Government of Karnataka
Department of Collegiate Education
Dr. G. Shankar Government Women's First Grade College & PG Study Centre, Ajjarkadu, Udupi

A Project Report on

"Sentimental Analysis of women safety in India on Instagram data using machine learning"

Submitted To:

Department of Computer Science
Dr. G. Shankar Government Women's First Grade College
and Post Graduate Study Centre, Ajjarkadu, Udupi

Submitted by

Ms. APOORVA U05DG21S0108
Ms. KAVYASHRI U05DG21S0154
Ms. PAVITHRA B S U05DD21S0133
Ms. RUMANA U05DG21S0156

Under the Guidance of

Mrs. Sushma S
Department of Computer-Science
Dr. G. Shankar Government Women's First Grade College
Post-Graduate Study Centre, Ajjarkadu, Udupi

Dr. G. Shankar Government Women's First Grade College and Post-Graduate Study Centre Ajjarkadu, Udupi

"Sentimental Analysis of women safety in India on Instagram data using machine learning"

Submitted To

Department of Computer Science
Dr. G. Shankar Government Women's First Grade College
and Post Graduate Study Centre, Ajjarkadu, Udupi

Submitted by

Ms. APOORVA **U05DG21S0108**

Ms. KAVYASHRI **U05DG21S0154**

Ms. PAVITHRA B S **U05DD21S0133**

Ms. RUMANA **U05DG21S0156**

Project Supervisor

Mrs. Sushma S
Department of Computer-Science
Dr. G. Shankar Government Women's First Grade College
Post-Graduate Study Centre, Ajjarkadu, Udupi

Government of Karnataka
Department of Collegiate Education
Dr. G. Shankar Government Women's First Grade College
& PG Study Centre, Ajjarkadu, Udupi

CERTIFICATE

This is to certify that the project report entitled **"Sentimental Analysis of women safety in India on Instagram data using machine learning"** is an authenticated record of the project work carried out by Ms. Apoorva (Reg.No. U05DG21S0108), Ms. Pavithra B S (Reg.No. U05DG21S0133), and Ms. Kavyashri (Reg.No.U05DG21S0154), and Ms. Rumana (Reg.No.U05DG21S0155)of **III BCA VI Semester** in partial fulfillment of the requirement for the award of the degree of **Bachelor of Computer Applications** of Mangalore University during the year 2023-2024.

Forwarded to Principal for Approval
Lecturer:

Mrs. Sushma.S
Project Guide's Name

Approved and Forwarded to Mangalore University

Place: UDUPI
Date:

Dr. Manibushan D souza
Head of Department

Prof. Bhaskar Shetty S
Principal

Examiners

Name	**1.**
Signature	**1.**
Name	**2.**
Signature	**2.**

DECLARATION

We hereby declare that the project report titled as “Sentimental Analysis of women safety in India on Instagram data using machine learning " has been prepared by us during the year 2023-2024 under the valuable guidance and supervision of Mrs. Sushma. S Assistant professor and project guide at Dr. G Shankar Government Women's First Grade College & P.G Study Centre Ajjarkadu, Udupi in partial fulfillment of the requirement for the award of Bachelor's degree in computer application from University of Mangalore for the academic year 2023-2024.

We also declare that this project is the result of our own effort and has not been submitted to any other University for the award of any degree or diploma.

Ms.Apoorva	Ms. Pavithra B S	Ms.Kavyashri	Ms.Rumana
U05DG21S0108	U05DG21S0133	U05DG21S0154	U05DG21S0156

ACKNOWLEDGEMENT

It gives us immense pleasure to write an acknowledgement so for this project. We express our gratitude to our institution for providing us with good infrastructure, laboratory facility, qualified and inspiring staff whose guidance was of great help in successful completion of this project.

We would like to sincerely thank our project guide Mrs. Sushma.S for her guidance, support and constructive suggestions for the betterment of the project. We would thank our dignified Head Of Department Prof Dr. Mani Bhushan Dsouza for giving us an opportunity to embark upon this topic and continued encouragement throughout the preparation of the project. We thank our beloved Principal Dr. Bhaskar Shetty S for his ample support and encouragement.

There is no doubt that despite of our strenuous efforts; error might remain in the project. We take all the responsibility for any lack of clarity, occasional erratum or inaccuracies that might occur.

Thank you

Table of contents

Chapter 1: Introduction:

1.1 Introduction …..……2-3

1.2 Problem Statement ……..4

1.3 Importance…….4-5

1.4 Objectives of the Study ……5

Chapter 2: Scope Study

2.1 Introduction ……..7

2.2 Related works on "Analysis of Women Safety in India Cities Using Machine Learning on Tweets ……7-8

2.3 Related works on **"**Women's Safety Analysis on social media using Machine Learning" …….8-10

2.4 Summary ……10

Chapter 3: Methodology of Study

3.1 Introduction…….12

3.2 Existing System……..12

3.3 Proposed Model…….12-13

3.4 Specification of Proposed work……13-15

3.4.1Performance Evaluation Metrics…….14-15

3.5 Designing AND Implementation……...16-25

3.5.1Python Libraries……..17-18

3.5.2Implementation Code……..18-25

3.6 Summary ……26

Chapter 4: Analysis and Interpretation

4.1 Introduction ……...28

4.2 Performance Evaluation of Text Blob……...28-34

Chapter 5: Conclusions and Future Scope

5.1 Introduction ……..36

5.2 Conclusion …….36

5.3 Learning outcome …….....37

5.4 Future Scope ……37

APPINDIX

I. Tools of Data Collection

- Software ……..39
- Data Sets …….39

II. Bibliography

- Bibliography ……39-40

III. Photographs

- Screenshots ……..40-45

CHAPTER 1 : INTRODUCTION

Sentimental Analysis of women safety in India on Instagram data Using machine learning

CHAPTER 1: INTRODUCTION

1.1 Introduction:

Cybercrime can significantly impact women's confidence and well-being. Online bullying and harassment can make women feel scared and unsure about using the internet, leading to anxiety over the potential misuse of their personal information. This fear can discourage women from sharing their thoughts or interacting with others online, resulting in feelings of loneliness and isolation. The pervasive nature of cybercrime undermines women's sense of security and confidence in their ability to stay safe online.

To address these issues, it is crucial to systematically analyze and understand the sentiments expressed on Instagram regarding women's safety in Indian cities. Leveraging machine learning techniques allows for the classification and analysis of comments and posts related to women's safety. By examining the sentiment of these posts, we can gain valuable insights into the prevalent concerns, attitudes, and trends surrounding women's safety discourse online.

Analyzing Instagram data offers a crucial lens into the pervasive issue of harassment, cyberbullying, and gender-based violence that women encounter on social media platforms. By examining this data, we can uncover the nuances of these threats, ranging from derogatory comments to more overt forms of abuse, and comprehend their prevalence within the online community. This understanding is essential for developing targeted interventions and policies aimed at combating online abuse and ensuring the safety and well-being of users.

Additionally, employing machine learning algorithms to analyze Instagram data allows us not only to gain a deeper understanding of these complex issues but also to explore innovative research methodologies and foster interdisciplinary collaborations. The intersection of technology and gender studies provides a fertile ground for new approaches to digital safety, utilizing the power of data analysis to inform evidence-based interventions.

In this project, we focus on using machine learning techniques to analyze Instagram data. We recognize that behaviors such as staring at women and making inappropriate comments constitute forms of violence and harassment, yet these practices are often normalized, particularly in urban environments. Research conducted in India highlights the prevalence of such incidents, with studies showing that many women feel unsafe in public spaces, especially when surrounded by strangers in major metropolitan cities like Delhi, Chennai, and Mumbai.

Ultimately, our effort to understand and address women's safety on Instagram through machine learning not only illuminates the challenges they face but also serves as a catalyst for advancing the broader field of digital Safety. This initiative aims to foster meaningful change at the intersection of technology and gender equity, promoting a safer and more inclusive online environment.

The aim of this project is to analyze women's safety in Indian cities by applying machine learning techniques to data collected from Instagram. The focus is on leveraging Instagram data to gain insights into the prevalence of harassment, cyberbullying, and gender-based violence experienced by women on the platform.

By utilizing data from Instagram, we aim to understand the prevalence of various safety concerns faced by women, including harassment, discrimination, and threats to personal safety. Instagram, with its vast user generated content, serves as a valuable resource for examining these issues through real-world user experiences.

The project employs machine learning algorithms to categorize the safety concerns that women face. This categorization will help in understanding the different types of challenges encountered by women on Instagram, such as harassment and discrimination. The insights gained from this analysis can inform targeted interventions to enhance women's safety online.

Through sentiment analysis of Instagram comments and posts tagged with relevant hashtags like "lady harassment" or "lady well-being," the project aims to uncover prevailing sentiments, concerns, and trends in the discourse surrounding women's safety online. By extracting emotional content from these comments, the project will classify them into categories such as positive, negative, or neutral sentiments, providing a comprehensive view of public perceptions and experiences related to women's safety.

This project's significance lies in its potential to contribute to the development of effective strategies to combat online abuse and promote digital safety for women. By analyzing Instagram data, we aim to enhance our understanding of the challenges women face in digital spaces, informing efforts to create safer online environments. The project's findings can also foster interdisciplinary collaborations and innovative research methodologies at the intersection of gender studies and technology.

Ultimately, by leveraging machine learning techniques on Instagram data, the project seeks to empower women and promote their safety and well-being in Indian cities. The insights gained will not only help in addressing online abuse but also contribute to broader initiatives aimed at ensuring women's safety in both digital and physical spaces.

Through this project, we hope to advance the understanding of women's safety issues in Indian cities and provide a data-driven foundation for policies and interventions that can make digital spaces safer and more inclusive for women.

1.2. Problem Statement

Cybercrime can make women feel less confident about themselves. When women face online bullying or harassment, it can make them feel scared and unsure about using the internet. They might worry about their personal information being shared without their permission, which can make them feel even more anxious. This fear can make women less likely to share their thoughts or interact with others online, which can make them feel lonely and isolated. Cybercrime makes women feel less sure about themselves and their ability to stay safe online, which can really affect their confidence.

