A

Mini Project

On

ANALYSIS OF SOCIETY OPINION ON WOMEN SAFETY THROUGH SOCIAL MEDIA PLATFORMS

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

BANDARI SAI KUMAR (197R1A05C9)
GUDIMETLA SRIKANTH (197R1A05E0)
JOGANNAGARI SAINATH REDDY (197R1A05E2)

Under the Guidance of

A.UDAY KIRAN

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGCAct.1956, Kandlakoya (V), Medchal Road, Hyderabad-501401.

2019-2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled **"ANALYSIS OF SOCIETY OPINION ON WOMEN SAFETY THROUGH SOCIAL MEADIA PLATFORMS "** being submitted by **B.SAIKUMAR(197R1A05C9),G.SRIKANTH(197R1A05E0)&J.SAINATHREDDY(197R 1A05E2)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by them under our guidance and supervision during the year 2022-23.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

A.Uday Kiran   **Dr. A. Raji Reddy**
(Associate Professor)   DIRECTOR
 INTERNAL GUIDE

Dr. K. Srujan Raju   **EXTERNAL EXAMINER**
      HOD

Submitted for viva voice Examination held on

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We take this opportunity to express my profound gratitude and deep regard to my guide **A.Uday Kiran,** Associate Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to the Project Review Committee (PRC) **Dr. Punyaban Patel, Ms. Shilpa, Dr.M . Subha Mastan Rao & J. Narasimharao** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to **Dr. K. Srujan Raju,** Head, Department of Computer Science and Engineering for providing encouragement and support for completing this project successfully.

We are obliged to **Dr. A. Raji Reddy,** Director for being cooperative throughout the course of this project. We also express our sincere gratitude to Sri. **Ch. Gopal Reddy,** Chairman for providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of **CMR Technical Campus** who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement, without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

B. SAI KUMAR	**(197R1A05C9)**
G. SRIKANTH	**(197R1A05E0)**
J. SAINATH REDDY	**(197R1A05E2)**

ABSTRACT

Women and girls have been experiencing a lot of violence and harassment in public places in various cities starting from stalking and leading to abuse harassment or abuse assault. This project basically focuses on the role of social media in promoting the safety of women in Indian cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram and some other newly introduced social media platforms. Twitter and other Twitter handles which include hash tag messages that are widely spread across the whole globe sir as a platform for women to express their views about how they feel while we go out for work or travel in a public transport and what is the state of their mind when they are surrounded by unknown men and whether these women feel safe or not?

This project basically focuses on the role of social media in promoting the safety of women in Indian cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram and some other newly introduced social media platforms. There exists several opinion-oriented information gathering and analytics systems that aim to extract people's opinion regarding different topics.The raw data is picked up from Kaggle website.

LIST OF FIGURES/TABLES

LIST OF SCREENSHOTS

TABLE OF CONTENTS

1. INTRODUCTION

1. INTRODUCTION

1.1 PROJECT SCOPE

This project is titled "Analysis of society opinion on women safety through social media platforms". This project is to analyse women safety using social networking messages and by applying machine learning algorithms on it. Now-a-days almost all peoples are using social networking sites to express their feelings and if any women feel unsafe in any area, then she will express negative words in her post/tweets/messages and by analysing those messages we can detect which area is more unsafe for women's.

1.2 PROJECT PURPOSE

The main purpose of this project is to identify Analysis of Women Safety in Indian Cities Using Machine Learning on Tweet

1.3 PROJECT FEATURES

The main features of this project are that this model classifies the given data into multiple statements so that the recognization of the abuse statement becomes very easy with the help of NLP and NLTK.NLP is called as Natural Language Processing and NLTK is called as Natural Language Tool Kit.In this project we are using Lexicon-Based Learning.

2. SYSTEM ANALYSIS

2. SYSTEM ANALYSIS

SYSTEM ANALYSIS

System Analysis is the important phase in the system development process. The System is studied to the minute details and analyzed. The system analyst plays an important role of an interrogator and dwells deep into the working of the present system. In analysis, a detailed study of these operations performed by the system and their relationships within and outside the system is done. A key question considered here is, "what must be done to solve the problem?" The system is viewed as a whole and the inputs to the system are identified. Once analysis is completed the analyst has a firm understanding of what is to be done.

2.1 PROBLEM DEFINITION

As People communicate and share their opinion actively on social medias including Facebook and Twitter, Social network can be considered as a perfect platform to learn about people's opinion and sentiments regarding different events. There exists several opinion-oriented information gathering and analytics systems that aim to extract people's opinion regarding different topics. Since Twitter contains short texts, people tend to use different words and abbreviations. These phrases are difficult to extract their sentiment by current NLP systems easily. Therefore, many researchers have used deep learning and machine learning techniques to extract and mine the polarity of the phrase.

2

2.2 EXISTING SYSTEM

People often express their views freely on social media about what they feel about the Indian society and the politicians. The tweets or comments about safety of women and stories of standing up against abuse harassment further motivates other women data on the same social media website or application like Twitter.

Earlier we used data mining techniques to extract data from twitter.The efficiency of this is very less.

2.2.1 DISADVANTAGES OF EXISTING SYSTEM

- Twitter and Instagram point and most of the people are using it to express their emotions and also their opinions about what they think about the Indian cities and Indian society.

- There are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning.

- Also there are another categorization Janta presented with categories of statistical knowledge-based and age wise differentiation approaches.

2.3 PROPOSED SYSTEM

As the efficiency of data mining in extracting is less,we proposing the new system by the combination of NLTK(Natural language toolkit) and NLP(Natural language processing). NLTK is a platform used for building python programs that work with human language data for applying in statistical natural language processing.It contains text processing libraries for tokenization,parsing,classification and semantic reasoning. (NLP)Natural language processing helps computers communicate with humans in their own language and scales other languages related tasks NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment and determine which parts are important.

3

2.3.1 ADVANTAGES OF THE PROPOSED SYSTEM

- The data set that was obtained through Twitter about the status of women safety in Indian society.

- There will be quick punishments if the comment is very abusive.

