A Project Report

on

ANALYSIS OF WOMEN SAFETY ON SOCIAL MEDIA

submitted in partial fulfilment of the requirements for the award of the degree of

MASTER OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING

by

20J41D5801 Ms. TALASILA ALEKHYA

under the esteemed guidance of

Dr A RAMA SWAMY REDDY

Professor

Principal of MREC

Department of Computer Science and Engineering

MALLA REDDY ENGINEERING COLLEGE

(Approved by AICTE, New Delhi and Affiliated to JNTUH, Hyderabad)

Campus 1, Dulapally Road Maisammaguda Post via. Kompally Rangareddy Dt Secunderabad, Hyderabad, Telangana 500100

May, 2022

DECLARATION

We hereby declare that the work presented in this project entitled **"ANALYSIS OF WOMEN SAFETY ON SOCIAL MEDIA"** submitted towards completion of Project Work in III year of M.Tech., CSE at 'MALLA REDDY ENGINEERING COLLEGE', Hyderabad is an authentic record of our original work carried out under the guidance of Dr. A RAMA SWAMY REDDY, Professor, Principal of MREC.

Sign. with date:

Ms. TALASILA ALEKHYA

(20J41D5801)

MALLA REDDY ENGINEERING COLLEGE

(Approved by AICTE, New Delhi and Affiliated to JNTUH, Hyderabad)

Campus 1, Dulapally Road Maisammaguda Post via. Kompally Rangareddy Dt Secunderabad, Hyderabad, Telangana 500100

Department of Computer Science and Engineering

Certificate

This is to certify that the Project Work report on "**ANALYSIS OF WOMEN SAFETY ON SOCIAL MEDIA**" is a bonafide work carried out by Ms. TALASILA ALEKHYA (20J41D5801) in the partial fulfilment for the award of M.Tech. degree in **Computer Science and Engineering, MALLA REDDY ENGINEERING COLLEGE, Maisammaguda, Hyderabad, Affiliated to JNTUH, Hyderabad** under my guidance and supervision. The results embodied in the project work have not been submitted to any other University or Institute for the award of any degree or diploma.

Head of the Department	**Guide**
Dr. N. Lakshmipathi Anantha	**Dr A RAMA SWAMY REDDY**
Professor and HOD,	**Professor**
Department of CSE	**Principal of MREC**

External Examiner

Acknowledgements

We would like to express our sincere thanks to **Dr. A RAMA SWAMY REDDY, Principal, Malla Reddy Engineering College,** for providing the working facilities in the college.

Our sincere thanks and gratitude to our **Dr. D Krishna Madhuri, Associate Professor,** Department of CSE, **Malla Reddy Engineering College**, for all the timely support and valuable suggestions during the period of our project.

We are extremely thankful and indebted to our internal guide, **Dr A RAMA SWAMY REDDY, Professor, Principal, Malla Reddy Engineering College**, for his constant guidance, encouragement and moral support throughout the project.

Finally, we would also like to thank our Project Coordinator, all the faculty and staff of CSE Department who helped us directly or indirectly, parents and friends for their cooperation in completing the project work.

Ms. TALASILA ALEKHYA

(20J41D5801)

ABSTRACT

Nowadays women are experiencing lots of violence such as harassment in places in several cities. This starts from stalking which then leads to abusive harassment or also called abuse assault. In this project we mainly focus on the role of social media which can be used to promote the safety of women in India, given the special reference to the participation of many social media websites or applications such as Twitter platform. This project also focuses on developing the responsibilities among the common people on the various parts of Indian cities so that the safety of women around them is ensured. So, this project aims at finding safety percentage of a city in the ideology of women safety by finding polarity of comments.

LIST OF FIGURES

Contents

1. INTRODUCTION

Twitter in this modern era has emerged as an ultimate microblogging social network consisting of over hundred million users and generate over five hundred million messages known as ‘Tweets’ every day. Twitter with such a massive audience has magnetized users to emit their perspective and judgmental about every existing issue and topic of internet, therefore twitter is an informative source for all the zones like institutions, companies and organizations. On twitter, users will share their opinions and perspective in the tweets section. This tweet can only contain 140 characters, thus making the users compact their messages with the help of abbreviations, slang, shot forms, emoticons, etc. In addition to this, many people express their opinions by using polysemy and sarcasm also. Hence twitter language can be termed as the unstructured. From the tweet, the sentiment behind the message is extracted. This extraction is done by using the sentimental analysis procedure. Results of the sentimental analysis can be used in many areas like sentiments regarding a particular brand or release of a product, analyzing public opinions on the government policies, people thoughts on women, etc. In order to perform classification of tweets and analyze the outcome, a lot of study has been done on the data obtained by the twitter. We also review some studies on machine learning in this paper and research on how to perform sentimental analysis using that domain on twitter data. The paper scope is restricted to machine learning algorithms and models. Staring at women and passing comments can be certain types of violence and harassments and these practices, which are unacceptable, are usually normal especially on the part of urban life. Many researches that have been conducted in India shows that women have reported sexual harassment and other practices as stated above. Such studies have also shown that in popular metropolitan cities like Delhi, Pune, Chennai and Mumbai, most women feel they are unsafe when surrounded by unknown people. On social media, people can freely express what they feel about the Indian politics, society and many other thoughts. Similarly, women can also share their experiences if they have faced any violence or sexual harassment and this brings innocent people together in order to stand up against such incidents. From the analysis of tweets text collection obtained by the twitter, it includes names of people who has harassed the women and also names of women or innocent people who have stood against such violent acts or unethical behavior of men and thus making them uncomfortable to walk freely in public.

1.1 Objective

This project is to analyze 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 analyzing those messages we can detect which area is more unsafe for women's.

1.2 Methodology

In the Sentiment Analysis the following steps are major to identify the positive, negative or neutral of the twitter post. They are:

i. Collecting the Dataset.

ii. Pre-Processing the Dataset.

iii. Feature Extraction.

iv. Apply Classifier.

i. Data Collection: The data is gathered from twitter using API. Application program interface (API) is utilized to gather information. Twitter website is a source which consists of user's tweets.

ii. Data Preparation: In Data Pre-processing removing noisy, unrelated data, inconsistent and incomplete data from the dataset. Generally, in twitter we have to remove URLs, special characters, retweets, hash tags.

iii. Feature Extraction: In this work, we used Bag of Words to extract features from text documents. After extraction, these features are used for training machine learning algorithms. It makes a jargon of the apparent multitude of novel words happening in all the reports in the preparation set. Bag of words features containing term frequencies of each word in each document, i.e., the number of occurrence and not sequence or order of words matters. This can be done by Count Vectorizer method in Python.

iv. Classification: A classification problem is applied if the output variable is a label or category, such as "Rainy" or "Sunny" or "disease" and "no disease" or in our work "Positive" or "Negative".

1.3 Dataset

• The dataset we are using is “MeToo Tweets.csv”

• There are 15k rows and 3 columns.