However, there is a need to analyze and understand the sentiments expressed on Instagram regarding women's safety in Indian cities more systematically. Leveraging machine learning techniques can enable the classification and analysis of comments and posts related to women's safety on Instagram. By examining the sentiment of these posts, we can gain insights into the prevalent concerns, attitudes, and trends surrounding women's safety discourse online.

Identifying Safety Threats : Understanding the types of safety threats and risks that women encounter on Instagram, including but not limited to harassment, abuse, stalking, and hate speech.

Analyzing Patterns and Trends: Uncovering patterns, trends, and sentiment analysis related to women's safety concerns in Instagram data, such as prevalent topics, harmful narratives, and discriminatory behaviors.

Developing Classification Models: Building machine learning models to classify and categorize women's safety-related content on Instagram, enabling the identification of harmful content and the promotion of positive and empowering narratives.

Informing Interventions: Providing actionable insights and recommendations based on data-driven analysis to inform interventions, policies, and strategies aimed at enhancing women's safety and well-being on Instagram.

Ethical Considerations: Addressing ethical considerations related to data privacy, consent, bias, and algorithmic fairness in the analysis of social media data to ensure responsible and ethical research practices.

1.3. Importance

Analyzing Instagram data provides a critical window into the pervasive issue of harassment, cyberbullying, and gender-based violence that women face on social media platforms. It allows us to uncover the nuances of these threats, from derogatory comments to more overt forms of abuse, and understand their prevalence within the online community. This insight is vital for crafting targeted interventions and policies aimed at combating online

abuse and ensuring the safety and well-being of users.

Moreover, by employing machine learning algorithms to dissect Instagram data, we not only gain a deeper understanding of these complex issues but also open avenues for innovative research methodologies and interdisciplinary collaborations. The intersection of technology and gender studies offers fertile ground for exploring new approaches to digital safety, leveraging the power of data analysis to inform evidence-based interventions.

Within the scope of this project, our focus remains on utilizing machine learning techniques to analyze Instagram data. We recognize that behaviors like staring at women and making inappropriate comments constitute forms of violence and harassment, yet sadly, these practices are often normalized, particularly in urban environments. Research conducted in India underscores the prevalence of such incidents, with studies revealing that a significant number of women feel unsafe in public spaces, especially when surrounded by strangers in major metropolitan cities such as Delhi, Chennai, and Mumbai.

In essence, our endeavor to understand and address women's safety on Instagram through machine learning not only sheds light on the challenges they face but also serves as a catalyst for advancing the broader field of digital safety and fostering meaningful change at the intersection of technology and gender equity.

1.4. Objectives of the Study

- Conduct an analysis of women's safety in Indian cities by applying machine learning techniques to Instagram data, focusing on prevalent issues like harassment, cyberbullying, and gender-based violence experienced by women on the platform.
- Utilize machine learning algorithms to categorize safety concerns faced by women on Instagram, including harassment, discrimination, and threats to personal safety, based on user-generated content.
- Perform sentiment analysis of Instagram comments and posts tagged with relevant hashtags such as "lady harassment" or "lady well-being" to uncover prevailing sentiments, concerns, and trends surrounding women's safety discourse online.
- Extract emotional content from Instagram comments and classify them into categories such as positive, negative, or neutral sentiments to provide insights into the public's perceptions and experiences related to women's safety.
- Contribute to the development of effective strategies to combat online abuse, promote digital safety for women, and advance our understanding of the challenges women face in digital spaces through interdisciplinary collaborations and innovative research methodologies.

Chapter 2: Scope Study

Chapter 2: Scope Study

2.1. Introduction:

In recent years, social media platforms like Facebook, Instagram, and Twitter have become hubs for people to freely share their opinions on Indian society and the claims made by politicians regarding the safety of women in Indian cities. On these platforms, individuals, especially women, have the space to express their views and narrate their personal experiences with sexual harassment. These stories not only shed light on the reality of women's safety but also serve as sources of motivation for others. When women share their stories of standing up against harassment, it inspires others to do the same. This ripple effect extends beyond just women, as men also join in raising their voices against those responsible for making Indian cities unsafe for women.

Researchers are utilizing advanced methods, such as machine learning, to analyze the massive amount of data generated on these platforms. By categorizing sentiments expressed in posts using techniques like hybrid and lexicon-based learning, they aim to gain insights into public opinions on women's safety. Additionally, categorizations such as statistical, knowledge based, and age-wise differentiation approaches are employed to understand the nuances of these sentiments. Through procedures like data extraction, analysis, and interpretation, valuable information is gleaned from social networking data. Moreover, the accuracy of analyzing and predicting trends on platforms like Instagram can be enhanced by studying behavioral patterns within social networks. Overall, social media serves as a powerful tool for both expressing emotions and opinions and for driving meaningful discussions and actions towards improving women's safety in Indian cities.

2.2. Literature review 1:

Title: "Analysis of Women Safety in India Cities Using Machine Learning on Tweets"
Year: 06, April 2022
Authors: Salla Anisha and
Srinidhi Ghankot Dr. BV
Ramana Murthy
Publication Name: Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Telangana, India

There are certain types of harassment and Violence that are very aggressive including staring and passing comments and these unacceptable practices are usually seen as a normal part of the urban life. There have been several studies that have been conducted in cities

across India and women report similar type of sexual harassment and passing off comments by other unknown people. The study that was conducted across most popular Metropolitan cities of India including Delhi, Mumbai and Pune, it was shown that 60 % of the women feel unsafe while going out to work or while traveling in public transport. Women have the right to the city which means that they can go freely whenever they want whether it be too an Educational Institute, or any other place women want to go. But women feel that they are unsafe in places malls, shopping malls on their way to their job location because of the several unknown Eyes body shaming and harassing these women point Safety or lack of concrete consequences in the life of women is the main reason of harassment of girls.

There are instances when the harassment of girls was done by their neighbors while they were on the way to school or there was a lack of safety that created a sense of fear in the minds of small girls who throughout their lifetime suffer due to that one instance that happened in their lives where they were forced to do something unacceptable or was sexually harassed by one of their own neighbor or any other unknown person. Safest cities approach women safety from a perspective of women rights to the affect the city without fear of violence or sexual harassment. Rather than imposing restrictions on women that society usually imposes it is the duty of society to imprecise the need of protection of women and also recognizes that women and girls also have a right same as men have to be safe in the City.

Analysis of twitter texts collection also includes the name of people and name of women who stand up against sexual harassment and unethical behavior of men in Indian cities which make them uncomfortable to walk freely. The data set that was obtained through Twitter about the status of women safety in Indian society was for the processed through machine learning algorithms for the purpose of smoothing the data by removing zero values and using Laplace and porter's theory is to developer method of analyzation of data and remove retweet and redundant data from the data set that is obtained so that a clear and original view of safety status of women in Indian society is obtained.

2.3. Literature review 2:

Title: "Women's Safety Analysis on Social Media Using Machine Learning"
Year: 2022
Authors: Mrs. B Ganga Bhavani, Ms. V Bhavani, Boddu Pavan Ganesh, Bonthu Sai Kiran, Bonthu Venkata Naveen, and Kallakuri Baladitya
Publication Name: BVCE

The increasing prevalence of social media platforms like Twitter has led to a significant focus on their potential to provide insights into public safety concerns, particularly regarding women's safety in urban areas. The study "Women's Safety Analysis on Social Media using Machine Learning" by Bhavani et al. (2022) delves into this topic, utilizing sentiment analysis to gauge women's perceptions of safety based on their social media activity. The authors highlight the pervasive issue of harassment in public spaces, noting that despite the normalization of such behavior, it remains a significant violation of women's civil liberties.

Sentiment analysis, a method for extracting emotional content from text, is central to this study. This technique categorizes tweets into positive, negative, or neutral sentiments using machine learning algorithms. Bhavani et al. employ tools like the TWEEPY Python program to gather tweets and the Natural Language Toolkit (NLTK) for preprocessing, which involves cleaning the text by removing special symbols and stop words. The analysis further incorporates the TextBlob library to determine the polarity of tweets, aiding in the classification of public sentiment towards women's safety.

The study references prior research that underscores the utility of social media data for sentiment analysis.

Agarwal et al. (2009) and Barbosa and Feng (2010) emphasize the robustness of sentiment detection in noisy

And biased data environments like Twitter. Bermingham and Smeaton (2010) and Gamon (2004) discuss the challenges and advantages of sentiment analysis in brief texts, which are characteristic of social media posts. These foundational studies inform the methodology adopted by Bhavani et al., particularly in handling the unique challenges posed by Twitter data.

The methodology section of Bhavani et al.'s study outlines a comprehensive approach, integrating algorithms like Term Frequency-Inverse Document Frequency (TF-IDF) and Decision Trees to enhance the accuracy of sentiment classification. This dual-algorithm approach aims to mitigate issues of fake reviews and malicious tweets, which can distort the true sentiment landscape. By training the Decision Tree algorithm on both genuine and fake tweets, the study seeks to ensure the reliability of its findings.

Furthermore, the study examines the broader implications of sentiment analysis on social media, as discussed by Gupta et al. (2017) and Sahayak et al. (2015), who explore the application of machine learning algorithms to Twitter data. The potential of sentiment analysis to inform public safety policies is also highlighted by Kumar and Aggarwal (2019), who analyze women's safety in Indian cities using tweets. This contextual framework supports Bhavani et al.'s assertion that sentiment analysis can provide actionable insights into public safety concerns.

In conclusion, Bhavani et al.'s research contributes to the growing body of literature on the use of social media data for sentiment analysis, particularly in the context of women's safety. By leveraging machine learning techniques and addressing the challenges of data authenticity, the study offers a nuanced understanding of how social media can reflect and influence public perceptions of safety. This work underscores the potential of sentiment analysis as a tool for social good, offering a pathway for future research and practical applications in public safety and urban planning.