- Easy to integrate

2.4 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and a business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. Three key considerations involved in the feasibility analysis:

- EconomicFeasibility
- TechnicalFeasibility
- SocialFeasibility

4

2.4.1 ECONOMIC FEASIBILITY

The developing system must be justified by cost and benefit. Criteria to ensure that effort is concentrated on a project, which will give best, return at the earliest. One of the factors, which affect the development of a new system, is the cost it would require. The following are some of the important financial questions asked during preliminary investigation:

- The costs conduct a full system investigation.
- The cost of the hardware and software.
- The benefits in the form of reduced costs or fewer costly errors.

Since the system is developed as part of project work, there is no manual cost to spend for the proposed system. Also all the resources are already available, it give an indication that the system is economically possible for development.

2.4.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

2.4.3 BEHAVIORAL FEASIBILITY

This includes the following questions:

- Is there sufficient support for the users?
- Will the proposed system cause harm?

The project would be beneficial because it satisfies the objectives when developed and installed. All behavioral aspects are considered carefully and conclude that the project is behaviorally feasible

5

2.5 HARDWARE & SOFTWARE REQUIREMENTS

2.5.1 HARDWARE REQUIREMENTS:

Hardware interfaces specify the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements.

- Processor   : Pentium IV or higher processor
- Hard disk   : minimum 512MB space in Hard Disk.
- RAM   : 256 MB RAM
- Input devices : Keyboard, mouse.

2.5.2 SOFTWARE REQUIREMENTS:

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements,

- Operating system : Windows 8 and above.
- Languages   : Python, tkinter.
- Tools   : Python IDEL3.7 version, vscode.

3. ARCHITECTURE

3. ARCHITECTURE

3.1 PROJECT ARCHITECTURE

This project architecture shows the procedure followed for classification, starting from input to final prediction.
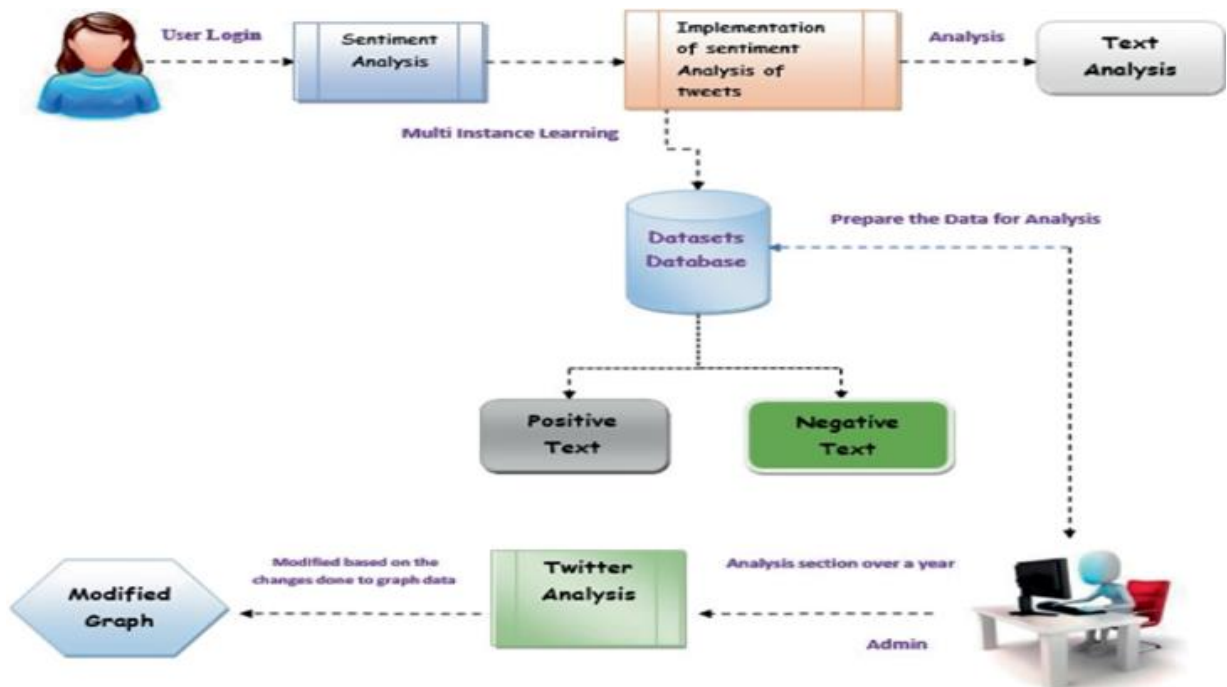

Figure 3.1: Analysis of society opinion on women safety through social media platforms

7

3.2 DESCRIPTION

Every user data such as credentials, new tweets, re-tweets and tweet score will be stored in the database for the admin to monitor and perform the analysis. The sentiment analysis is applied on the user data in order to monitor and confirm whether any tweets are abusive to women or not. Admin performs this analysis on each and every user tweets to provide safety for the women. Sentimental analysis will be implemented on the tweets of user that are stored in the database. Admin can now prepare the data to perform the analysis. The tweets made by every user of the application will be called as the initial input for the sentiment analysis and hence they will be the dataset. Along with this, text analysis graph can also be shown. Admin will store the filters in the database. Filters are the keywords for which the tweet context will be searched for in order to declare as abusive or not. There can be two types of filters – positive keyword and negative keyword. Positive keywords are those words which are abusive or disrespect the women by any means. Negative keywords are the words which are normal and will not abuse the women.