• This dataset includes many fields about each tweet like Text, Favorite_count and Retweet_ count.

https://www.kaggle.com/hollyhetherington/metootweets?select=MeToo_tweets.csv

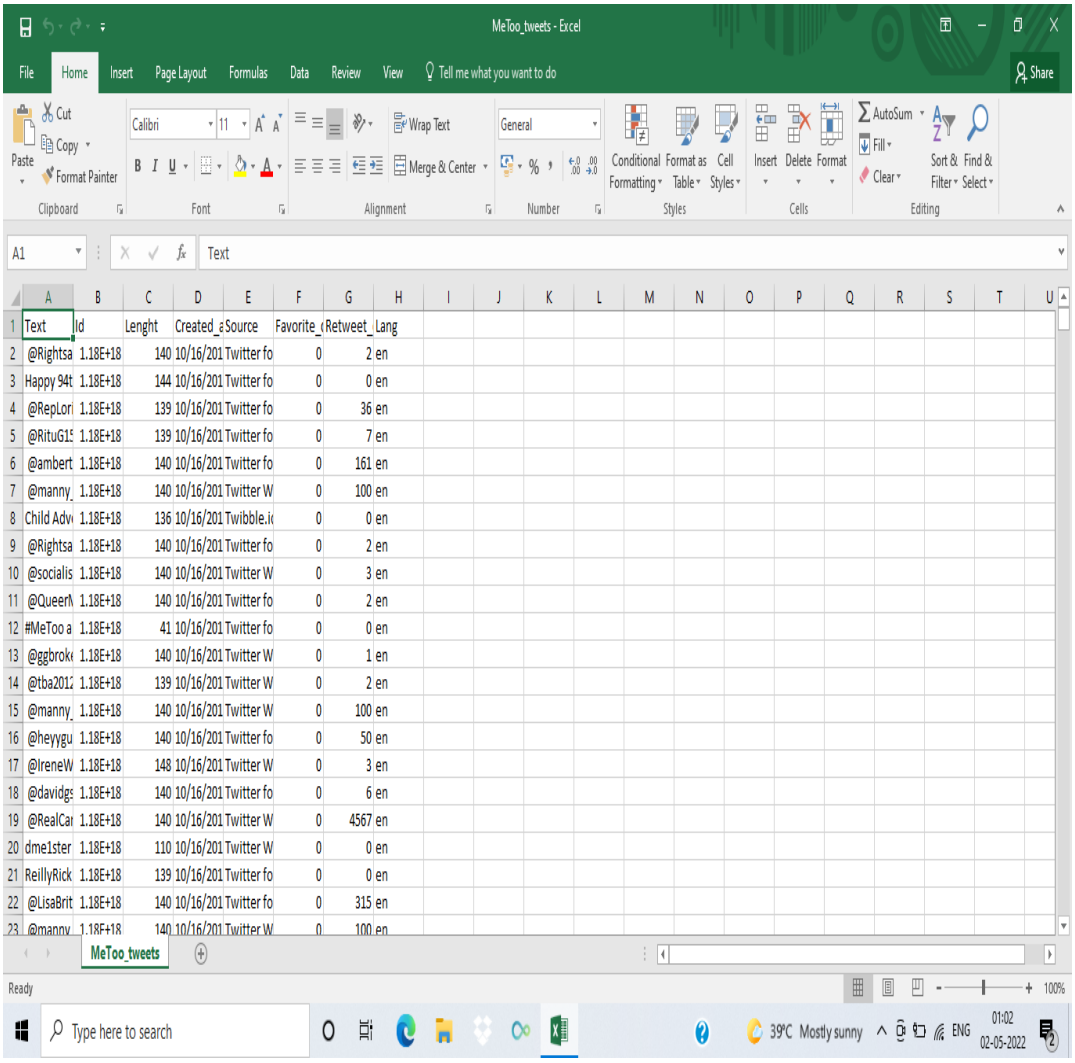

Fig 1: Dataset

2. Theoretical Analysis of the proposed project

2.1 Requirements Gathering

2.1.1 Software Requirements

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation. The appropriation of requirements and implementation constraints gives the general overview of the project in regards to what the areas of strength and deficit are and how to tackle them.

- Python idle 3.7 version (or)
- Anaconda 3.7 (or)
- Jupiter(or)
- Google Collaboratory

2.1.2 Hardware Requirements

Minimum hardware requirements are very dependent on the particular software being developed by a given Enthought Python / Canopy / VS Code user. Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

- Operating system : Windows, Linux
- Processor : Minimum intel i3
- Ram : Minimum 4gb
- Hard disk : Minimum 250gb

3. DESIGN

3.1 Introduction

Software design sits at the technical kernel of the software engineering process and is applied regardless of the development paradigm and area of application. Design is the first step in the development phase for any engineered

product or system. The designer's goal is to produce a model or representation of an entity that will later be built. Beginning, once system requirements have been specified and analyzed, system design is the first of the three technical activities -design, code and test that is required to build and verify software.

The importance can be stated with a single word "Quality". Design is the place where quality is fostered in software development. Design provides us with representations of software that can assess quality. Design is the only way that we can accurately translate a customer's view into a finished software product or system. Software design serves as a foundation for all the software engineering steps that follow. Without a strong design we risk building an unstable system – one that will be difficult to test, one whose quality cannot be assessed until the last stage.

During design, progressive refinement of data structure, program structure, and procedural details are developed, reviewed and documented. System design can be viewed from either technical or project management perspective. From the technical point of view, design is comprised of four activities–architectural design, data structure design, interface design and procedural design.

3.2 Architecture Diagram

Architecture diagram is a graphical representation of a set of concepts that are part of an architecture, including their principles, elements and components.