2.4. Summary

Both studies underscore the importance of utilizing social media data to analyze and address women's safety concerns in urban areas. Salla Anisha et al.'s research focuses on the prevalence of harassment and the societal failure to protect women, employing machine learning techniques to analyze Twitter data for insights into public perceptions of safety. Bhavani et al.'s study complements this by employing sentiment analysis to categorize women's safety-related tweets, using advanced algorithms to ensure data authenticity and reliability.

Collectively, these studies highlight the significant role of machine learning and social media analysis in understanding and improving women's safety in Indian cities. They advocate for recognizing women's right to safe urban environments and the need for societal and policy interventions to combat harassment and violence effectively. By leveraging the insights gained from social media data, both studies contribute to a growing body of literature that seeks to inform public safety policies and urban planning with the goal of creating safer cities for women.

Chapter 3: Methodology of Study

Chapter 3: Methodology of Study

3.1. Introduction

Sentiment analysis is a crucial task in natural language processing, aimed at identifying and categorizing opinions expressed in text into positive, negative, or neutral sentiments. This process is particularly valuable for understanding public sentiment towards various topics, products, or services. In this context, we explore the major steps involved in conducting sentiment analysis on Instagram comments, adapting the process from its common application to instagram posts.

3.2. Existing System

The existing system for the project report involves the analysis of women's safety using social networking messages and applying machine learning algorithms on them. The process begins with the collection of tweets from Twitter using the Tweepy package in Python. These tweets are then used to detect women's sentiments. This is achieved by preprocessing the tweets using NLTK (natural language toolkit) to remove special symbols and stop words, making the text cleaner. Following this, the TEXTBLOB corpora package and dictionary are used to count the positive, negative, and neutral polarity of the tweets.
Moreover, the tweets are categorized based on their polarity values, with those having a value less than 0 considered negative, those between 0 and 0.5 considered neutral, and those above 0.5 considered positive.

This system aims to provide insights into women's sentiments and safety by analyzing the content of social media posts, particularly on Twitter, using machine learning and natural language processing techniques.

3.3. Proposed Model

The proposed model aims to analyze and evaluate the sentiment of Instagram comments related to women's safety issues. It categorizes comments into specific safety-related categories such as Rape, Stalking, Sexual Abuse, and Gender Discrimination, and performs sentiment analysis to understand public perception in these areas. The model reads Instagram comments based on the selected safety-related categories and cleans the data by removing stopwords, punctuation, and converting all text to lowercase for standardization. TextBlob is used for sentiment analysis, classifying comments as positive, negative, or neutral based on polarity scores. The results of the sentiment analysis, along with their percentages, are displayed, and a pie chart is used to visualize the sentiment distribution.

The user interface of the model includes a Combobox for category selection and buttons that enable actions such as uploading images, reading comments, cleaning comments, running machine learning algorithms, and displaying graphs. A Text widget is used to display the comments and results. The polarity score, which ranges from -1 (negative) to 1 (positive) and 0 (neutral), determines the sentiment of individual comments.

The formula for calculating the percentage of each sentiment category is:

Percentage of Sentiment Category=(Count of Sentiment Category/Total Number of Comments) ×100

Here, the count of sentiment categories (positive, negative, neutral) refers to the number of comments classified into each sentiment, and the total number of comments is the overall count of all comments analyzed. The percentage of sentiment category provides an overall view of the distribution of sentiments across the analyzed comments. Higher percentages of negative sentiments in specific categories can indicate areas of concern for women's safety. By cleaning the data, applying machine learning algorithms, and visualizing the results, the model provides insights into sentiment distribution and helps identify potentially unsafe areas for women.

3.4 Specification of Proposed work

1. Data Collection: Collect Instagram comments using the API based on the selected categories.
2. Data Cleaning: Process the comments to remove noise and standardize the text.
3. Sentiment Analysis: Apply TextBlob to classify the cleaned comments into positive, negative, or neutral sentiments.
4. Results Calculation: Count the number of comments in each sentiment category and calculate the percentages.
5. Results Display: Show the sentiment analysis results and their percentages. Visualize the sentiment distribution using a pie chart.
6. User Interface Interaction: Use the Combobox for category selection, and buttons for various actions to manage the workflow.

3.4.1 Performance Evaluation Metrics

Steps and Formulas:-

1. Read and Clean Comments
   - Reading Comments: Retrieve comments from the selected category.
   - Cleaning Comments: Remove stopwords, punctuation, and convert the text to lowercase for uniformity.

2. Sentiment Analysis with TextBlob
   - Use TextBlob to determine the sentiment polarity of each comment.
   - Polarity Score:

   Negative (-1 to 0): Indicates negative sentiment.

   Neutral (0): Indicates neutral sentiment.

   Positive (0 to 1): Indicates positive sentiment.

3. Count Sentiment Categories
   - Positive Comments Count (pos): Number of comments with polarity > 0.5.
   - Neutral Comments Count (neu): Number of comments with 0.2 < polarity <= 0.5.
   - Negative Comments Count (neg): Number of comments with polarity <= 0.2.

4. Calculate Percentages

   Total Comments Count (total): Total number of comments analyzed. Percentage Calculation Formula:

   Percentage of a Sentiment Category=(Count of a Sentiment Category/Total Number of Comments)×100 Percentage of a Sentiment Category=(Total Number of Comments/Count of a Sentiment Category)×100

Formulas for Each Sentiment Category:-

1. Positive Percentage:

   Positive Percentage=(postotal/Total)×100 postotal: Number of positive comments.

   total: Total number of comments.

2. Neutral Percentage:

   Neutral Percentage=(neutotal/Total)×100

   neutotal: Number of neutral comments.

   total: Total number of comments.

3. Negative Percentage:

   Negative Percentage=(negtotal/Total)×100

   negtotal: Number of negative comments.

   total: Total number of comments

Components Explanation:-

Combobox for Category Selection:

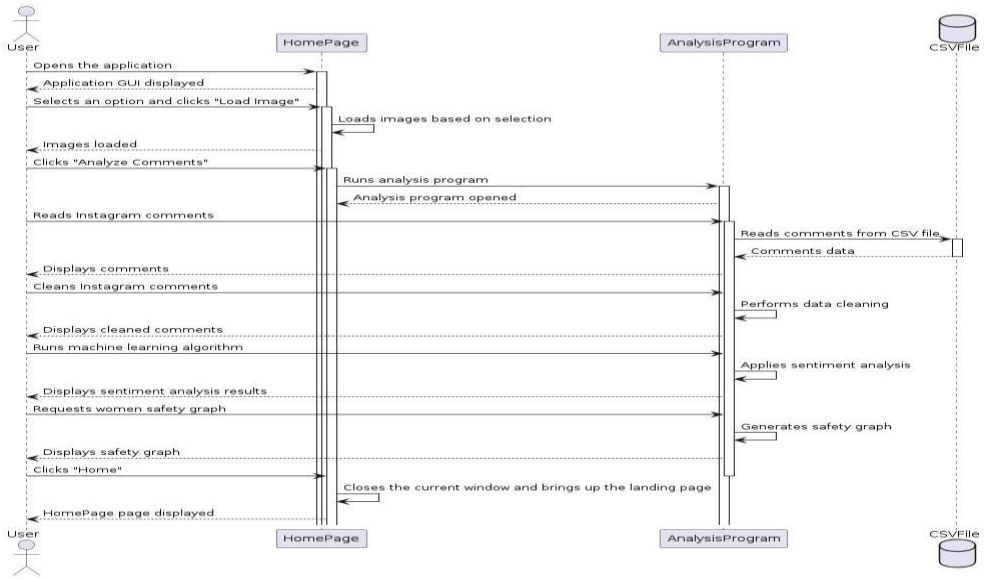
- Allows the user to select the category of comments (e.g., Rape, Stalking, Sexual abuse, Gender - discrimination).

Buttons for Actions:

- Upload Images Button: Loads images related to the selected category.
- Read Comments Button: Reads the comments from the selected category.
- Clean Comments Button: Cleans the read comments by removing stopwords, punctuation, and converting to lowercase.
- Run ML Algorithm Button: Applies TextBlob for sentiment analysis.
- Display Graphs Button: Visualizes the sentiment distribution using a pie chart Text Widget:
- Displays the comments and the results of the sentiment analysis.

3.5 Designing and Implementation

Sequence diagram

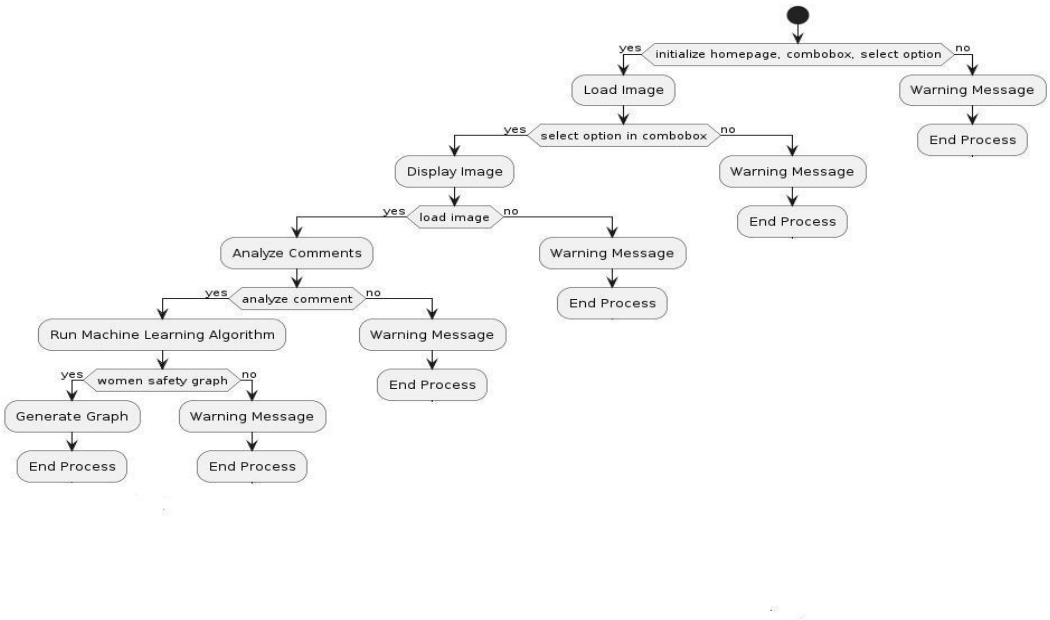


Implementation

Algorithm

1. Start
2. Import Required Libraries
3. Create an instance of the HomePage class, which serves as the main interface for the application
4. Enter the main loop to handle user interactions and events.
5. Selecting a category from the combobox and clicking buttons to upload images or comments.
6. Display a loading message while the images are loading.
7. Based on the selected category, load and display corresponding images in a grid layout.
8. When the user clicks the "Analyze Comment" button, open the analysis program window.
9. Create an instance of the Analysis Program class, passing the selected category and associated CSV file path.
10. Enter the main loop of the analysis program window.
11. The user interacts with the analysis program by clicking buttons to read comments, clean comments, run machine learning algorithms, and generate graphs.
12. Close the analysis program window when the user clicks the "Home" button to return to the landing page.
13. End Main Loop
14. stop

