing datasets, models, and intermediate results.

3. Memory (RAM):

- Adequate memory to support data processing and machine learning tasks.

4. Processor:

- Multi-core processors for parallel processing of data and training machine learning models.

Software Requirement Specification:

1. Programming Languages and Libraries:

- Python: Python is the primary programming language for data preprocessing, analysis, and model development.
- NLTK (Natural Language Toolkit): For text processing and sentiment analysis.
- Scikit-learn: For implementing machine learning models.
- TensorFlow or PyTorch: For deep learning if you plan to use advanced models like BERT.
- Pandas: For data manipulation and analysis.
- Numpy: For numerical computations.
- Matplotlib and Seaborn: For data visualization.

2. Sentiment Analysis Tools:

- VADER (Valence Aware Dictionary and sEntiment Reasoner): A pre-built sentiment analysis tool for understanding the sentiment in text data.

- TextBlob: Another library for sentiment analysis and part-of-speech tagging.

3. Data Collection and Preprocessing:

- Tweepy: To access and collect tweets from Twitter.
- Pandas: For data cleaning and preprocessing.

4. Text Preprocessing:

- NLTK: For tasks like tokenization, stop word removal, and stemming.
- Spacy: An alternative to NLTK for text preprocessing tasks.

5. Machine Learning Models:

- Scikit-learn: For implementing traditional machine learning models like Decision Trees, Random Forests, and SVM.
- XGBoost or LightGBM: For gradient boosting models, if necessary.
-Transformers (Hugging Face): If using BERT or other transformer-based models for sentiment analysis.

6. Web Development (Optional):

- If you plan to create a web-based interface for users to interact with the sentiment analysis tool:
- Flask or Django: Python web frameworks.
- HTML/CSS: For creating the user interface.
- JavaScript: For adding interactivity to the web application.

7. Version Control:

- Git: For tracking changes in your codebase.

8. Integrated Development Environment (IDE):

- Jupyter Notebook or Visual Studio Code: Popular IDEs for Python.

9. Database (Optional):

- SQLite or MySQL: If you need to store metadata about the tweets or user interactions with your tool.

10. Deployment and Hosting (Optional):

- Cloud Services (e.g., AWS, Google Cloud, or Azure): For hosting web applications or machine learning models.

11. Collaboration and Documentation:

- GitHub: For code repository and collaboration.
- Jupyter Notebook or Markdown: For documenting your project and analysis.

12. Project Management (Optional):

- Trello or Asana: For project planning and task management.

These hardware and software requirements provide a foundation for conducting sentiment analysis on COVID-19 tweets in India. Depending on the scope and scale of your project, you may need to adapt and scale these requirements accordingly.

CHAPTER 6

DESIGN ANALYSIS

1. Data Collection and Preprocessing:

- Data Sources: Explain the sources from which you collected tweets related to COVID-19. These may include social media platforms, APIs, or web scraping tools. Specify why these sources were chosen.
-Data Filtering: Describe any criteria used to filter the dataset, such as selecting tweets with location tags related to India. Explain why this filtering is essential for the project's focus.

2. Data Preprocessing:

-Text Cleaning: Detail the steps taken to clean the text data. This includes removing special characters, converting text to lowercase, and handling missing or irrelevant data.
- Stopword Removal: Explain the importance of removing stopwords (common words like "and" or "the") to reduce noise in the text data.
- Stemming/Lemmatization: Discuss the use of stemming or lemmatization to reduce words to their base form and why it is applied.

3. Sentiment Analysis Model:
- Choice of Model: Clarify the choice of the sentiment analysis model (e.g., VADER, supervised machine learning, deep learning). Discuss why this model is suitable for your analysis.
- Labeling and Classification: Explain the sentiment labels used (positive, negative, neutral) and how the model classifies tweets into these categories.
-Training Data: If applicable, mention any training data used to fine-tune the sentiment analysis model.

4. Feature Engineering:
-TF-IDF Vectorization: Describe the use of TF-IDF (Term Frequency-Inverse Document Frequency) vectorization to convert text data into numerical features. Explain why TF-IDF was chosen over other techniques.
-Dimensionality Reduction: If employed, discuss techniques like Principal Component Analysis (PCA) or feature selection methods to reduce the dimensionality of the TF-IDF matrix.

5. Evaluation Metrics:
- Choice of Metrics: Specify the evaluation metrics used to assess the model's performance (e.g., accuracy, precision, recall, F1 score).
- Justification: Explain why these specific metrics were chosen and how they provide a comprehensive view of the model's effectiveness.

6. Implementation:
- Tools and Libraries: List the programming languages, libraries (e.g., Python, NLTK, scikit-learn), and frameworks used for implementation.
- Workflow: Describe the workflow, including data loading, preprocessing, model training, evaluation, and visualization.

7. Challenges and Solutions:
- Data Quality Issues: Discuss any challenges related to data quality and how they were addressed during preprocessing.
- Class Imbalance: If present, explain how you handled class imbalance in the sentiment labels (e.g., oversampling, undersampling).

8. Future Improvements:
- Enhancements: Suggest potential improvements to the design, such as using more advanced sentiment analysis models, incorporating external data sources, or addressing misinformation detection.

9. Ethical Considerations:
-Privacy: Discuss any privacy concerns related to collecting and analyzing user-generated social media data. Explain how you ensured the ethical use of data.

10. Limitations:
-Scope: Highlight the limitations of the project, such as its focus on Twitter data and its applicability to the Indian context. Acknowledge any constraints.

This section provides a detailed account of the key decisions and methodologies used in your sentiment analysis project. It offers readers a clear understanding of your project's design choices and their implications for the analysis and results.