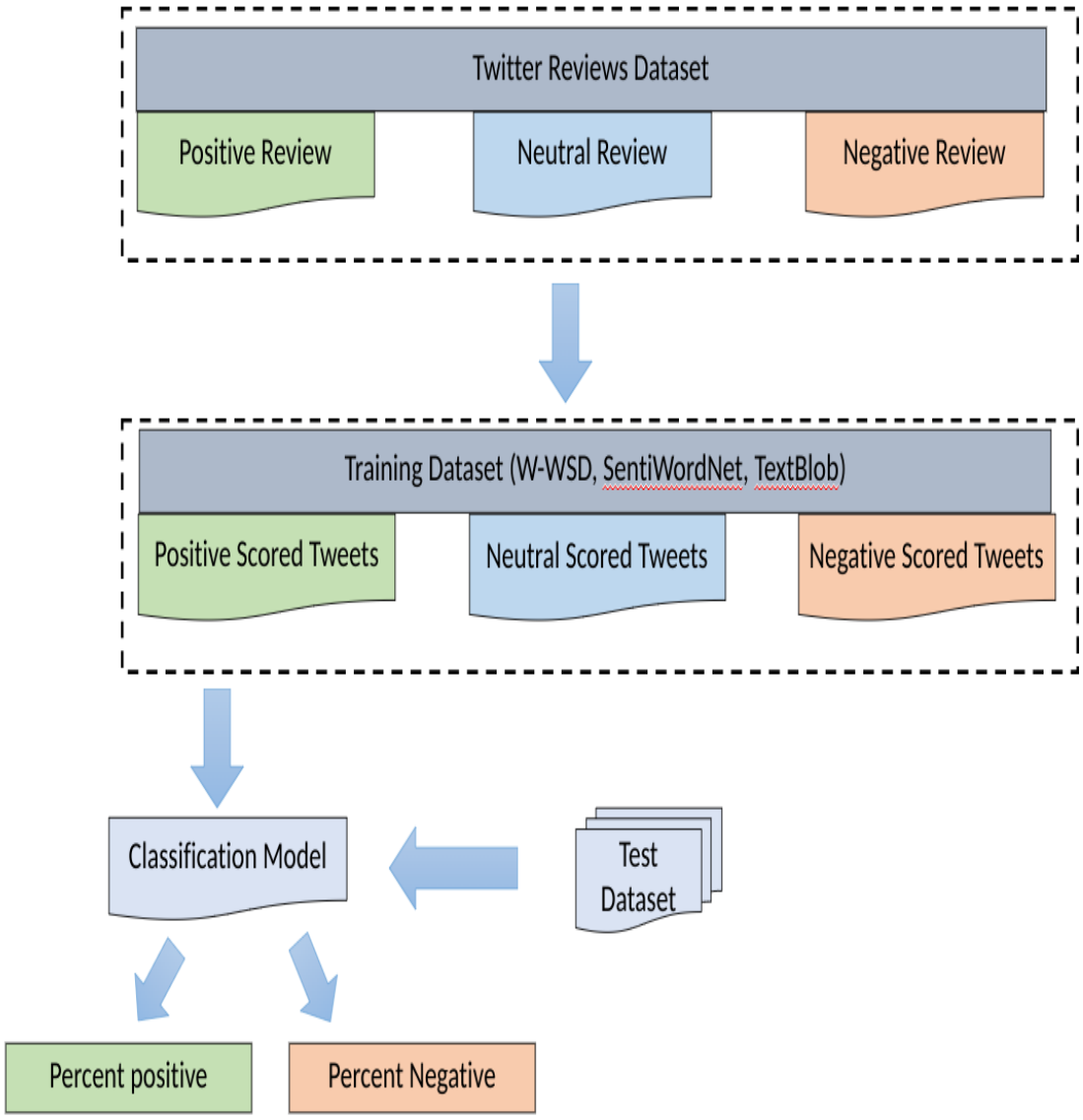

Fig 2: Architecture Diagram

3.3 Block Diagram

A block diagram is a diagram of a system in which the principal parts or functions are represented by blocks connected by lines that show the relationships of the blocks. They are heavily used in engineering hardware design, electronic design, software design, and process flow diagrams.

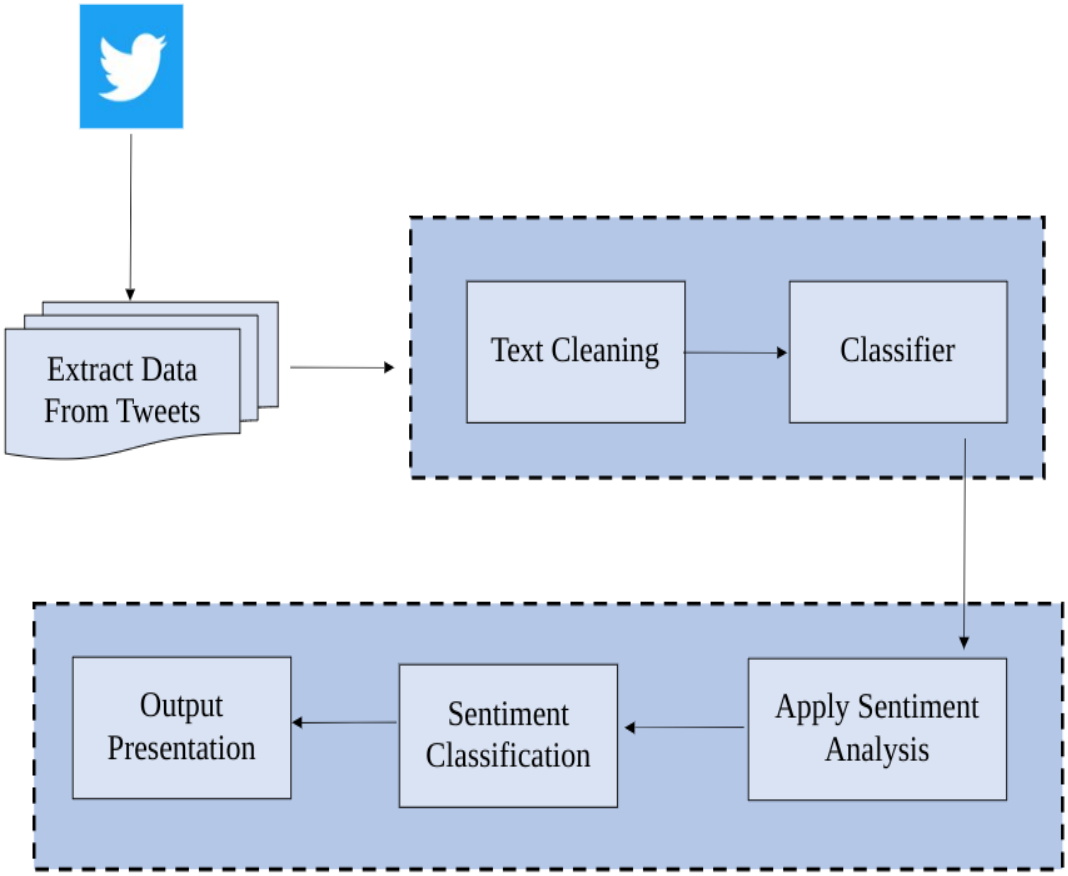

Fig 3: Block Diagram

1) INFORMATION EXTRACTION: This aides in separating the tweet message from the interpersonal organization stage and this messages additionally additional information like tweets likes, abhorrence's and remarks.

2) TEXT CLEANING: Post the information extraction the data must be passed to the classifier. The classifier Clean tweets by eliminating the stop words, non-text-based substance and commotion like monotonous letters.

3) SENTIMENT ANALYSIS: Once the classifier cleans the dataset, the information is prepared for the nostalgic investigation measure. AI approaches includes in Training the dataset and afterward testing that prepared dataset. Prepared and tried information are helpful for the classifier to play out the calculation. Greatest Entropy, Navies Bayes order, Support Vector Machine and Bayesian Networks are a portion of the calculation

which can be utilized to prepare the classifier. Testing information which is utilized to distinguish the effectiveness of the slant classifier.

4) SENTIMENT CLASSIFICATION: In this progression, the dataset is prepared for the arrangement. Every single sentence of the tweet will be analyzed and assessment will be framed appropriately, for subjectivity. Emotional articulation sentences are held are characterized into acceptable, terrible or like, abhorrence or positive and negative. For target articulation sentences are dismissed.

5) OUTPUT PRESENTATION: Once the calculation is finished, the yield of the investigation can be envisioned by making various sorts of charts. Time arrangement, Bar diagrams and Pie outlines are a portion of the models which can be utilized to show the yield. To quantify the supposition of the tweets regarding Positive, Negative and Neutral Bar charts can be utilized. Likewise, to gauge regarding likes, disdains, normal length of tweet for a specific period, Time arrangement can be utilized. To acquire the underlying wellspring of the tweet, pie outlines can be utilized.