Flowchart Diagram



3.5.1 python libraries

1. tk: Tkinter

   import tkinter as tk

2. ttk: Tkinter Themed Widgets

   from tkinter import ttk

3. messagebox: Tkinter MessageBox

   from tkinter import messagebox

4. ImageTk: PIL (Python Imaging Library) ImageTk Module

   from PIL import ImageTk, Image

5.Image: PIL (Python Imaging Library) Image Module

  from PIL import ImageTk, Image

6. pd: pandas

   import pandas as pd

7. TextBlob: textblob

 from textblob import TextBlob

8. punctuation: string
from string import punctuation

9. stopword: nltk.corpus.stopwords
from nltk.corpus import stopwords

10. threading: threading
import threading

11. matplotlib: matplotlib
import matplotlib

12. plt: matplotlib.pyplot
import matplotlib.pyplot as plt

```
import tkinter as tk  # tk: Tkinter
from tkinter import ttk  # ttk: Tkinter Themed Widgets
from tkinter import messagebox  # messagebox: Tkinter MessageBox
from PIL import ImageTk, Image  # ImageTk: PIL ImageTk Module, Image: PIL Image Module
import pandas as pd  # pd: pandas
from textblob import TextBlob  # TextBlob: textblob
from string import punctuation  # punctuation: string
from nltk.corpus import stopwords  # stopwords: nltk.corpus.stopwords
import threading  # threading: threading
import matplotlib  # matplotlib: matplotlib
matplotlib.use('TkAgg')  # Initialize Matplotlib for use with Tkinter
import matplotlib.pyplot as plt  # plt: matplotlib.pyplot
```

3.5.2 implementation code

Here are the implementation blocks extracted from the given code, along with captions and brief descriptions:

1. Landing Page Initialization

```
class LandingPage(tk.Tk):
    def __init__(self):
        super().__init__()
```

```
self.title("Analysis of Women Safety in India Using Instagram Data")
self.geometry("1000x300")
self.configure(bg="#f0f0f0")
```

Description: This initializes the main landing page window with the title, size, and background color.

2. **Title Label**

```
self.label_title = tk.Label(self, text="Analysis of Women Safety in India Using Instagram Data", font=("Helvetica", 24, "bold"), bg="#f0f0f0")
self.label_title.pack(pady=5)
```

Description: This sets up the title label on the landing page.

3. Combobox for Analysis Option Selection

```
self.label_choose_option = tk.Label(self, text="Select an option:", font=("Helvetica", 14), bg="#f0f0f0")
self.label_choose_option.pack()
self.combobox = ttk.Combobox(self, values=["Rape", "Stalking", "Sexual abuse", "Gender discrimination"], font=("Helvetica", 12))
self.combobox.pack(pady=5)
```

Description: This creates a label and a combobox for selecting an analysis option.

4. Upload Button

```
self.upload_button = tk.Button(self, text="Load Image", command=self.display_images, font=("Helvetica", 12))
self.upload_button.pack(pady=5)
```

Description: This creates a button to load images.

5. Frame for Displaying Images

```
self.grid_frame = tk.Frame(self, bg="#f0f0f0")
self.grid_frame.pack(pady=5)
```

Description: This creates a frame where images will be displayed.

6. Analyze Button

```
self.upload_comment_button = tk.Button(self, text="Analyze Comments", command=self.open_analysis_program, font=("Helvetica", 12))
self.upload_comment_button.pack(pady=5)
```

Description: This creates a button to start comment analysis.

7. Display Images Method

```
def display_images(self):
    selected_value = self.combobox.get()
    if not selected_value:
        messagebox.showwarning("Warning", "Please select a value from the combobox.")
        return
    images = None
     if selected_value == "Rape":
        images = [r"D:\final project\Rape1.jpeg",r"D:\final project\Rape5.jpg",r"D:\final
       project\Rape3.jpeg",r"D:\final project\Rape4.jpeg"]
    elif selected_value == "Stalking":
        images = [r"D:\final project\stk1.jpg",r"D:\final project\stk2.png",r"D:\final
        project\stk3.jpg",r"D:\final project\stk4.jpg"]
    elif selected_value == "Sexual abuse":
        images = [r"D:\final project\sexual5.jpg",r"D:\final project\sexual6.jpg",r"D:\final
      project\sexual3.jpg",r"D:\final project\sexual2.webp"]
    elif selected_value == "Gender discrimination":
        images = [r"D:\final project\gender1.jpg",r"D:\final project\gender2.jpg",r"D:\final
        project\gender3.jpg",r"D:\final project\gender4.jpg"]
    if images is None:
        messagebox.showwarning("Warning", "No images found for the selected option.")
        return
    loading_label = tk.Label(self.grid_frame, text="Loading...", font=("Helvetica", 12),
    bg="#f0f0f0")
    loading_label.grid(row=0, column=0, columnspan=2, padx=10, pady=10)
    self.after(1000, lambda: self._display_images(images, loading_label))
```

Description: This method handles image loading and displaying based on the selected combobox value.

8. Display Images Helper Method

```
def _display_images(self, images, loading_label):
    for widget in self.grid_frame.winfo_children():
        widget.destroy()
    row = 0
    col = 0
    for img_path in images:
```

```
        img = Image.open(img_path)
        img = img.resize((200, 200), Image.LANCZOS)
        photo = ImageTk.PhotoImage(img)
        label = tk.Label(self.grid_frame, image=photo)
        label.image = photo
        label.grid(row=row, column=col, padx=10, pady=10)
        col += 1
        if col > 1:
            col = 0
            row += 1
    loading_label.destroy()
```

Description: This helper method handles the actual display of images after loading.

9. Open Analysis Program Method

```
def open_analysis_program(self):
    selected_value = self.combobox.get()
    if not selected_value:
        messagebox.showwarning("Warning", "Please select a value from the combobox.")
        return
    csv_files = {
        "Rape": r"D:\final project\Rape.csv",
        "Stalking": r"D:\final project\stalking.csv",
        "Sexual abuse": r"D:\final project\sexual_abusee(1).csv",
        "Gender discrimination": r"D:\final project\genderdiscrimination(1).csv"
    }
    if selected_value in csv_files:
        self.run_analysis(selected_value, csv_files[selected_value])
    else:
        messagebox.showwarning("Warning", "Invalid selection")
```

Description: This method opens the analysis window based on the selected combobox value.

10. Analysis Program Initialization

```
class AnalysisProgram:
    def __init__(self, selected_value, csv_file, landing_page_instance):
        self.main = tk.Toplevel()
        self.landing_page_instance = landing_page_instance
```

```
self.main.title("Analysis of Women Safety in India Using Instagram Data")
self.main.geometry("1000x1000")
self.selected_value = selected_value
self.filename = csv_file
self.insta_list = []
self.clean_list = []
self.pos = 0
self.neu = 0
self.neg = 0
self.font = ('times', 14, 'bold')
self.title = tk.Label(self.main, text='Analysis of Women Safety in India Using
 Instagram Data')
self.title.config(bg='antiquewhite', fg='black')
self.title.config(font=self.font)
       self.title.config(height=3, width=130)
self.title.place(x=0, y=5)
      self.home_button = tk.Button(self.main, text="Home",
     command=self.go_to_landing_page)
       self.home_button.place(x=30, y=150)
self.home_button.config(font=self.font)
self.read_button = tk.Button(self.main, text="Read Instagram Comments",
 command=self.read)
self.read_button.place(x=135, y=150)
self.read_button.config(font=self.font)
self.clean_button = tk.Button(self.main, text="Clean Instagram Comments",
 command=self.clean)
self.clean_button.place(x=420, y=150)
self.clean_button.config(font=self.font)
self.ml_button = tk.Button(self.main, text="Run Machine Learning Algorithm",
command=self.machine_learning)
self.ml_button.place(x=710, y=150)
self.ml_button.config(font=self.font)
self.graph_button = tk.Button(self.main, text="Women Safety Graph",
command=self.graph)
self.graph_button.place(x=1050, y=150)
self.graph_button.config(font=self.font)
self.font1 = ('times', 12, 'bold')
```

```
        self.text = tk.Text(self.main, height=25, width=167)
        self.scroll = tk.Scrollbar(self.text)
        self.text.configure(yscrollcommand=self.scroll.set)
        self.text.place(x=10, y=200)
       self.text.config(font=self.font1)
        self.main.config(bg='darkgrey')
        self.main.mainloop()
```

Description: This initializes the analysis program window with necessary buttons and text areas.

11. Go to Landing Page Method

```
def go_to_landing_page(self):
    self.main.destroy()
    self.landing_page_instance.update()
    self.landing_page_instance.deiconify()
```

Description:This method returns the user to the landing page.

12. Read Comments Method

```
def read(self):
        self.text.delete('1.0', tk.END)
    self.insta_list.clear()
    if self.filename:
            train = pd.read_csv(self.filename, encoding='iso-8859-1')
        for i in range(len(train)):
            insta = train.at[i, 'Comments']
            self.insta_list.append(insta)
            self.text.insert(tk.END, insta + "\n")
    else:
        messagebox.showwarning("Warning", "Please select a CSV file first.")
```

Description: This method reads Instagram comments from the specified CSV file.