CHAPTER 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 theobjectives 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

```
1  data.head()
```

	number	sentiment	text
0	3204	sad	agree the poor in india are treated badly thei...
1	1431	joy	if only i could have spent the with this cutie...
2	654	joy	will nature conservation remain a priority in ...
3	2530	sad	coronavirus disappearing in italy show this to...
4	2296	sad	uk records lowest daily virus death toll since...

```
[ ]  1  data.tail()
```

Snap 1:Table head

```
rf_classifier = RandomForestClassifier(random_state=42)
accuracy_rf, precision_rf, recall_rf, f1_rf, conf_matrix_rf = evaluate_classifier(rf_classifier, X_train_tfidf, y_train, X_test_tfidf, y_test)

# Initialize and evaluate Naive Bayes
nb_classifier = MultinomialNB()
accuracy_nb, precision_nb, recall_nb, f1_nb, conf_matrix_nb = evaluate_classifier(nb_classifier, X_train_tfidf, y_train, X_test_tfidf, y_test)

# Initialize and evaluate Support Vector Machine (SVM)
svm_classifier = SVC(kernel='linear', random_state=42)
accuracy_svm, precision_svm, recall_svm, f1_svm, conf_matrix_svm = evaluate_classifier(svm_classifier, X_train_tfidf, y_train, X_test_tfidf, y_

# Initialize and evaluate Logistic Regression
lr_classifier = LogisticRegression(max_iter=1000, random_state=42)
accuracy_lr, precision_lr, recall_lr, f1_lr, conf_matrix_lr = evaluate_classifier(lr_classifier, X_train_tfidf, y_train, X_test_tfidf, y_test)

# Print evaluation metrics for each classifier
print("Metrics for Other Classifiers:")
print(f"Decision Tree - Accuracy: {accuracy_dt:.2f}, Precision: {precision_dt:.2f}, Recall: {recall_dt:.2f}, F1 Score: {f1_dt:.2f}")
print(f"Random Forest - Accuracy: {accuracy_rf:.2f}, Precision: {precision_rf:.2f}, Recall: {recall_rf:.2f}, F1 Score: {f1_rf:.2f}")
print(f"Naive Bayes - Accuracy: {accuracy_nb:.2f}, Precision: {precision_nb:.2f}, Recall: {recall_nb:.2f}, F1 Score: {f1_nb:.2f}")
print(f"SVM - Accuracy: {accuracy_svm:.2f}, Precision: {precision_svm:.2f}, Recall: {recall_svm:.2f}, F1 Score: {f1_svm:.2f}")
print(f"Logistic Regression - Accuracy: {accuracy_lr:.2f}, Precision: {precision_lr:.2f}, Recall: {recall_lr:.2f}, F1 Score: {f1_lr:.2f}")

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
Metrics for Other Classifiers:
Decision Tree - Accuracy: 0.52, Precision: 0.52, Recall: 0.52, F1 Score: 0.52
Random Forest - Accuracy: 0.63, Precision: 0.63, Recall: 0.63, F1 Score: 0.62
Naive Bayes - Accuracy: 0.68, Precision: 0.69, Recall: 0.68, F1 Score: 0.68
SVM - Accuracy: 0.69, Precision: 0.69, Recall: 0.69, F1 Score: 0.69
Logistic Regression - Accuracy: 0.70, Precision: 0.70, Recall: 0.70, F1 Score: 0.70
```

Snap 2

Initial results 1

Decision Tree - Accuracy: 0.52, Precision: 0.52, Recall: 0.52, F1 Score: 0.52
Random Forest - Accuracy: 0.63, Precision: 0.63, Recall: 0.63, F1 Score: 0.62
Naive Bayes - Accuracy: 0.68, Precision: 0.69, Recall: 0.68, F1 Score: 0.68
SVM - Accuracy: 0.69, Precision: 0.69, Recall: 0.69, F1 Score: 0.69
Logistic Regression - Accuracy: 0.70, Precision: 0.70, Recall: 0.70, F1 Score: 0.70

```python
if __name__ == "__main__":
    while True:
        # Take user input
        user_input = input("Enter text (or 'exit' to quit): "

        # Check if the user wants to exit
        if user_input.lower() == 'exit':
            break

        # Perform sentiment analysis
        sentiment = analyze_sentiment(user_input)
        print(f"Sentiment: {sentiment}")

```

```
Enter text (or 'exit' to quit): sad
Sentiment: Negative
Enter text (or 'exit' to quit): happy
Sentiment: Positive
Enter text (or 'exit' to quit): neutral
Sentiment: Neutral
Enter text (or 'exit' to quit): exit
```

Snap 3:Sentimental analysis

Here we got **89 percent accuracy**

```python
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

# Make predictions
predictions = rf_classifier.predict(tfidf_matrix)

# Calculate evaluation metrics
accuracy = accuracy_score(data['sentiment'], predictions)
precision = precision_score(data['sentiment'], predictions, average='weighted')
recall = recall_score(data['sentiment'], predictions, average='weighted')
f1 = f1_score(data['sentiment'], predictions, average='weighted')

print(f"Accuracy: {accuracy:.2f}")
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1 Score: {f1:.2f}")
```

```
Accuracy: 0.89
Precision: 0.89
Recall: 0.89
F1 Score: 0.89
```

Snap 4: final accuracy

Results:

Accuracy: 0.89

Precision: 0.89

Recall: 0.89

F1 Score: 0.89

CHAPTER 9

CONCLUTION

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.

REFERENCE

1. Dataset from kaggle.
2. Code executed in colab & visual studio