3.4 UML Diagrams

UML, short for Unified Modeling Language, is a standardized modeling language consisting of an integrated set of diagrams, developed to help system and software developers for specifying, visualizing, constructing, and documenting the artifacts of software systems, as well as for business modeling and other non-software systems. For example, use case diagrams, activity diagrams, ER diagrams, Sequence diagrams, class diagrams etc.

3.4.1 Use Case Diagram

To model a system, the most important aspect is to capture the dynamic behavior. Dynamic behavior means the behavior of the system when it is running/operating. Only static behavior is not sufficient to model a system rather dynamic behavior is more important than static behavior. In UML, there are five diagrams available to model the dynamic nature and use case diagram is one of them. Now as we have to discuss that the use case diagram is dynamic in nature, there should be some internal or external factors for making the interaction.

These internal and external agents are known as actors. Use case diagrams consist of actors, use cases and their relationships. The diagram is used to model the

system/subsystem of an application. A single use case diagram captures a particular functionality of a system. Hence to model the entire system, a number of use case diagrams are used.

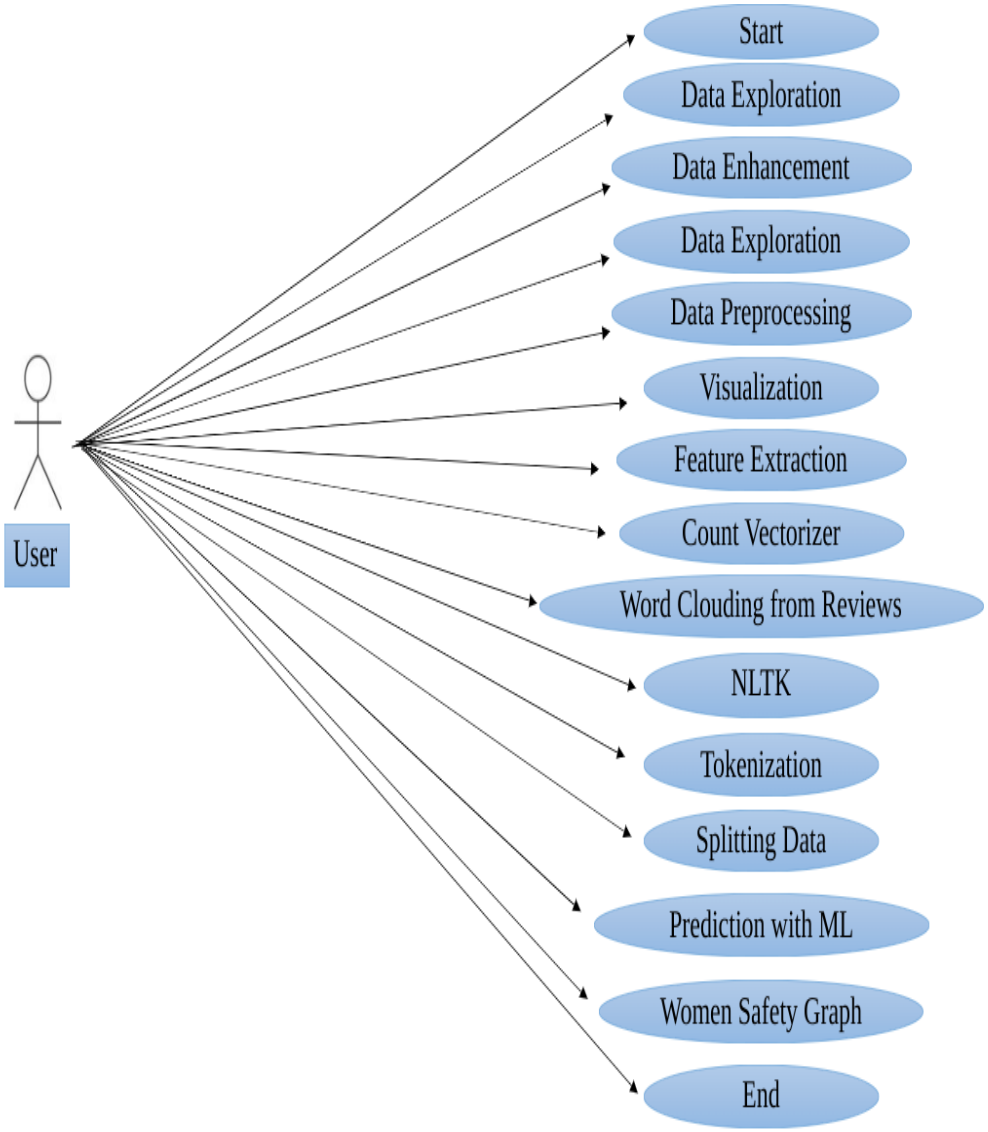

Fig 4: Use Case Diagram

3.4.2 Sequence Diagram

Sequence Diagrams represent the objects participating in the interaction horizontally and time vertically. A Use Case is a kind of behavioral classifier that represents a declaration of an offered behavior. Each use case specifies some behavior, possibly including variants that the subject can perform in collaboration with one or more actors. Use cases define the offered behavior of the subject without reference to its internal

structure. These behaviors, involving interactions between the actor and the subject, may result in changes to the state of the subject and communications with its environment. A use case can include possible variations of its basic behavior, including exceptional behavior and error handling.