13. Clean Comments Method

```
def clean(self):
    self.text.delete('1.0', tk.END)
    self.clean_list.clear()
    if self.insta_list:
        for i in range(len(self.insta_list)):
```

```
            insta = self.insta_list[i]
            insta = insta.strip("\n")
            insta = insta.strip()
            insta = self.tweet_cleaning(insta.lower())
            self.clean_list.append
```

Description: This method cleans the Instagram comments for analysis.

14.Build machine learning algorithm

```
def machine_learning(self):
    self.text.delete('1.0', tk.END)
    self.pos = 0
    self.neu = 0
    self.neg = 0
    if self.clean_list:
        for i in range(len(self.clean_list)):
            insta = self.clean_list[i]
            blob = TextBlob(insta)
            if blob.polarity <= 0.2:
                self.neg += 1
                        self.text.insert(tk.END, insta + "\n")
                self.text.insert(tk.END, "Predicted Sentiment : NEGATIVE\n")
                self.text.insert(tk.END, "Polarity Score      : " + str(blob.polarity) + "\n")
                 self.text.insert(tk.END,
'============================================================
============================================================\n')
            if 0.2 < blob.polarity <= 0.5:
                self.neu += 1

                        self.text.insert(tk.END, insta + "\n")
                self.text.insert(tk.END, "Predicted Sentiment : NEUTRAL\n")
                self.text.insert(tk.END, "Polarity Score      : " + str(blob.polarity) + "\n")
                self.text.insert(tk.END,
'============================================================
============================================================\n')
            if blob.polarity > 0.5:
                self.pos += 1
                self.text.insert(tk.END, insta + "\n")
```

```
                self.text.insert(tk.END, "Predicted Sentiment : POSITIVE\n")
                self.text.insert(tk.END, "Polarity Score      : " + str(blob.polarity) + "\n")
                self.text.insert(tk.END,
'=============================================================================
============================================================\n')
        else:
            messagebox.showwarning("Warning", "Please clean comments first.")
```

Description:This function performs sentiment analysis on cleaned Instagram comments and updates a Tkinter text widget with the results.

15 .Sentiment Distribution Pie Chart in Tkinter using Matplotlib

```
def graph(self):
    if self.clean_list:
        label_X = ['Positive', 'Negative', 'Neutral']
        category_X = [self.pos, self.neg, self.neu]
        self.text.delete('1.0', tk.END)
        self.text.insert(tk.END, "Safety Factor\n\n")
        self.text.insert(tk.END, 'Positive : ' + str(self.pos) + "\n")

            self.text.insert(tk.END, 'Negative : ' + str(self.neg) + "\n")
        self.text.insert(tk.END, 'Neutral  : ' + str(self.neu) + "\n\n")
        self.text.insert(tk.END, 'Length of tweets  : ' + str(len(self.clean_list)) + "\n")
        self.text.insert(tk.END, 'Positive : ' + str(self.pos) + ' / ' + str(len(self.clean_list)) + '
        = ' + str(self.pos / len(self.clean_list) * 100) + '%\n')
          self.text.insert(tk.END, 'Negative : ' + str(self.neg) + ' / ' + str(len(self.clean_list))
         + ' = ' + str(self.neg / len(self.clean_list) * 100) + '%\n')
        self.text.insert(tk.END, 'Neutral  : ' + str(self.neu) + ' / ' + str(len(self.clean_list)) + '
         = ' + str(self.neu / len(self.clean_list) * 100) + '%\n')
            plt.pie(category_X, labels=label_X, autopct='%1.1f%%')
        plt.title('Women Safety & Sentiment Graph')
        plt.axis('equal')
        plt.show()
    else:
        messagebox.showwarning("Warning", "Please process machine learning first.")
```

Description: This function visualizes the sentiment distribution of processed Instagram comments as a pie chart using Matplotlib within a Tkinter GUI.

3.6.Summary

The methodology of the study focuses on conducting sentiment analysis on Instagram comments related to women's safety issues. The process involves several key steps, including data collection, data cleaning, sentiment analysis, results calculation, and results display. The study aims to categorize comments into specific safety-related categories such as rape, stalking, sexual abuse, and gender discrimination, and analyze the sentiments expressed within these categories.

The existing system analyzed women's safety using social networking platform messages and machine learning algorithms, originally focusing on Twitter data but now utilizing Instagram data. By examining Instagram comments, patterns and sentiments indicating safety concerns can be identified, providing insights into the safety of different areas for women.

The proposed model includes objectives, model components, formulas, variables, interpretation of metrics, and the process flow. The model utilizes techniques such as text cleaning, sentiment analysis using TextBlob, and visualization of results using pie charts to provide insights into sentiment distribution and identify potentially unsafe areas for women.

The design and implementation section describe the structure and functionality of the application. It includes the initialization of the landing page, selection of analysis options through a combobox, loading of images, analysis of comments, and display of results. The analysis program window allows users to read, clean, and analyze comments, as well as visualize safety-related graphs.

Overall, the methodology and implementation aim to provide a comprehensive analysis of women's safety perceptions based on Instagram comments, leveraging natural language processing and machine learning techniques.

Chapter 4: Analysis and Interpretation

Chapter 4: Analysis and Interpretation

4.1. Introduction

In our analysis and interpretation of social media data, particularly focusing on comments related to women's experiences on platforms like Instagram, we delve into the sentiments expressed and the underlying issues they highlight. By utilizing tools like sentiment analysis through the TextBlob library, we categorize comments into positive, negative, or neutral sentiments, aiming to understand the impact of online interactions on women's well-being.

Cybercrime and online harassment pose significant challenges for women, impacting their sense of safety and confidence in navigating the digital space. Instances of bullying, stalking, and privacy violations contribute to feelings of fear and unease among women, hindering their ability to freely express themselves online. Our analysis seeks to shed light on these issues, emphasizing the importance of creating a safer online environment where women feel empowered to share their thoughts and experiences without fear of harassment or intimidation.

Through our interpretation of the collected data, we aim to advocate for change and raise awareness about the prevalence of online harassment against women. By highlighting the need for empowerment, support, education, and awareness, we strive to foster a more inclusive and respectful online community where women can thrive without facing threats to their safety and well-being. Together, we can work towards creating a digital space that upholds the rights and dignity of all individuals, regardless of gender.

4.2.Performance Evaluation of TextBlob

Percentage of Sentiment Category=(Count of Sentiment Category/Total Number of Com ments)×100

- Example 1: Analysis of Rape Comments

Total Number of Comments: 500

Count of Positive Comments: 50

Count of Negative Comments: 300

Count of Neutral Comments: 150

Percentage Calculation:

Positive Comments: (50500)×100=10%(50050)×100=10%

Negative Comments: (300500)×100=60%(500300)×100=60%

Neutral Comments: (150500)×100=30%(500150)×100=30%

Metric Interpretation:

- Negative Comments (60%): A significant majority of the comments are negative, indicating strong concerns and dissatisfaction among users regarding safety in relation to rape.
- Neutral Comments (30%): A considerable portion of the comments are neutral, suggesting that many users are discussing the topic without expressing strong sentiments.
- Positive Comments (10%): A small percentage of comments are positive, which might indicate occasional positive remarks or support, but it's overshadowed by negative sentiments.

- Example 2: Analysis of Stalking Comments

Total Number of Comments: 400

Count of Positive Comments: 40

Count of Negative Comments: 280

Count of Neutral Comments: 80

Percentage Calculation:

Positive Comments: (40400)×100=10%(40040)×100=10%

Negative Comments: (280400)×100=70%(400280)×100=70%

Neutral Comments: (80400)×100=20%(40080)×100=20%

Metric Interpretation:

- Negative Comments (70%): A very high proportion of negative comments suggests a serious concern about stalking and perceived lack of safety.
- Neutral Comments (20%): A smaller percentage of comments are neutral, indicating that while some users are discussing the topic, most are expressing strong opinions.
- Positive Comments (10%): As with Example 1, positive comments are minimal, indicating rare positive

feedback or hopeful remarks about improvements in safety.

- Example 3: Analysis of Sexual Abuse Comments

Total Number of Comments: 600

Count of Positive Comments: 120

Count of Negative Comments: 360

Count of Neutral Comments: 120

Percentage Calculation:

Positive Comments: (120600)×100=20%(600120)×100=20%

Negative Comments: (360600)×100=60%(600360)×100=60%

Neutral Comments: (120600)×100=20%(600120)×100=20%

Metric Interpretation:

- Negative Comments (60%): The majority of comments are negative, showing substantial concern and discontent related to sexual abuse.
- Neutral Comments (20%): Neutral comments are significant but less than negative, indicating discussions that may not carry strong sentiments.
- Positive Comments (20%): A higher proportion of positive comments compared to previous examples suggests some positive perceptions or reports of positive developments, though still overshadowed by negative sentiments.

- Example 4: Analysis of Gender Discrimination Comments

Total Number of Comments: 300

Count of Positive Comments: 30

Count of Negative Comments: 180

Count of Neutral Comments: 90

Percentage Calculation:

Positive Comments: (30300)×100=10%(30030)×100=10%

Negative Comments: (180300)×100=60%(300180)×100=60%

Neutral Comments: (90300)×100=30%(30090)×100=30%

Metric Interpretation:

- Negative Comments (60%): Most comments are negative, highlighting serious concerns about gender discrimination.
- Neutral Comments (30%): A significant number of neutral comments suggest many discussions without strong sentiments.
- Positive Comments (10%): A minimal percentage of positive comments implies rare positive feedback or
support for gender equality issues.