3.3 USE CASE DIAGRAM

In the use case diagram, we have basically one actor who is the user in the trained model. A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has. The use cases are represented by either circles or ellipses. The actors are often shown as stick figures.
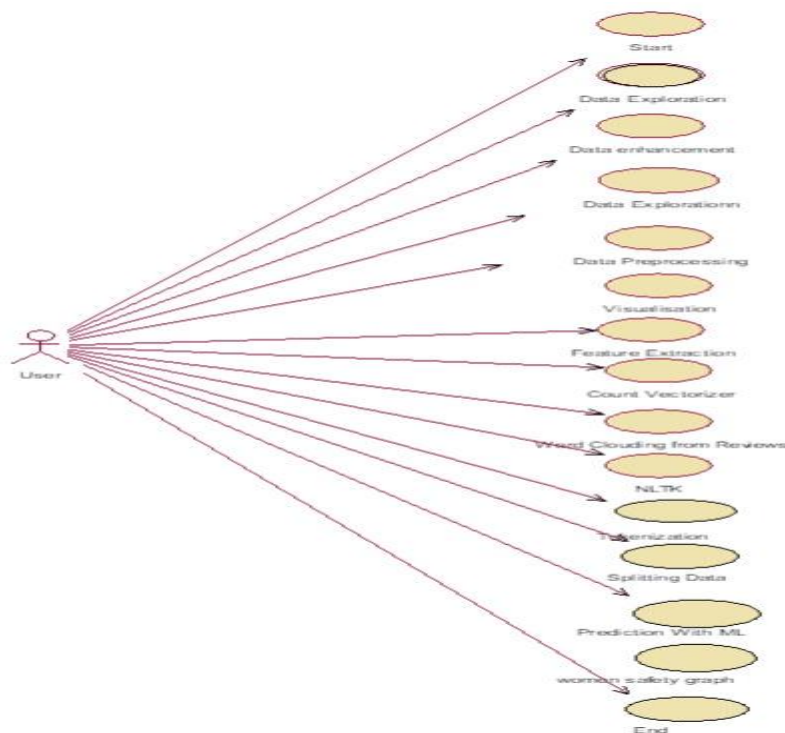

Figure 3.2: Use Case Diagram for Analysis of society opinion on women
safety through social meadia platforms

3.4 CLASS DIAGRAM

Class diagram is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations(or methods), and the relationships among objects.
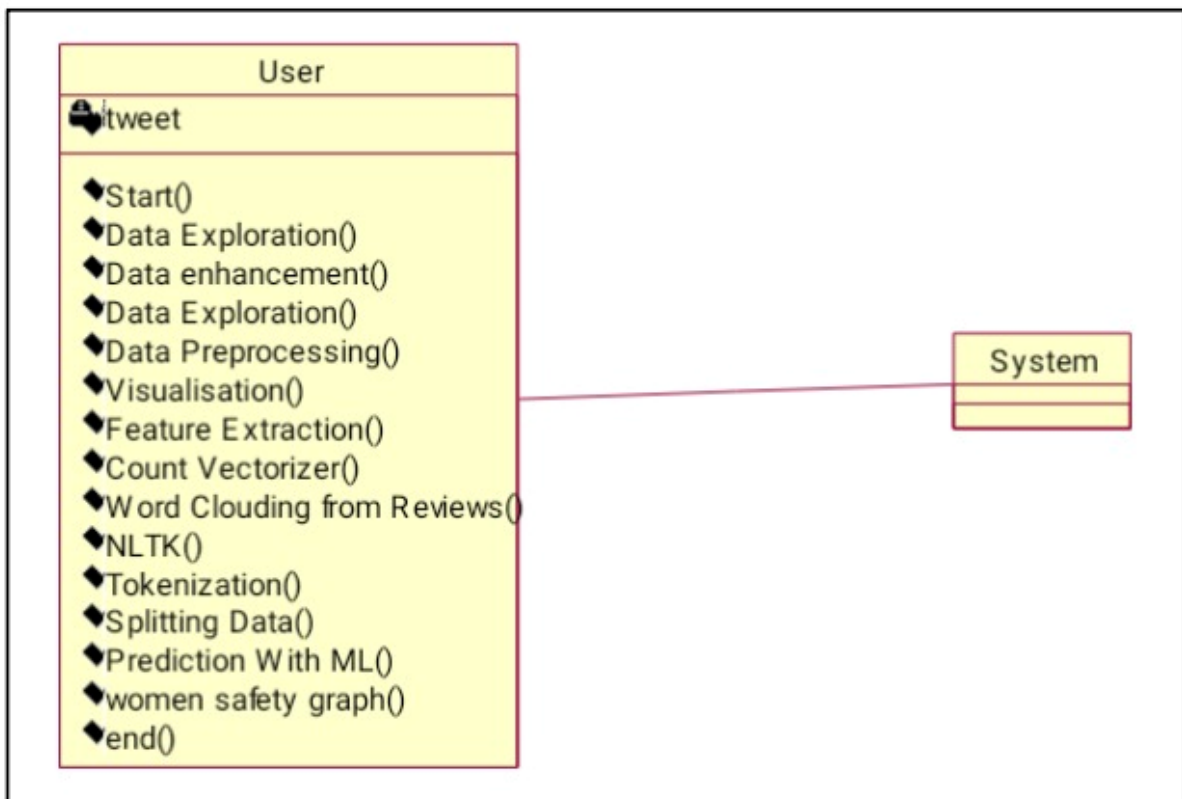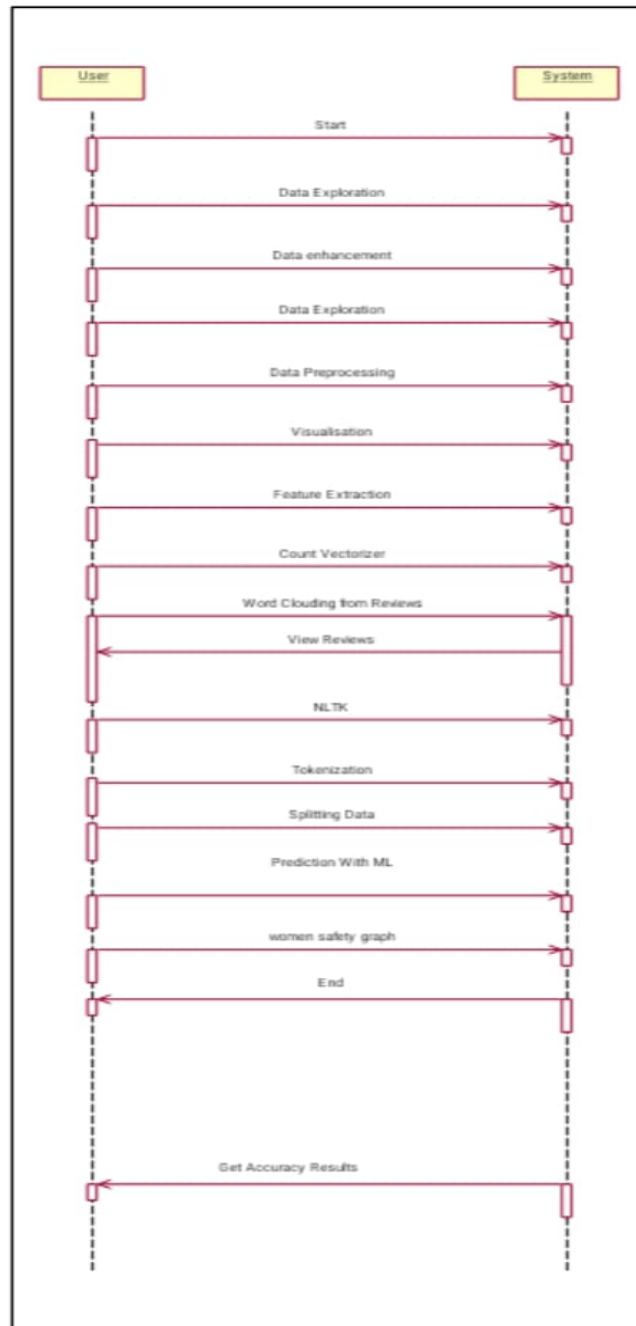