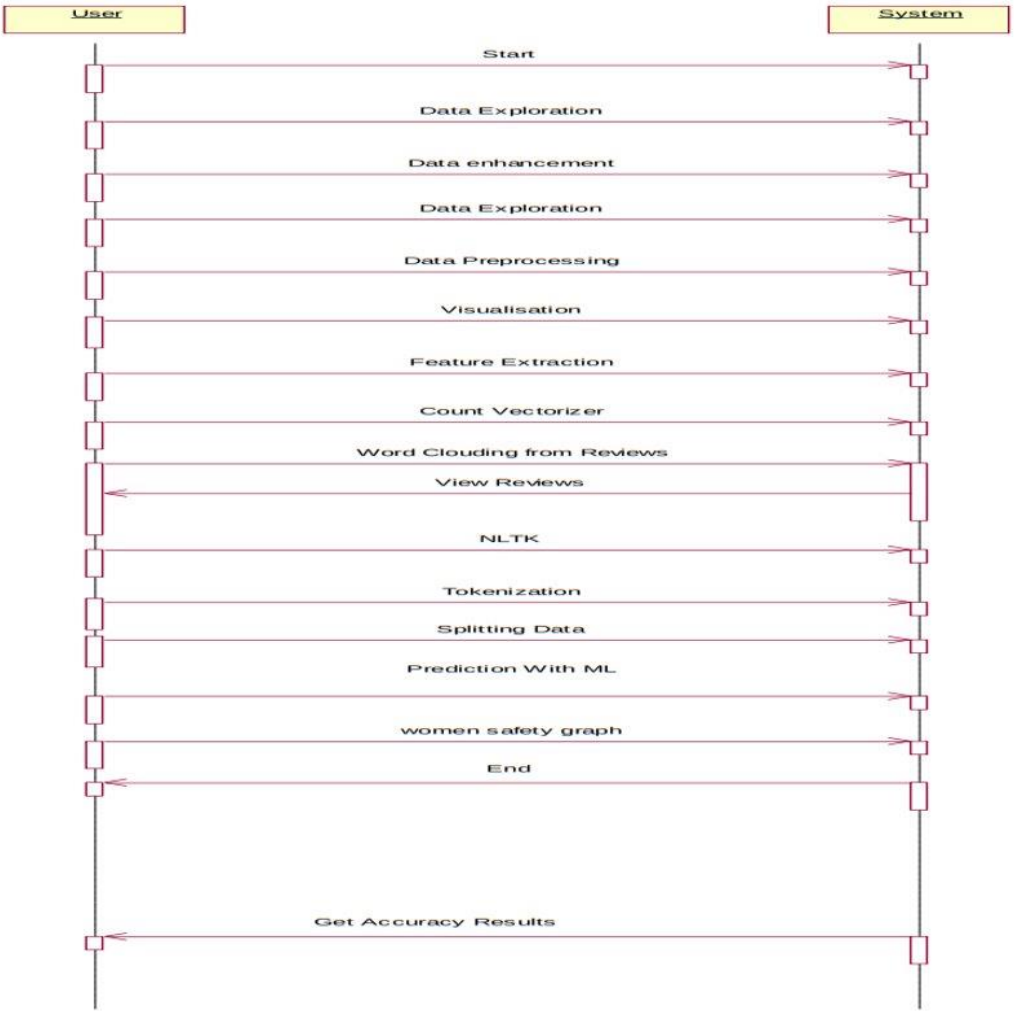

Fig 5: Sequence Diagram

3.4.3 Class Diagram

The class diagram is the main building block of object-oriented modelling. It is used for general conceptual modelling of the systematic of the application, and for detailed modelling translating the models into programming code. Class diagrams can also be

used for data modelling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.