- Example 5: Combined Safety Analysis

Total Number of Comments: 1000

Count of Positive Comments: 100

Count of Negative Comments: 600

Count of Neutral Comments: 300

Percentage Calculation:

Positive Comments: (1001000)×100=10%(1000100)×100=10%

Negative Comments: (6001000)×100=60%(1000600)×100=60%

Neutral Comments: (3001000)×100=30%(1000300)×100=30%

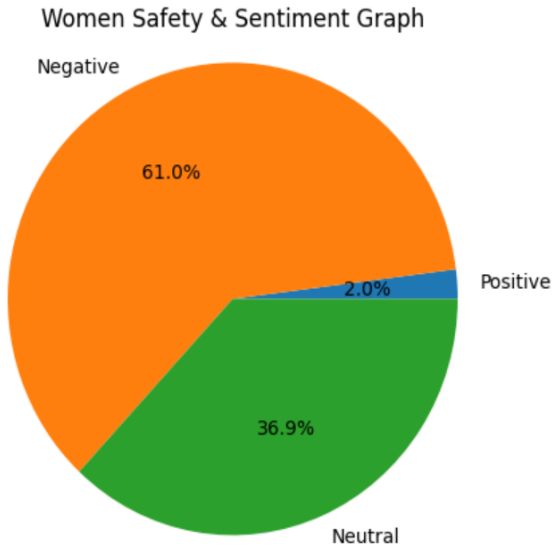
Metric Interpretation:

- Negative Comments (60%): The overall sentiment shows a majority of negative comments, indicating a widespread concern for women's safety across various issues.
- Neutral Comments (30%): A significant portion of neutral comments suggests many users discuss safety issues without expressing strong emotions.
- Positive Comments (10%): A small percentage of positive comments might indicate some positive changes or support but is largely outweighed by negative sentiments.

Interpretation

caption:

“Sentiment Analysis on Women's Safety Concerning Rape”



Description:

The pie chart visually represents the sentiment analysis concerning women's safety based on rape. It categorizes the sentiments into three distinct sections:

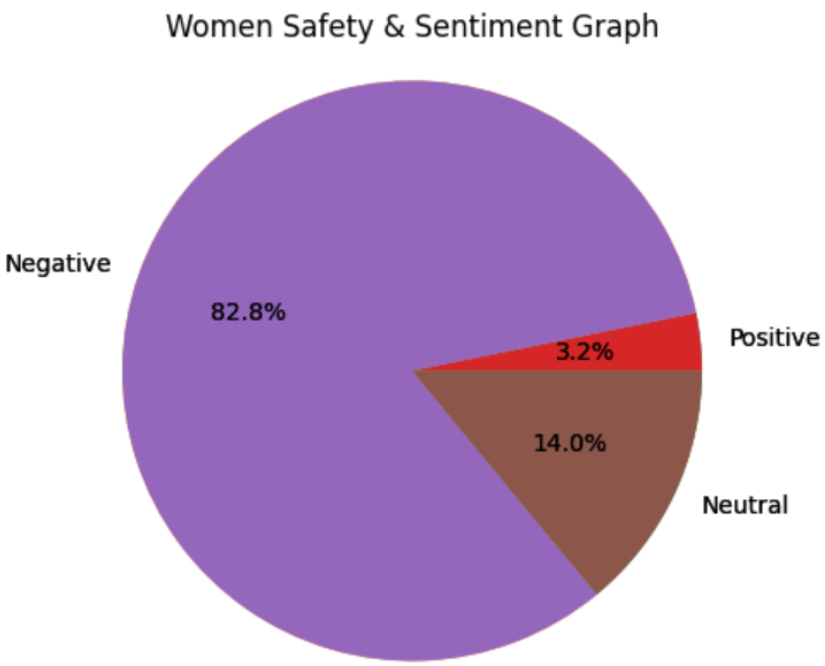
Negative: The largest portion of the chart, colored in orange, represents a negative sentiment, accounting for 61.0% of the responses.

Neutral: The second largest segment, shown in green, indicates a neutral sentiment, making up 36.9% of the responses.

Positive: The smallest section, colored in blue, reflects a positive sentiment, comprising only 2.0% of the responses.

caption:

“Sentiment Analysis on Women's Safety Concerning Stalking”



Description:

The pie chart visualizes the sentiment related to women's safety based on stalking. It is divided into three segments:

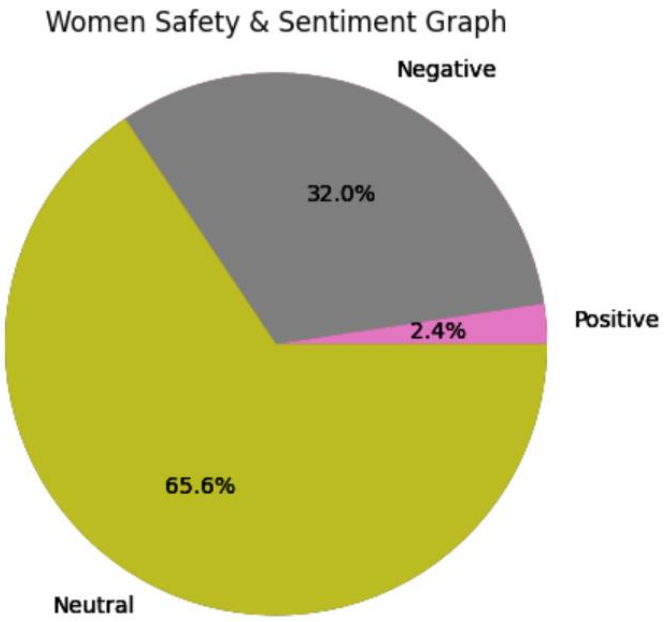
Negative: The largest segment, colored in orange, represents a negative sentiment, accounting for 82.8% of the responses.

Neutral: The second segment, colored in green, indicates a neutral sentiment, making up 14.0% of the responses.

Positive: The smallest segment, colored in blue, shows a positive sentiment, comprising only 3.2% of the responses.

caption:

"Sentiment Analysis on Women's Safety Concerning Sexual abuse”



Description:

The pie chart visually represents the sentiment analysis regarding women's safety and experiences related to sexual abuse. It categorizes the sentiments into three distinct sections:

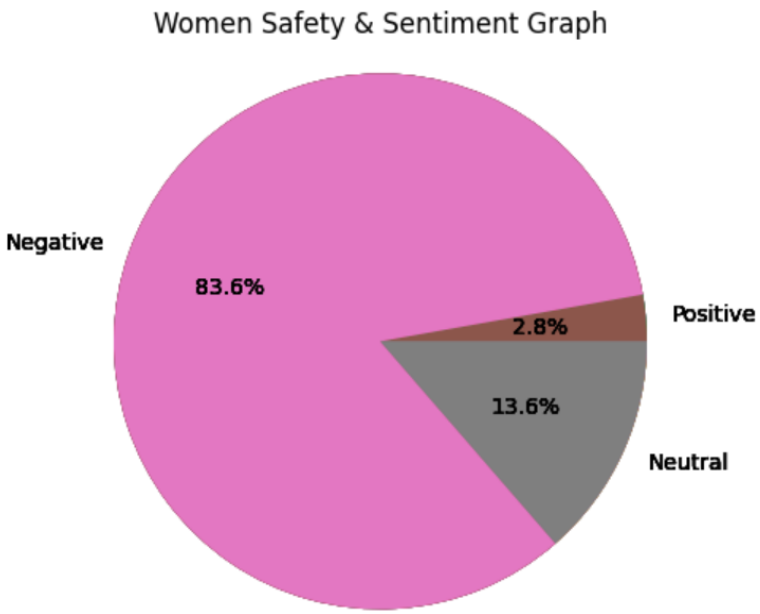
Neutral: The largest portion of the chart, colored in green, represents a neutral sentiment, accounting for 65.6%.

Negative: The second largest segment, shown in orange, indicates a negative sentiment, making up 32.0% of the chart.

Positive: The smallest section, depicted in blue, shows a positive sentiment, which is only 2.4%.

caption:

"Sentiment Analysis on Women's Safety Concerning Gender discrimination”



Description:

The pie chart visualizes the sentiment towards women's safety and experiences of gender discrimination. It is divided into three segments indicating different sentiments:

Negative: This segment is the largest, colored in orange, and accounts for 83.6% of the responses, indicating a predominantly negative sentiment.

Neutral: Colored in green, this segment represents 13.6% of the responses, suggesting a neutral viewpoint.

Positive: The smallest segment, colored in blue, comprises only 2.8% of the responses, indicating a very low level of positive sentiment regarding women's safety and gender discrimination.

Chapter 5: Conclusions and Future Scope

Chapter 5: Conclusions and Future Scope

5.1. Introduction:

In the digital age, social media platforms like Instagram have become integral parts of our daily lives, facilitating communication, connection, and expression. However, alongside the benefits, there exists a darker side characterized by online harassment and abuse, particularly targeting women.

This study delves into the prevalent problem of online harassment against women on Instagram, shedding light on its multifaceted nature and the profound effects it has on individuals and society as a whole.

This study shows that it's really important to help women feel strong and supported so they can use the internet safely. This research proves that it's super important to make sure women feel confident and have help so they can use the internet without feeling scared.

This study outlines a comprehensive future scope aimed at addressing the complex challenges of online harassment against women. From analyzing women's safety concerns in India to leveraging advanced technologies like sentiment analysis and machine learning, the goal is to enhance our understanding and response to online threats.

5.2. Conclusion:

- Prevalence of Online Harassment: Online harassment against women on Instagram is a big problem, with lots of hurtful comments and threats being made.
- Impact on Women's Well-being: Women feel really bad when they're harassed online. It makes them scared and anxious, and they might stop using Instagram because of it.
- Underreporting and Silent Suffering: Many times, women don't tell anyone when they're harassed online because they're worried about what might happen if they do. This means the problem often goes unnoticed and keeps happening.
- Need for Empowerment and Support: We need to help women feel stronger and give them places online where they feel safe. Providing things like helplines and support groups can make a big difference.
- Importance of Education and Awareness: It's super important to teach women and girls how to stay safe online. They need to know about their rights, how to keep their information private, and how to deal with mean people online.