Figure 3.3: Class Diagram for Analysis of society opinion on women
safety through social meadia platforms

3.5 SEQUENCE DIAGRAM

A sequence diagram shows object interactions arranged in time sequence. It depicts the objects involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence   10

CMRTC

diagrams are typically associated with use case realizations in the logical view of the system under development.

CMRTC

Figure 3.4: Sequence Diagram for Analysis of society opinion on women
safety through social media platforms

3.6 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. They can also include elements showing the flow of data between activities through one or more data stores.
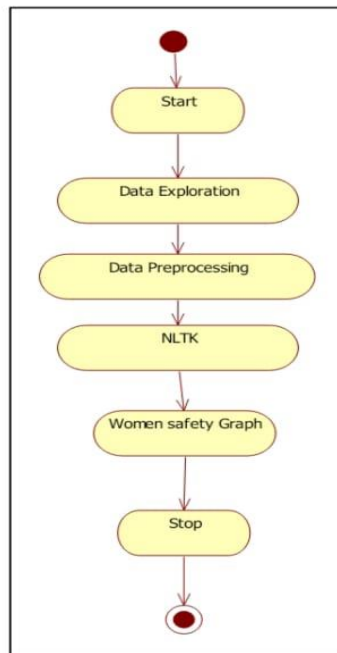

Figure 3.5: Activity Diagram for Analysis of society opinion on women
safety through social media platforms

4. IMPLEMENTATION

4.1 SAMPLE CODE

```python
import tkinter
from textblob import TextBlob
from tkinter import *
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from string import punctuation
from nltk.corpus import stopwords

main = tkinter.Tk()
main.title("Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets") #designing main screen
main.geometry("1300x1200")

global filename
tweets_list = []
clean_list = []
global pos, neu, neg

def tweetCleaning(doc):
    tokens = doc.split()
    table = str.maketrans('', '', punctuation)
    tokens = [w.translate(table) for w in tokens]
    tokens = [word for word in tokens if word.isalpha()]
    stop_words = set(stopwords.words('english'))
    tokens = [w for w in tokens if not w in stop_words]
    tokens = [word for word in tokens if len(word) > 1]
    tokens = ' '.join(tokens) #here upto for word based
    return tokens

def upload(): #function to upload tweeter profile
    global filename
    filename = filedialog.askopenfilename(initialdir="dataset")
    pathlabel.config(text=filename)
    text.delete('1.0', END)
    text.insert(END,filename+" loaded\n");

def read():
    text.delete('1.0', END)
    tweets_list.clear()
    train = pd.read_csv(filename,encoding='iso-8859-1')
    for i in range(len(train)):
        tweet = train.get_value(i, 'Text')
```

CMRTC

13

```python
        tweets_list.append(tweet)
        text.insert(END,tweet+"\n")
    text.insert(END,"\n\nTotal tweets found in dataset is :
"+str(len(tweets_list))+"\n\n\n")


def clean():
    text.delete('1.0', END)
    clean_list.clear()
    for i in range(len(tweets_list)):
        tweet = tweets_list[i]
        tweet = tweet.strip("\n")
        tweet = tweet.strip()
        tweet = tweetCleaning(tweet.lower())
        clean_list.append(tweet)
        text.insert(END,tweet+"\n")
    text.insert(END,"\n\nTotal tweets found in dataset is :
"+str(len(clean_list))+"\n\n\n")


def machineLearning():
    text.delete('1.0', END)
    global pos, neu, neg
    pos = 0
    neu = 0
    neg = 0
    for i in range(len(clean_list)):
        tweet = clean_list[i]
        blob = TextBlob(tweet)
        if blob.polarity <= 0.2:
            neg = neg + 1
            text.insert(END,tweet+"\n")
            text.insert(END,"Predicted Sentiment : NEGATIVE\n")
            text.insert(END,"Polarity Score     : "+str(blob.polarity)+"\n")

    text.insert(END,'=================================================
====================================\n')
        if blob.polarity > 0.2 and blob.polarity <= 0.5:
            neu = neu + 1
            text.insert(END,tweet+"\n")
            text.insert(END,"Predicted Sentiment : NEUTRAL\n")
            text.insert(END,"Polarity Score     : "+str(blob.polarity)+"\n")


    text.insert(END,'=================================================
====================================\n')
```

14

CMRTC

```python
        if blob.polarity > 0.5:
            pos = pos + 1
            text.insert(END,tweet+"\n")
            text.insert(END,"Predicted Sentiment : POSITIVE\n")
            text.insert(END,"Polarity Score      : "+str(blob.polarity)+"\n")

    text.insert(END,'=================================================
====================================\n')

def graph():
    label_X = []
    category_X = []
    text.delete('1.0', END)
    text.insert(END,"Saftey Factor\n\n")
    text.insert(END,'Positive : '+str(pos)+"\n")
    text.insert(END,'Negative : '+str(neg)+"\n")
    text.insert(END,'Neutral  : '+str(neu)+"\n\n")
    text.insert(END,'Length of tweets  : '+str(len(clean_list))+"\n")
    text.insert(END,'Positive : '+str(pos)+' / '+ str(len(clean_list))+' =
'+str(pos/len(clean_list))+'%\n')
    text.insert(END,'Negative : '+str(neg)+' / '+ str(len(clean_list))+' =
'+str(neg/len(clean_list))+'%\n')
    text.insert(END,'Neutral  : '+str(neu)+' / '+ str(len(clean_list))+' =
'+str(neu/len(clean_list))+'%\n')
    label_X.append('Positive')
    label_X.append('Negative')
    label_X.append('Neutral')
    category_X.append(pos)
    category_X.append(neg)
    category_X.append(neu)

    plt.pie(category_X,labels=label_X,autopct='%1.1f%%')
    plt.title('Women Saftey & Sentiment Graph')
    plt.axis('equal')
    plt.show()

font = ('times', 16, 'bold')
title = Label(main, text='Analysis of Women Safety in Indian Cities Using Machine
Learning on Tweets')
title.config(bg='brown', fg='white')
title.config(font=font)
title.config(height=3, width=120)
title.place(x=0,y=5)

font1 = ('times', 14, 'bold')
```