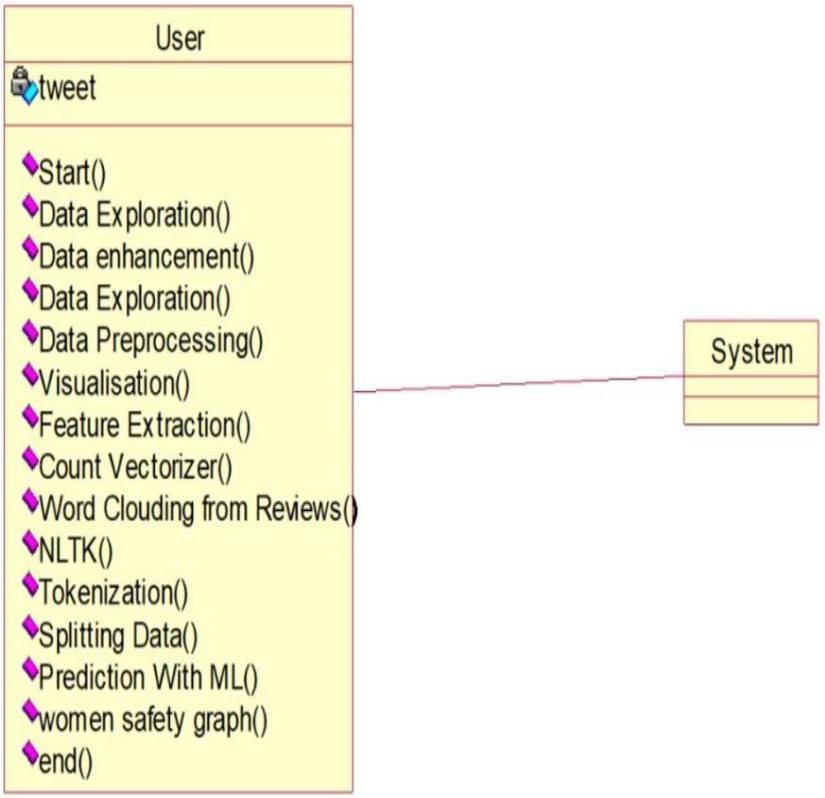

Fig 6: Class Diagram

4. IMPLEMENTATION

4.1 Coding

```
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 On Social Media")

main.geometry("1200x1200")

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)

return tokens

def upload():

global filename

filename = filedialog.askopenfilename(initialdir="dataset")
```

```
text.delete('1.0', END)

text.insert(END, filename + " loaded\n");

def read():

tweets_list.clear()

train = pd.read_csv(filename, encoding='iso-8859-1')

for i in range(len(train)):

tweet = train.get_value(i, 'Text')

tweets_list.append(tweet)

text.insert(END, tweet + "\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")

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 t

ext.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')

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 bar():

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.graph(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 on Social Media')

title.config(bg=blue, fg='white')

title.config(font=font)

title.config(height=3, width=120)

title.place(x=0, y=5)

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

uploadButton = Button(main, text="Upload Tweets Dataset", command=upload)

uploadButton.place(x=50, y=100)

uploadButton.config(font=font1)

readButton = Button(main, text="Read Tweets", command=read)

readButton.place(x=50, y=150)

readButton.config(font=font1)

text = Text(main, height=25, width=150)

scroll = Scrollbar(text)

text.configure(yscrollcommand=scroll.set)

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

cleanButton.config(font=font1)

text = Text(main, height=25, width=150) scroll = Scrollbar(text) text.configure(yscrollcommand=scroll.set)

mlButton = Button(main, text="Run Machine Learning Algorithm", 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='purple')

main.mainloop()

4.2 Test cases

Test Case ID	Test Scenario	Expected Result	Actual Result	Pass/Fail
TC01	Check whether the application is working	Output page should be opened	Has excepted	Pass
TC02	Check whether the upload button is working	Dataset has to accessible	Has excepted	Pass
TC03	Check whether dataset is getting uploaded	Dataset has to uploaded	Has excepted	Pass