5.3. Learning outcome:

- Data Collection and Cleaning:
  - The process of gathering raw data from various sources
  - cleaning and pre-processing data for analysis. This includes removing duplicate entries
- Programming and Tools:
  - Proficiency in programming languages like Python
  - Familiarity with tools and software such as Excel
- Machine Learning:
  - Knowledge of machine learning algorithms and their practical applications.
  - Experience in building, training, and validating machine learning models.
- Data Visualization:
  - The graphical representation of data to communicate insights and patterns effectively.
  - Ability to use visualization tools to tell a compelling data story.

5.4. Future Scope:

- Analyzing women's safety in India using Instagram data, focusing on issues like rape, stalking, sexual abuse, gender discrimination, and other related concerns.
- Future development includes expanding data analysis to incorporate platforms like Twitter, Facebook, and news articles for a more comprehensive understanding of women's safety concerns.
- Advancing sentiment analysis through sophisticated NLP techniques and machine learning algorithms to enhance comment classification and trend identification.
- Creating a system that constantly checks social media for signs of safety problems as they happen. This helps us react quickly to keep people safe.
- Using maps and location-based information to understand safety issues in different regions or cities. This helps authorities focus efforts where they're needed most and distribute resources effectively.
- Apply the system to analyze other societal issues like child abuse and animal abuse, helping to identify and address various forms of abuse and violence.

APPINDIX

APPINDIX

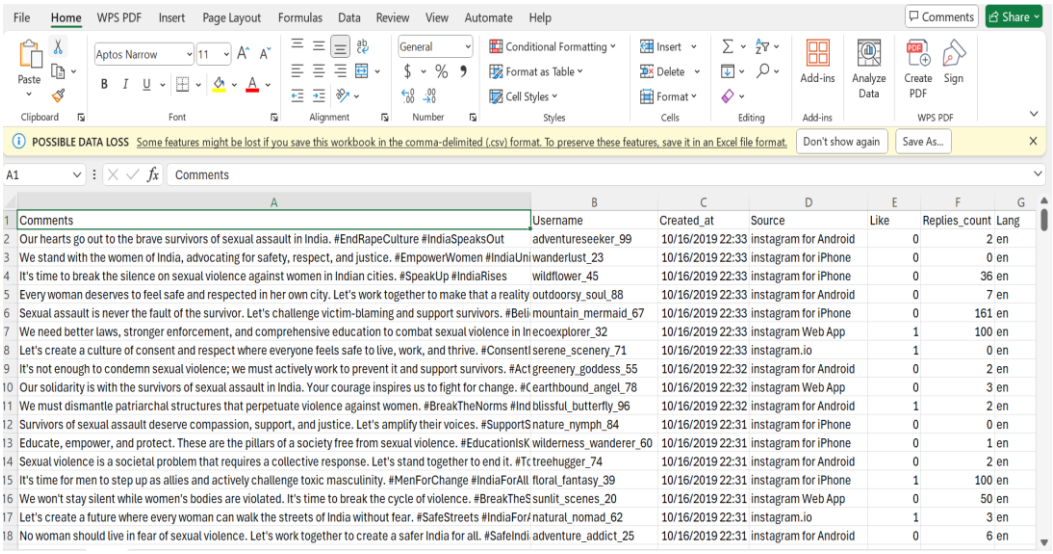
I. Tools of Data Collection

Software

- Python IDLE: Python IDLE (Integrated Development and Learning Environment) serves as the primary development environment for this project. The version of the IDLE is IDLE (Python 3.11 64-bit).
- Microsoft Excel 2016

Data Sets

- The dataset was obtained from Kaggle.com
- https://www.kaggle.com/hollyhetherington/metootweets?select=MeToo_tweets.csv
- Based on this dataset created 4 datasets



	A	B	C	D	E	F	G
1	Comments	Username	Created_at	Source	Like	Replies_count	Lang
2	Our hearts go out to the brave survivors of sexual assault in India. #EndRapeCulture #IndiaSpeaksOut	adventureseeker_99	10/16/2019 22:33	instagram for Android	0	2	en
3	We stand with the women of India, advocating for safety, respect, and justice. #EmpowerWomen #IndiaUni	wanderlust_23	10/16/2019 22:33	instagram for iPhone	0	0	en
4	It's time to break the silence on sexual violence against women in Indian cities. #SpeakUp #IndiaRises	wildflower_45	10/16/2019 22:33	instagram for iPhone	0	36	en
5	Every woman deserves to feel safe and respected in her own city. Let's work together to make that a reality	outdoorsy_soul_88	10/16/2019 22:33	instagram for Android	0	7	en
6	Sexual assault is never the fault of the survivor. Let's challenge victim-blaming and support survivors. #Beli	mountain_mermaid_67	10/16/2019 22:33	instagram for iPhone	0	161	en
7	We need better laws, stronger enforcement, and comprehensive education to combat sexual violence in In	ecoexplorer_32	10/16/2019 22:33	instagram Web App	1	100	en
8	Let's create a culture of consent and respect where everyone feels safe to live, work, and thrive. #Consenti	serene_scenery_71	10/16/2019 22:33	instagram.io	1	0	en
9	It's not enough to condemn sexual violence; we must actively work to prevent it and support survivors. #Act	greenery_goddess_55	10/16/2019 22:32	instagram for Android	0	2	en
10	Our solidarity is with the survivors of sexual assault in India. Your courage inspires us to fight for change. #C	earthbound_angel_78	10/16/2019 22:32	instagram Web App	0	3	en
11	We must dismantle patriarchal structures that perpetuate violence against women. #BreakTheNorms #Ind	blissful_butterfly_96	10/16/2019 22:32	instagram for Android	1	2	en
12	Survivors of sexual assault deserve compassion, support, and justice. Let's amplify their voices. #SupportS	nature_nymph_84	10/16/2019 22:31	instagram for iPhone	0	0	en
13	Educate, empower, and protect. These are the pillars of a society free from sexual violence. #EducationIsK	wilderness_wanderer_60	10/16/2019 22:31	instagram for iPhone	0	1	en
14	Sexual violence is a societal problem that requires a collective response. Let's stand together to end it. #Tc	treehugger_74	10/16/2019 22:31	instagram for Android	0	2	en
15	It's time for men to step up as allies and actively challenge toxic masculinity. #MenForChange #IndiaForAll	floral_fantasy_39	10/16/2019 22:31	instagram for iPhone	1	100	en
16	We won't stay silent while women's bodies are violated. It's time to break the cycle of violence. #BreakTheS	sunlit_scenes_20	10/16/2019 22:31	instagram Web App	0	50	en
17	Let's create a future where every woman can walk the streets of India without fear. #SafeStreets #IndiaFor/	natural_nomad_62	10/16/2019 22:31	instagram.io	1	3	en
18	No woman should live in fear of sexual violence. Let's work together to create a safer India for all. #SafeIndi	adventure_addict_25	10/16/2019 22:31	instagram for Android	0	6	en

II. Bibliography

Text Books:

1. Arun K Pujari – "Data Mining Techniques" 3rd Edition, Universities Press
2. Jiawei Han and Micheline Kambar – "Data Mining Concepts and Techniques" second edition
3. Nageswara Rao, R. (2019). Core Python Programming (3rd ed.). Dreamtech Press.

Websites:

- https://www.w3schools.com/python/
- https://www.pythontutorial.net/

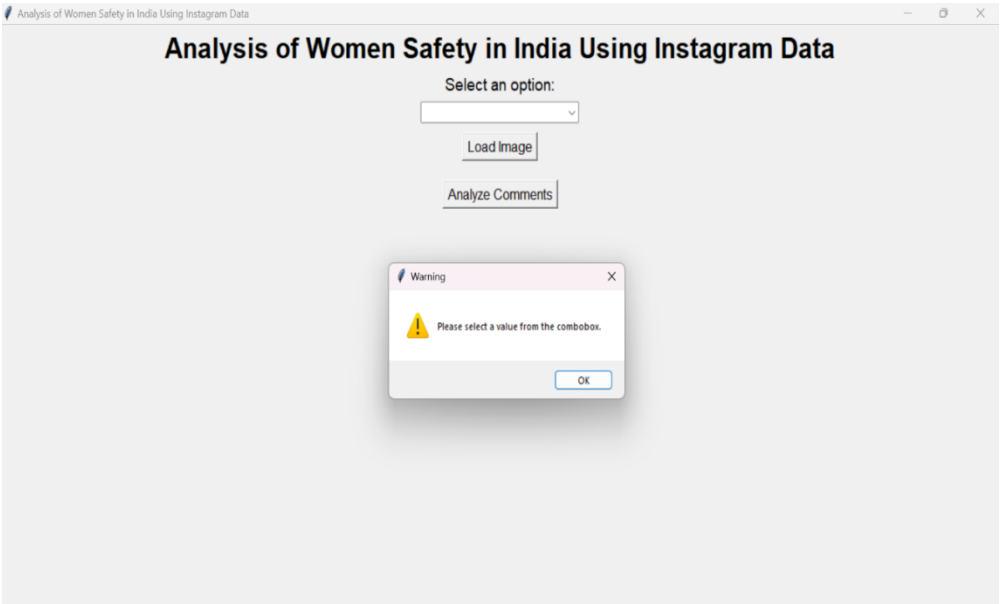
- https://courses.edx.org/assetv1:ColumbiaX+CSMM.101x+1T2017+type@asset+block@AI_edx_ml_5.1intro.pdf
- https://www.alpha-sense.com/blog/engineering/sentiment-score/