CMRTC

```python
uploadButton = Button(main, text="Upload Tweets Dataset", command=upload)
uploadButton.place(x=50,y=100)
uploadButton.config(font=font1)

pathlabel = Label(main)
pathlabel.config(bg='brown', fg='white')
pathlabel.config(font=font1)
pathlabel.place(x=370,y=100)

readButton = Button(main, text="Read Tweets", command=read)
readButton.place(x=50,y=150)
readButton.config(font=font1)

cleanButton = Button(main, text="Tweets Cleaning", command=clean)
cleanButton.place(x=210,y=150)
cleanButton.config(font=font1)

button = Button(main, text="natural language processing",
command=machineLearning)
mlButton.place(x=400,y=150)
mlButton.config(font=font1)

graphButton = Button(main, text="Women Saftey Graph", command=graph)
graphButton.place(x=730,y=150)
graphButton.config(font=font1)

font1 = ('times', 12, 'bold')
text=Text(main,height=25,width=150)
scroll=Scrollbar(text)
text.configure(yscrollcommand=scroll.set)
text.place(x=10,y=200)
text.config(font=font1)


main.config(bg='brown')
main.mainloop()
```
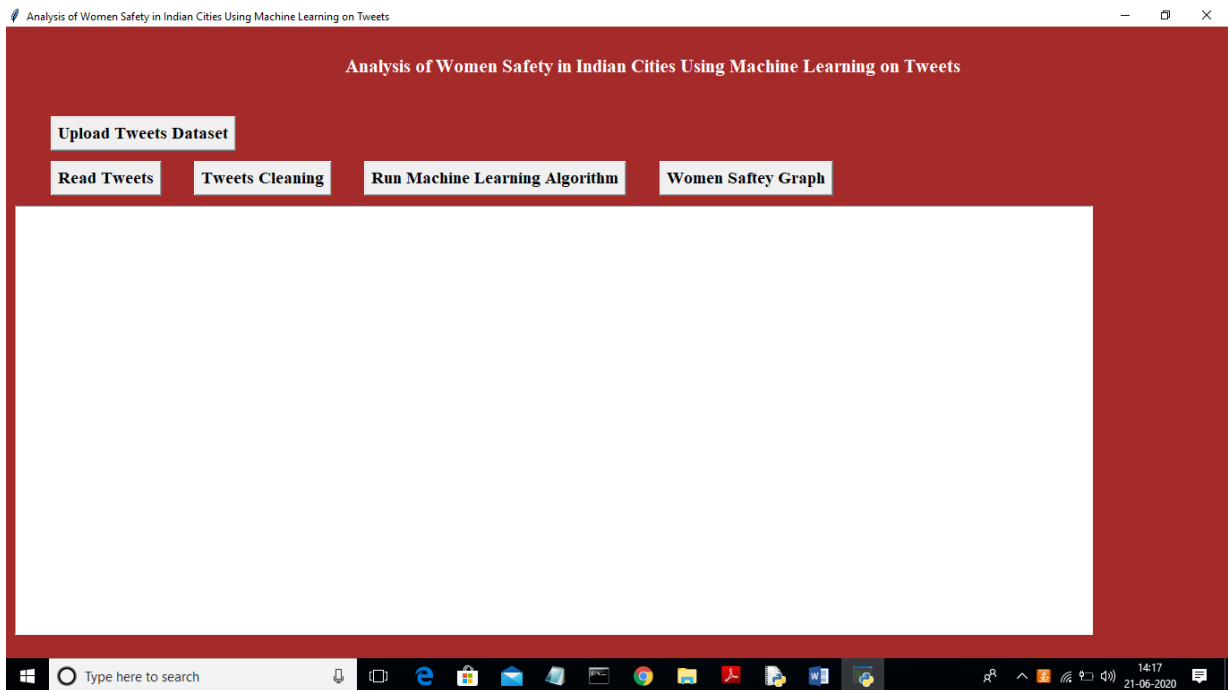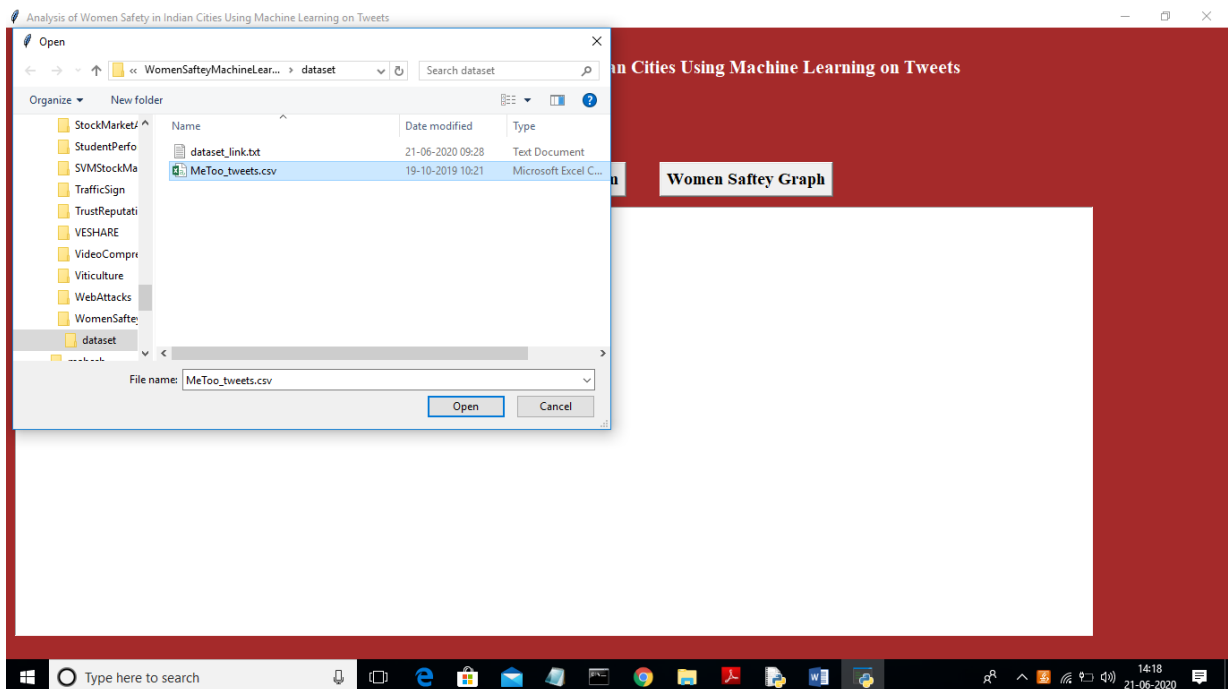
5. SCREENSHOTS