TC04	Check whether the read tweets is working	Tweets has to be read	Has excepted	Pass
TC05	Check whether the clean tweets is working	Tweets has to be cleaned	Has excepted	Pass
TC06	Check whether the ML algorithms are running	Polarity has to be calculated	Has excepted	Pass
TC07	Check whether getting the graph	Graph has to displayed	Has excepted	Pass

Fig 7: Test cases

4.3. Output Screenshots

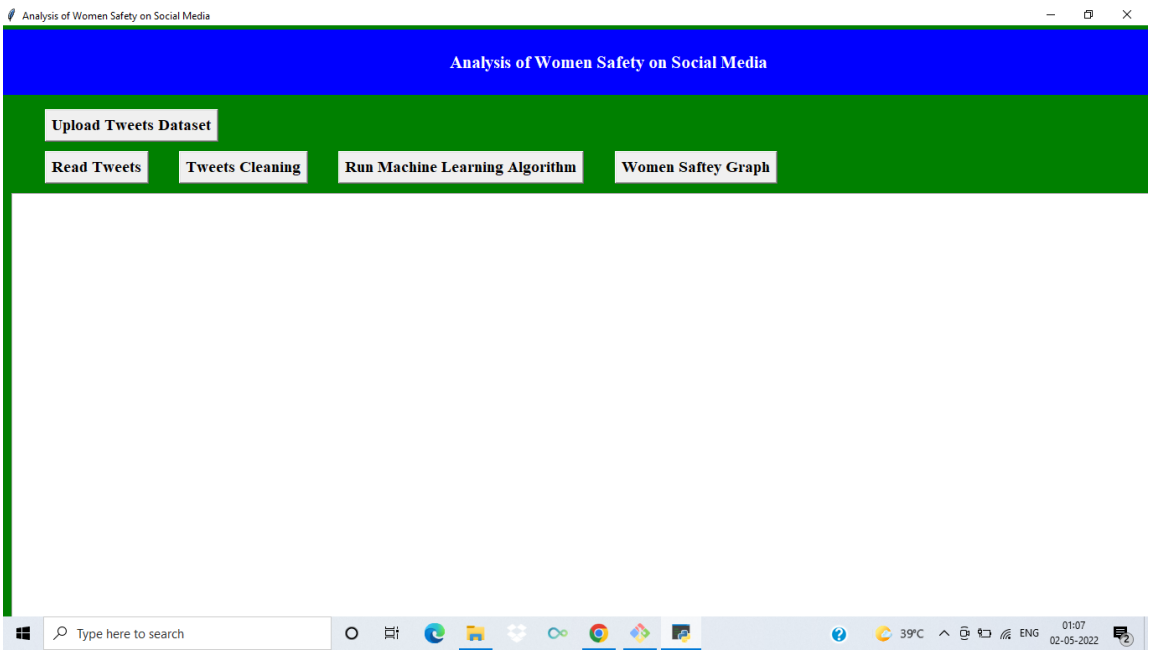

Fig 8: Output Window

In above screen click on 'Upload Tweets Dataset' button and upload tweets

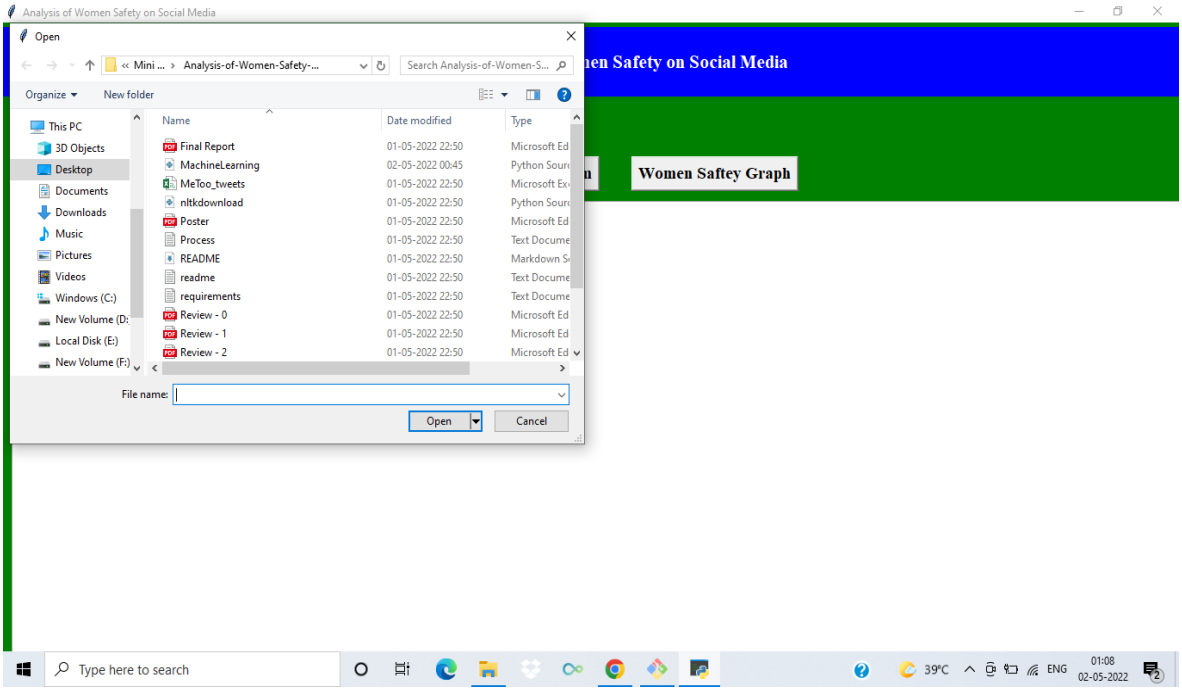

Fig 8: Uploading Dataset

In above screen uploading MeeToo_tweets.csv file and then click on 'Open' button to load dataset and to get below screen

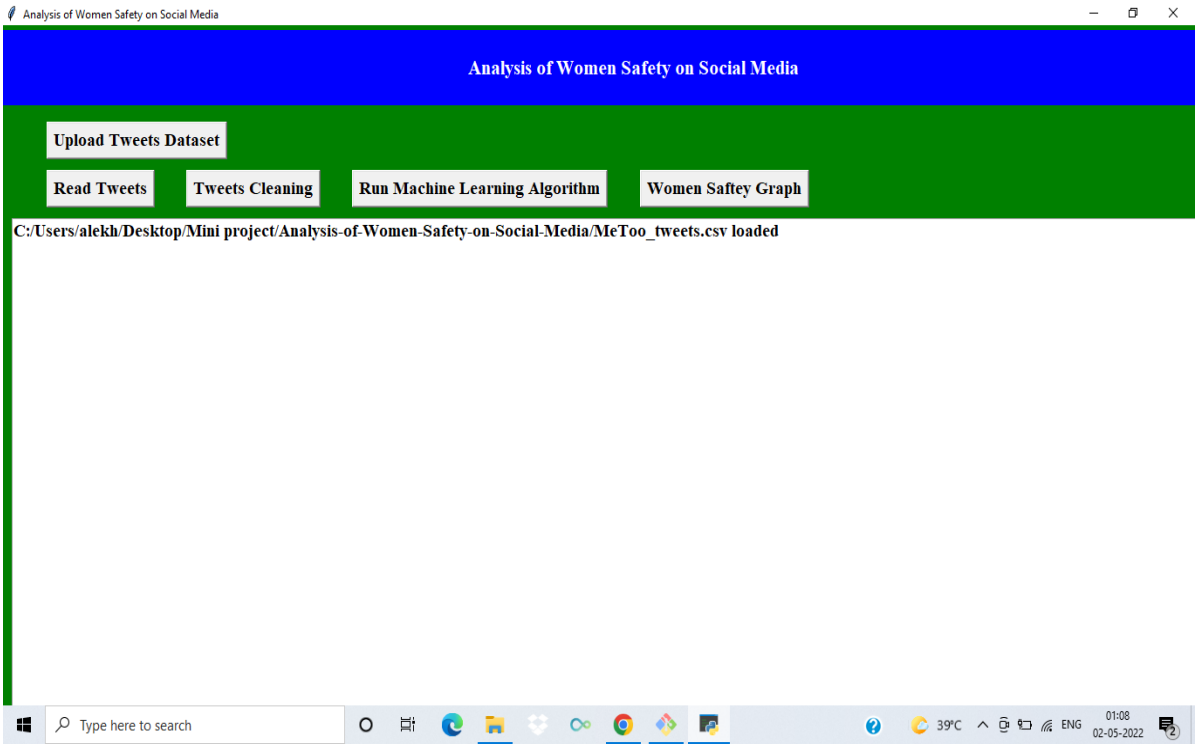

Fig 9: Dataset Uploaded

In the above screen tweets dataset loaded and now click on 'Read Tweets' button to read tweets from dataset.

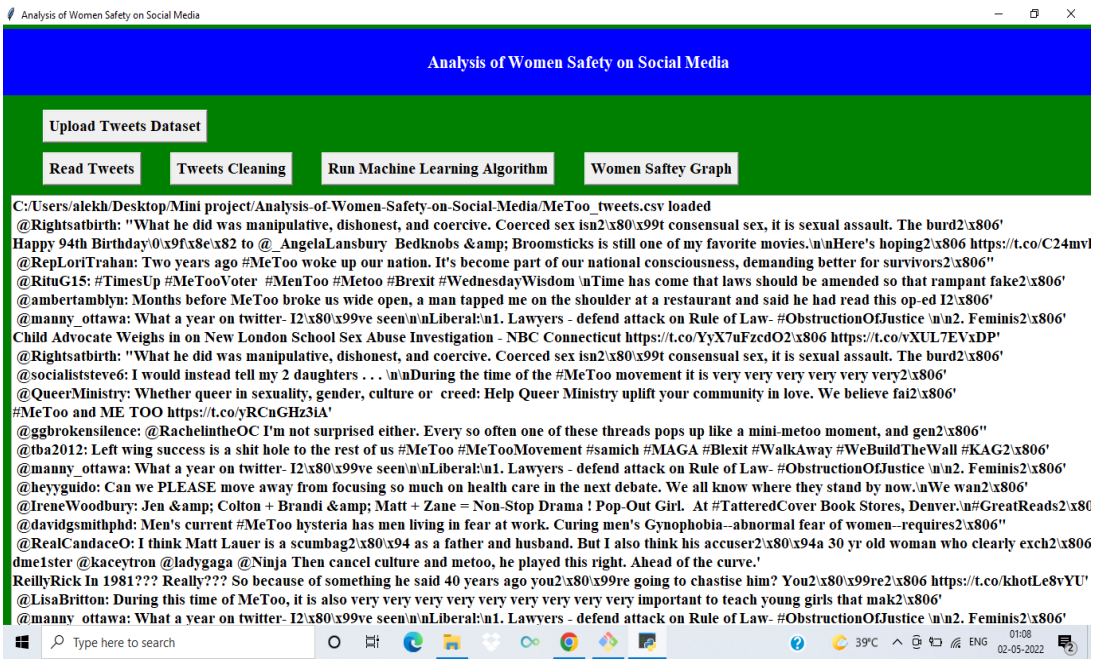

Fig 10: Reading Tweets

In the above screen each line represents one tweet and you can scroll down above screen text area to view all tweets. In the above screen we can see all tweets containing special symbols and stop words and to clean those tweets click on ‘Tweets Cleaning’ button.

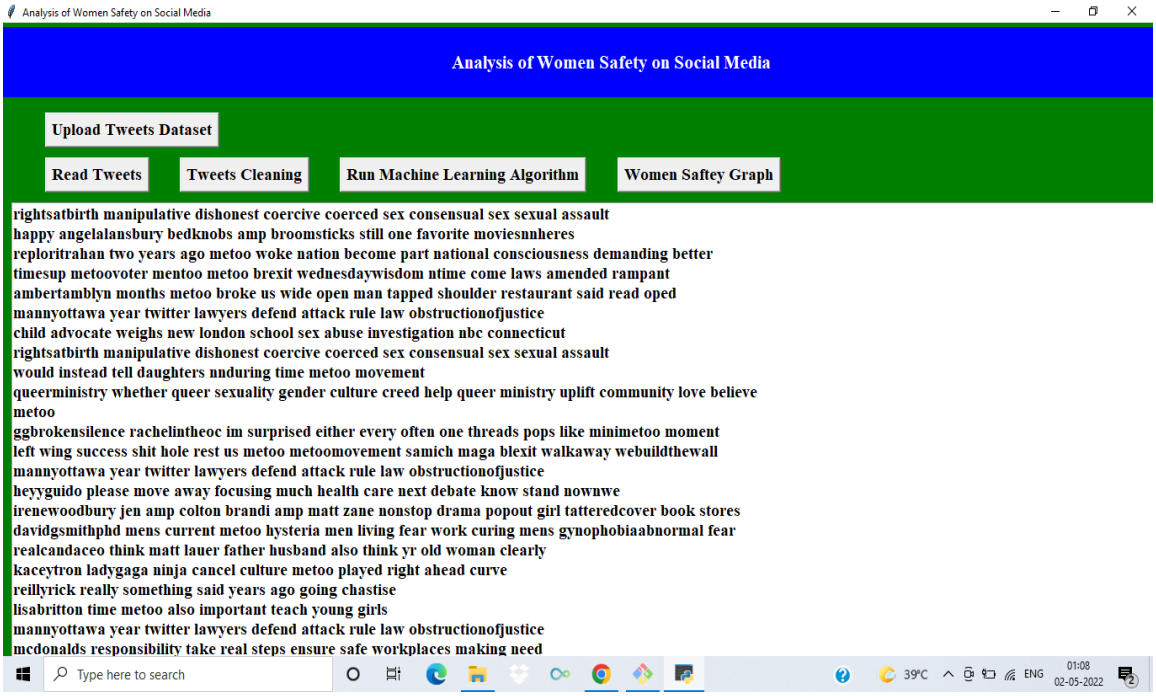

Fig 11: Cleaning Tweets

In the above screen we can see all special symbols and stop words removed from tweets and only clean words are there and now click on ‘Run Machine Learning Algorithm’ button to predict sentiments from tweets.

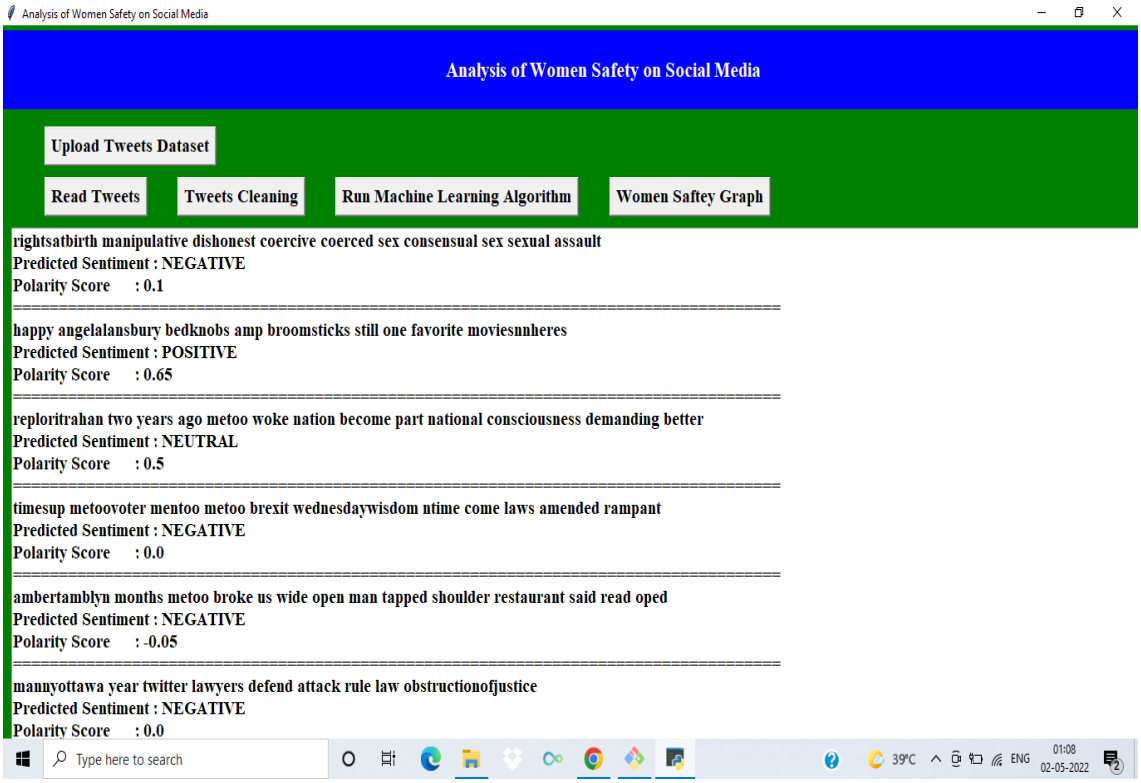

Fig 12: Running ML Algorithm

In the above screen each tweet has tweet text and then displays tweets sentiments with polarity score. Scroll down above text area to see all tweets.

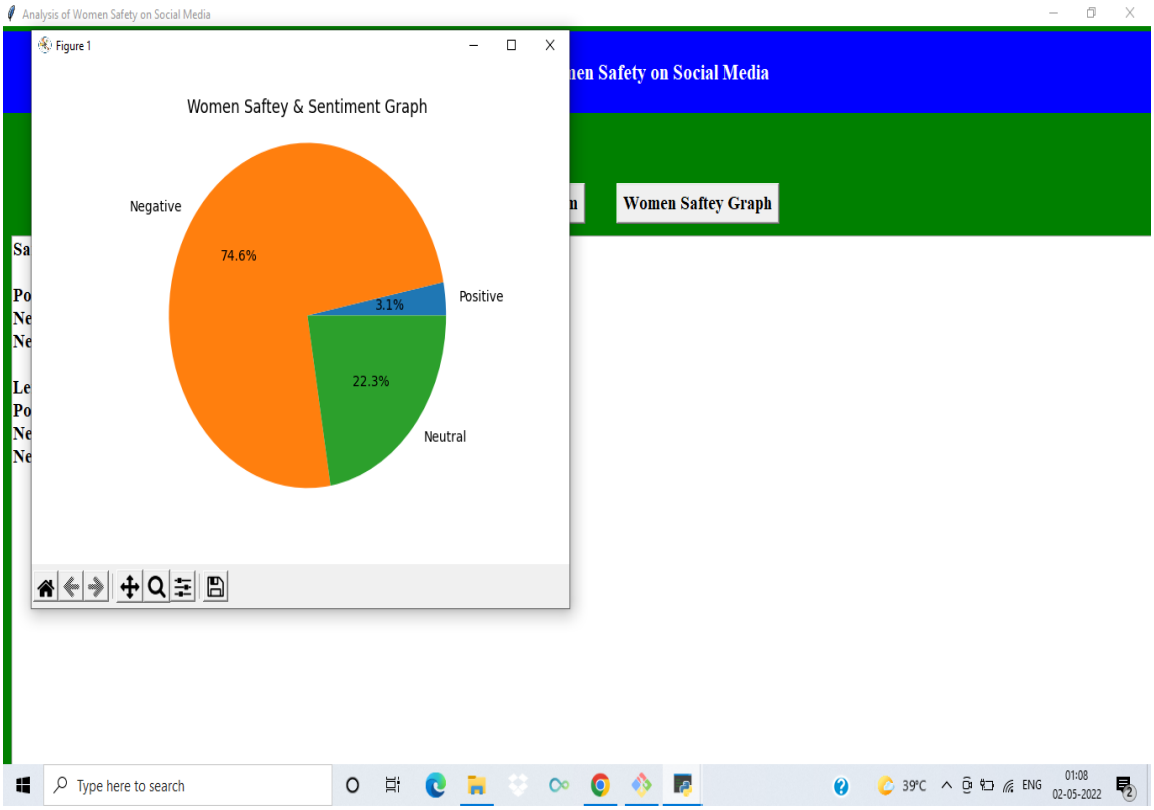

Fig 14: Implementing Pie chart

Now click on 'Women Safety Graph' button to get below results and by seeing that result user can easily understand whether area is safe or not. If the area is safe then more people will express either positive or neutral tweets and if not safe then more people will discuss negative tweets.

5. CONCLUSION AND FUTURE SCOPE