III Photographs

- **Screenshots**

Home page

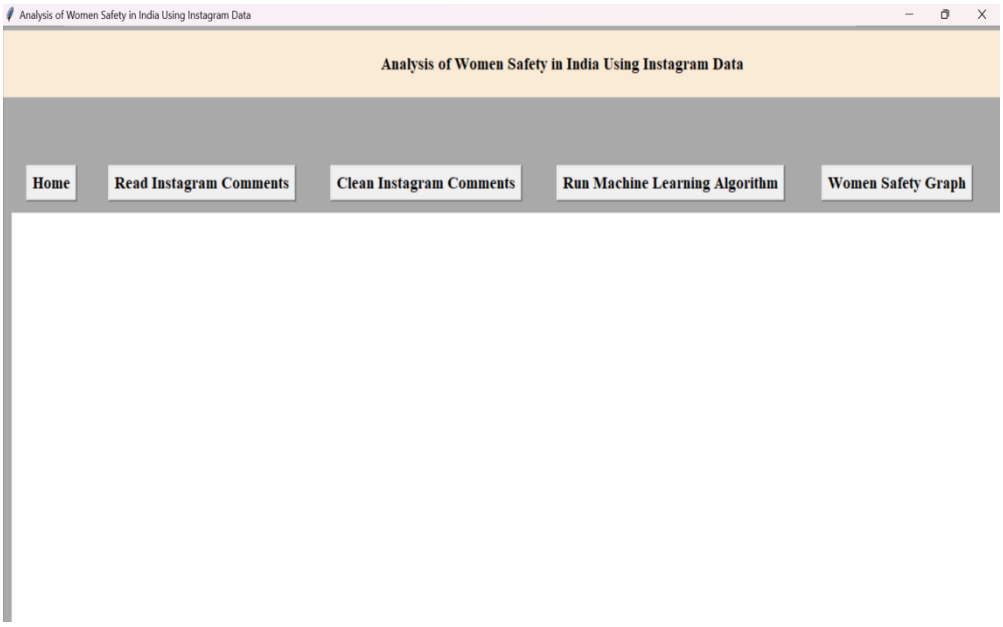
warning dialog box displayed with the message



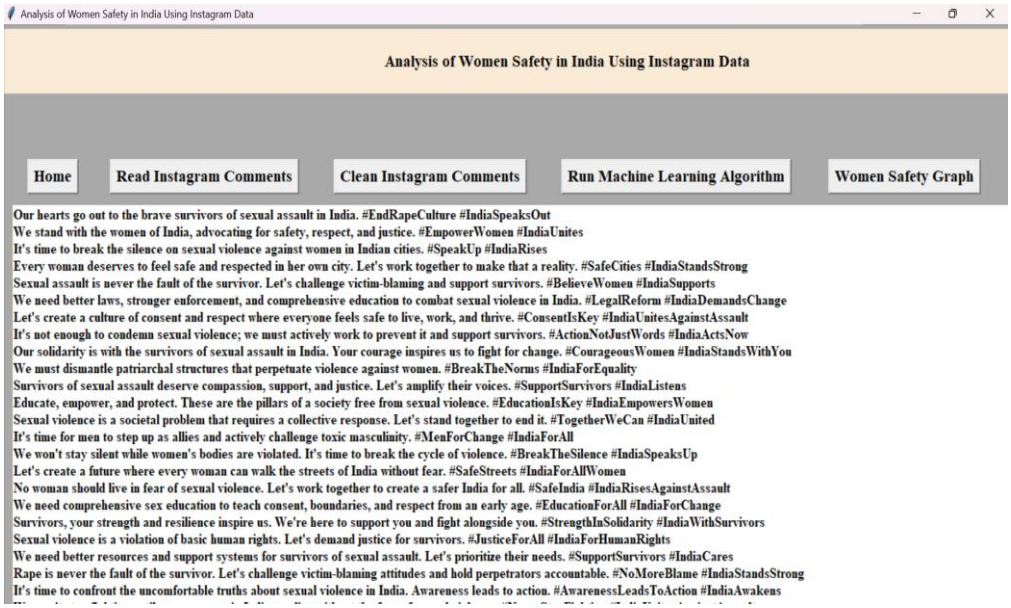
Select an option rape with combobox

Rape load images displayed

Second home page

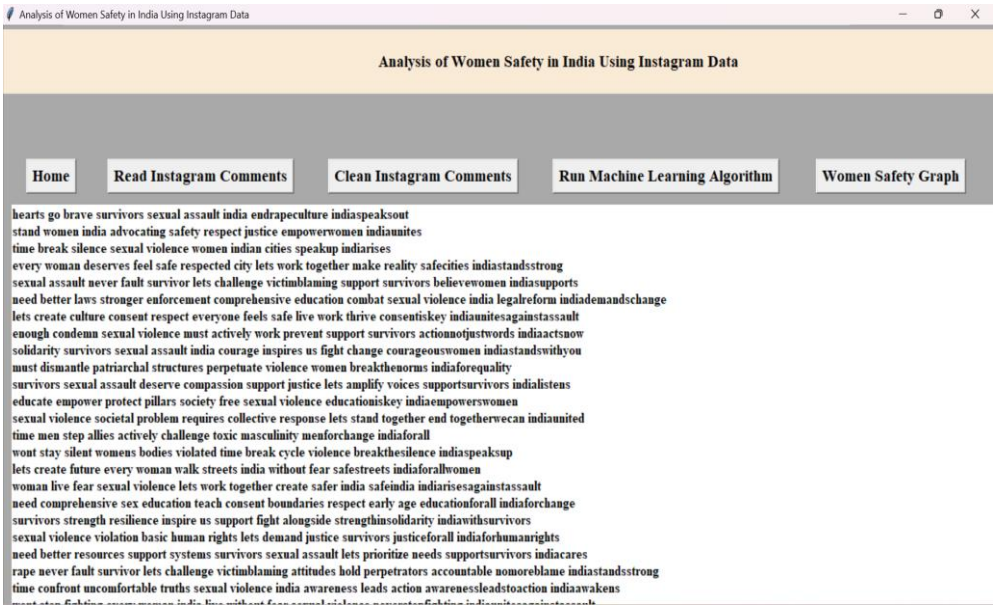


Read Instagram comments



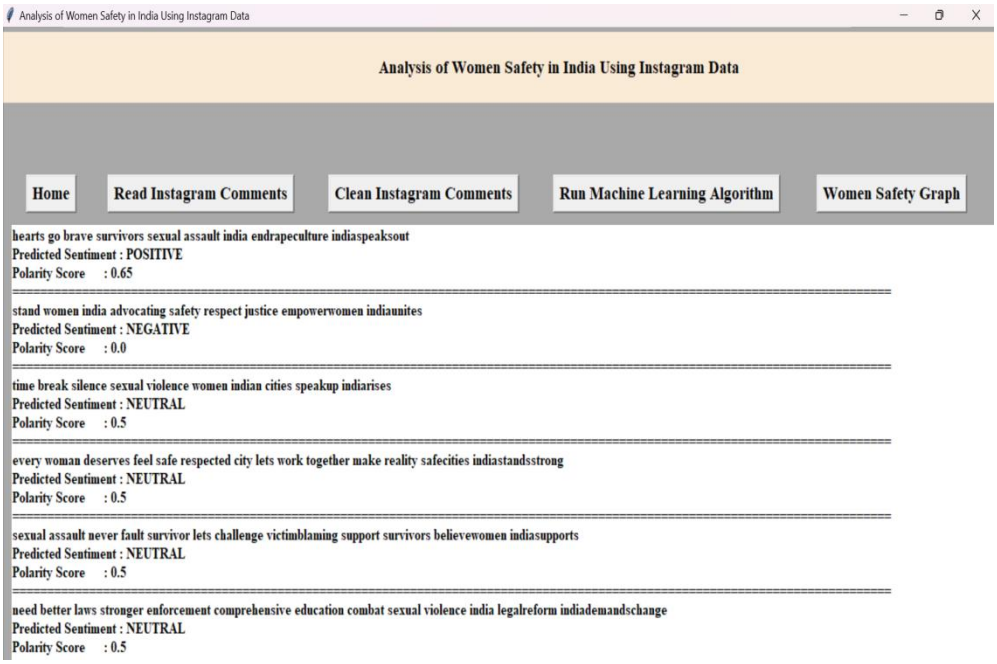
warning dialog box displayed with the message

Clean Instagram comments



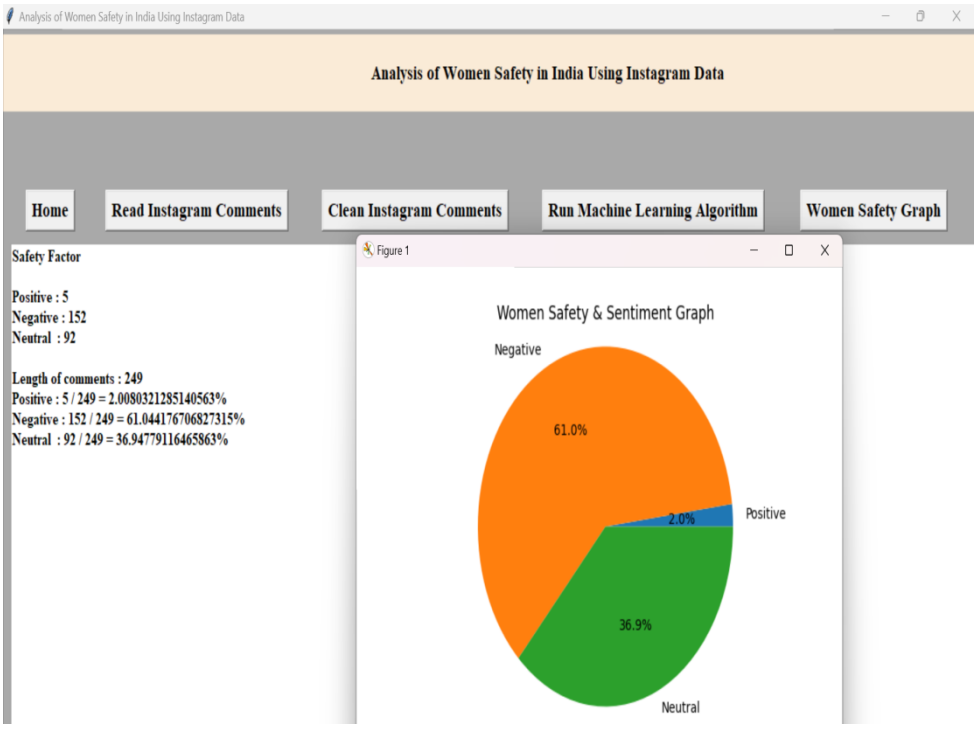
warning dialog box displayed with the message

Run machine learning algorithm



warning dialog box displayed with the message

Women Safety Graph



Thank You