Screenshot 5.1: uploading tweet dataset.

Screenshot 5.2: Uploading sample raw data file.

CMRTC
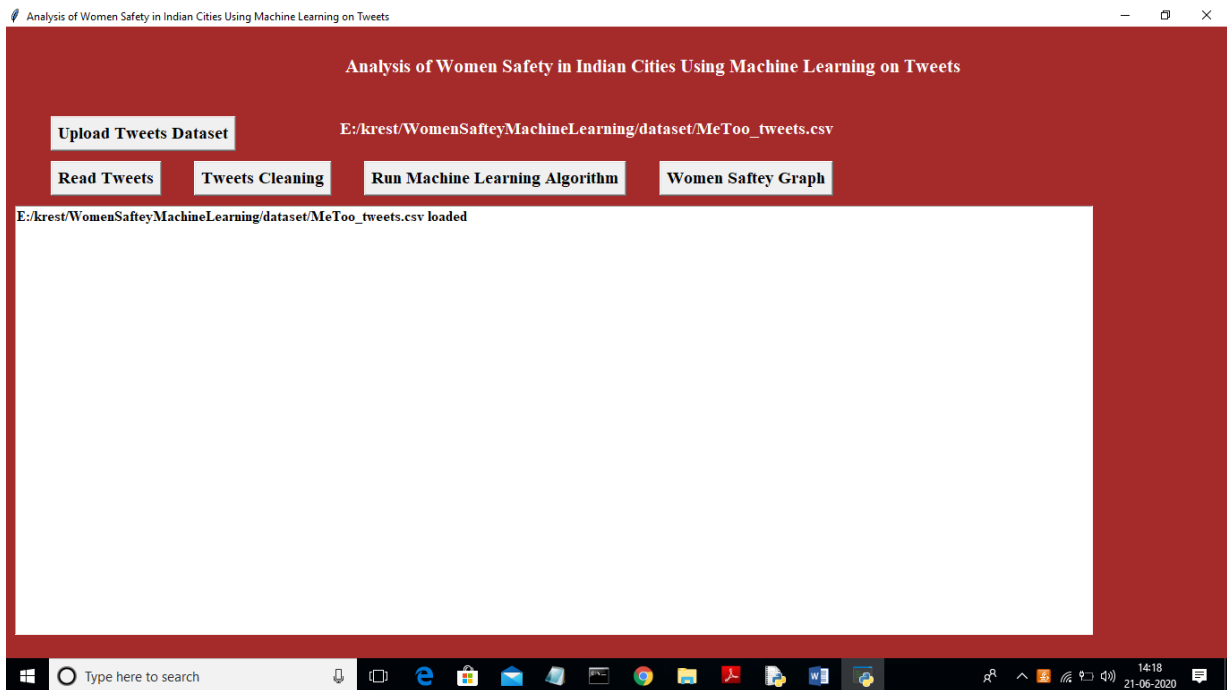
Screenshot 5.3: Uploaded file path name.
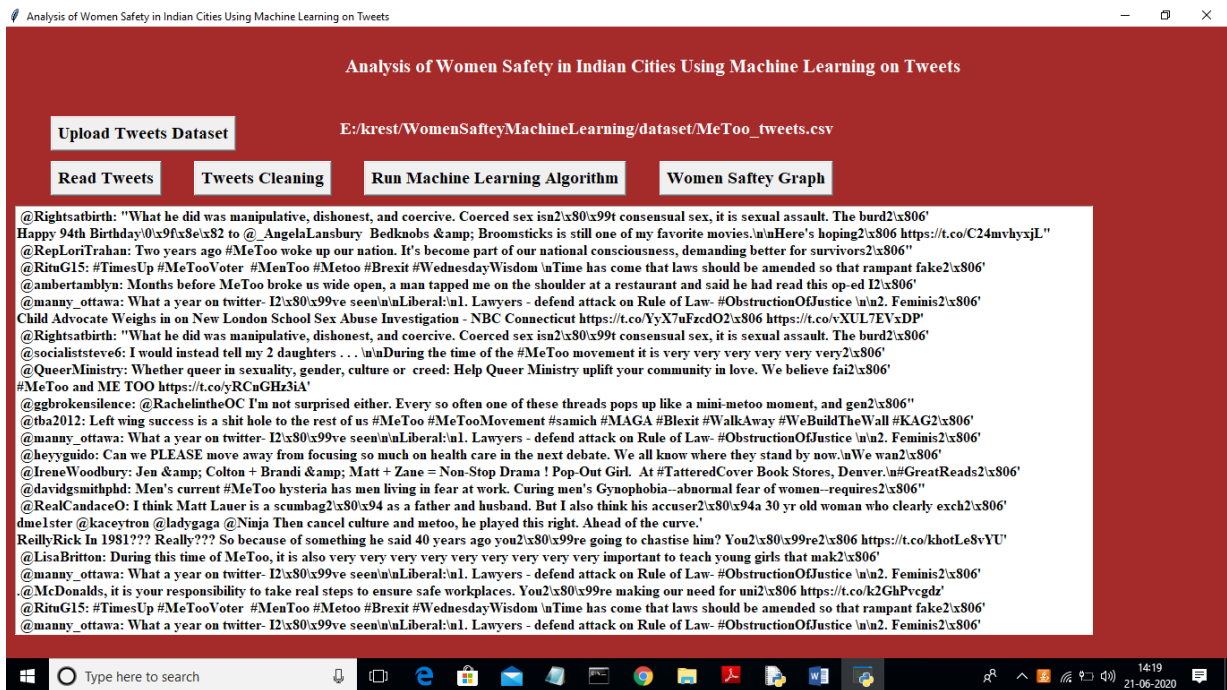
Screenshot 5.4: Reading data from the file.   18
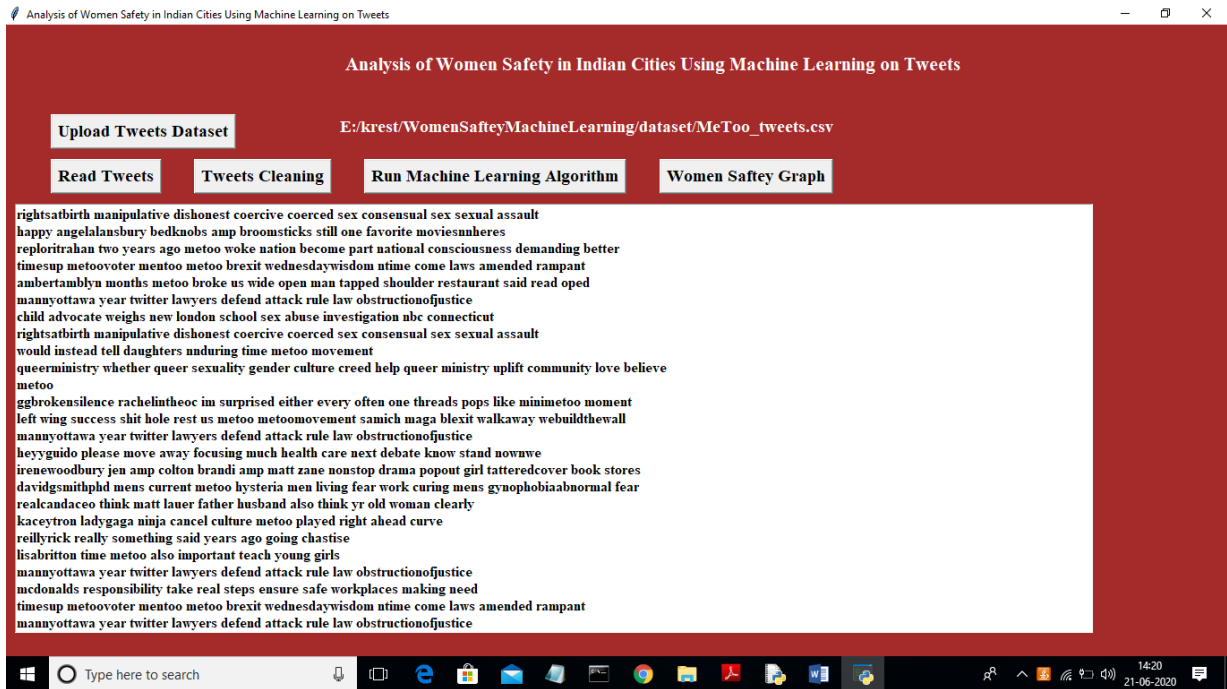
Screenshot 5.5:Cleaning of the raw data