AI calculation has been examined all through the venture. For the twitter information that incorporates a huge number of tweets and messages each day, AI calculation assists with getting sorted out and perform investigation. SPC calculation and direct math are a portion of the calculations which are powerful in dissecting the huge amount of information that give arrangement and convert into significant datasets. Subsequently we can perform AI calculations to accomplish nostalgic examination and carry more security to women by spreading the mindfulness. For the future improvement, we can reach out to apply these AI calculations on various web-based media stages like Facebook and Instagram additionally since in our undertaking just

twitter is thought of. The present philosophy which is proposed can be incorporated with the twitter application interface to arrive at a bigger degree and apply wistful investigation on huge number of tweets to give more wellbeing.

6. REFERENCES

[1] Deepak Kumar1, Shivani Aggarwal2 "Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets", (IEEE), 2019. [2] Teja KR, Kumar KA, Praveen GS, Harini DN. Analysis of Crimes Against Women in India Using Machine Learning Techniques. In Communication Software and Networks 2021 (pp. 499-510). Springer, Singapore. [3] Vikram Chandra1, Rampur Srinath2 "Analysis of Women Safety using Machine Learning on Tweets", (IRJET) 2020. [4] P. Anbumani and R. Dhanapal, "Review on Privacy Preservation Methods in Data Mining Based on Fuzzy Based Techniques," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). [5] Reyes-Menendez A, Saura JR, AlvarezAlons C. Understanding# World Environment Day user opinions in Twitter: A topic-based sentiment analysis approach. International journal of environmental research and public health. 2018 Nov;15(11):2537. [6] Gupta B, Negi M, Vishwakarma K, Rawat G & Badhani P (2017). "Research of Twitter sentiment evaluation making use of system mastering algorithms on Python." International Journal of Computer System Applications, one hundred sixty-five(nine) 0975-8887. [7] Mamgain N, Mehta E, Mittal A & Bhatt G (2016, March). "Belief evaluation of top schools in India making use of Twitter data." In Computational Strategies, in Details and Interaction Technologies (ICCTICT). [8] Bakshi RK, Kaur N, Kaur R, Kaur G. Opinion mining and sentiment analysis. In2016 3rd international conference on computing for sustainable global development (INDIACom) 2016 Mar 16 (pp. 452-455). IEEE.