CMRTC
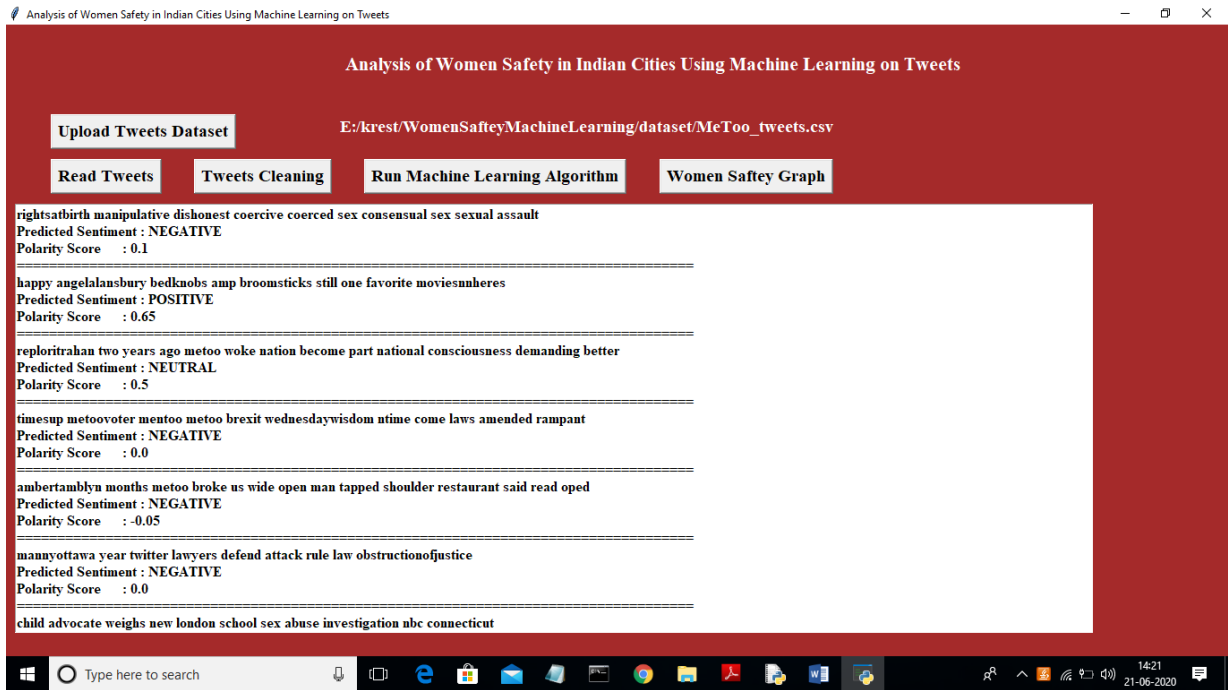
Screenshot 5.6: Applying machine learning algorithm
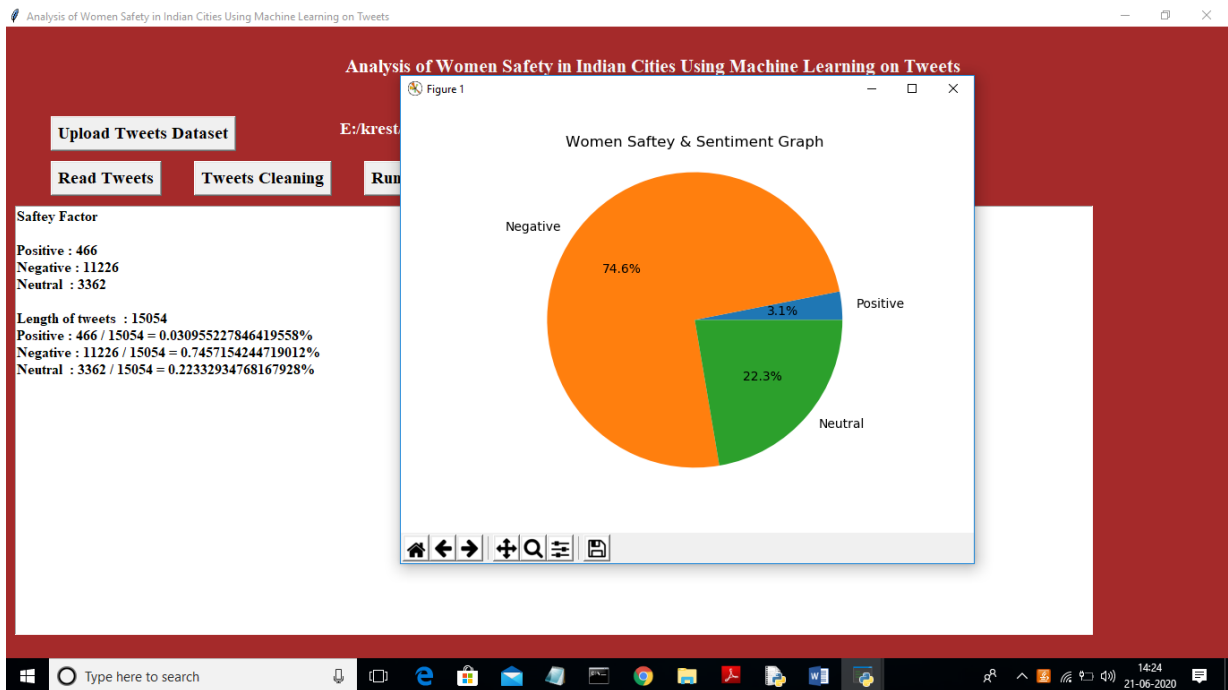
Screenshot 5.7:Final graph

6. TESTING

6. TESTING

6.1 INTRODUCTION TO TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

6.2 TYPES OF TESTING

6.2.1 UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .It is done after the completion of an individual unit before integration. This is a structural testing that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

21

6.2.2 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Integration tests demonstrate that although the components were individually satisfactory, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

6.2.3 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid   : identified classes of invalid input must Input be rejected.

Functions   : identified functions must be exercised.

Output   : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases.   22

CMRTC

6.3 TEST CASES
6.3.1 CLASSIFICATION

Test Case Id	Test Case Name	Test Case Description	Test Steps			Test Case Status	Test Priority
			Step	Expected	Actual		
01	Start the Application	Host the application and test if it starts making sure the required software is available	If it doesn't Start	We cannot run the application.	The application hosts success.	High	High
02	Home Page	Check the deployment environment for properly loading the application.	If it doesn't load.	We cannot access the application.	The application is running successfully.	High	High
03	User Mode	Verify the working of the application in freestyle mode	If it doesn't Respond	We cannot use the Freestyle mode.	The application displays the Freestyle Page	High	High
04	Data Input	Verify if the application takes input and updates	If it fails to take the input or store in The Database	We cannot proceed further	The application updates the input to application	High	High

7. CONCLUSION

7. CONCLUSION & FUTURE SCOPE

7.1 PROJECT CONCLUSION

By using this NLP(Natural language processing) machine will analyze the twitter data and get an idea about the status of women safety in Indian cities. The efficiency of this NLP is more when compared with other data analyzing techniques.The efficiency of this NLP is 60% to 80%. We will get the positive,negative and neutral percentages about the women status in our society in a bar graph.

7.2 FUTURE SCOPE

Throughout the project various algorithms have been discussed about deep learning and machine learning which can help in analyzing huge amount of data accumulated via tweeter to help determine the safety of women in the society. The machine learning algorithms used are very effective and work efficiently on various platforms when it comes to handling the large amount of data from social media platforms. These algorithms can really help make a dent in women safety and extracting information and create various datasets to work with. We look forward to work more and tweak it to work even more efficiently in the coming near future.

8. BIBLIOGRAPHY

8. BIBLIOGRAPHY

8.1 REFERENCES

[1]. Agarwal, Apoorv, FadiBiadsy, and Kathleen R. Mckeown. "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams."Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics.Association for Computational Linguistics, 2009.

[2]. Barbosa, Luciano, and JunlanFeng. "Robust sentiment detection on twitter from biased and noisy data." Proceedings of the 23rd international conference on computational linguistics: posters. Association for Computational Linguistics, 2010.

[3]. Bermingham, Adam, and Alan F. Smeaton. "Classifying sentiment in microblogs: is brevity an advantage?."Proceedings of the 19th ACM international conference on Information and knowledge management.ACM, 2010.

[4]. Gamon, Michael. "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis." Proceedings of the 20th international conference on Computational Linguistics.Association for Computational Linguistics, 2004.

[5]. Kim, Soo-Min, and Eduard Hovy. "Determining the sentiment of opinions."Proceedings of the 20th international conference on Computational Linguistics.Association for Computational Linguistics, 2004.

8.2 GITHUB LINK

https://github.com/sravya666/Face-recognition-audio-output/tree/master