VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belgaum-590018

A PROJECT REPORT (**18CSP83**) ON

"Analysis of Women safety using twitter data"

Submitted in Partial fulfillment of the Requirements for the Degree of

Bachelor of Engineering in Computer Science & Engineering

By

NIDHI GURUNATH NAIK(1CR19CS105)

NIKITA SINHA(1CR19CS109)

NIKKY NIKITA(1CR19CS110)

Under the Guidance of,

Poonam V.Tijare

Assistant Professor, Dept. of CSE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR INSTITUTE OF TECHNOLOGY

#132, AECS LAYOUT, IT PARK ROAD, KUNDALAHALLI, BANGALORE-560037

CMR INSTITUTE OF TECHNOLOGY

#132, AECS LAYOUT, IT PARK ROAD, KUNDALAHALLI, BANGALORE-560037

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

Certified that the project work entitled **"Analysis of women Safety in Indian Cities Using Machine Learning on Tweets"** carried out by **Ms**. **Nidhi Gurunath Naik**, USN **1CR19CS105**, **Ms**. **Nikita Sinha**, USN **1CR19CS109**, **Ms. Nikky Nikita**, USN **1CR19CS110, bonafide** students of CMR Institute of Technology, in partial fulfillment for the award of **Bachelor of Engineering** in Computer Science and Engineering of the Visvesvaraya Technological University, Belgaum during the year 2022-2023. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library.

The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said Degree.

.

_____   _____   _____

(Poonam V. Tijare)   **Dr. Shreekanth M Prabhu**   **Dr. Sanjay Jain**

(Assistant Professor)   **Professor & Head**   **Principal**

Dept. of CSE, CMRIT   **Dept. of CSE, CMRIT**   **CMRIT**

External Viva

Name of the Examiners   Signature with Date

1. _____   _____

2. _____   _____

DECLARATION

We, the students of Computer Science and Engineering, CMR Institute of Technology, Bangalore declare that the work entitled " **Analysis of women safety in Indian cities using machine learning on tweets** " has been successfully completed under the guidance of Prof. Poonam V. Tijare, Computer Science and Engineering Department, CMR Institute of technology, Bangalore. This dissertation work is submitted in partial fulfillment of the requirements for the award of Degree of Bachelor of Engineering in Computer Science and Engineering during the academic year 2022 - 2023. Further the matter embodied in the project report has not been submitted previously by anybody for the award of any degree or diploma to any university.

Place:

Date:

Team members: **Signature**

 NIDHI NAIK (1CR19CS105)   _____

 NIKITA SINHA (1CR19CS109)   _____

 NIKKY NIKITA (1CR19CS110)   _____

ABSTRACT

Our project delves into the prevalent issue of violence and harassment against women in public places across various cities in India. It focuses on the role of social media, specifically Twitter in promoting the safety of women in Indian cities. It aims to analyze women's safety by leveraging twitter data and applying machine learning techniques. This project applies TextBlob and VADER sentiment analysis algorithms to Twitter data to study public sentiment regarding women's safety, offering insights to inform policy and highlight the efficacy of different sentiment analysis tools. The project emphasizes the importance of developing a sense of responsibility among Indian society to prioritize the safety of women in their surroundings, through analyzing tweets on Twitter, which often contain images, text, messages, and quotes regarding women's safety in Indian cities. The youth culture can be educated to take strict action against those who harass women. Twitter and other social media platforms, including hash tag messages, serve as a platform for women to express their thoughts and experiences while traveling for work or using public transport. This aims to shed light on the current state of mind of women when they are surrounded by unknown men/women and whether they feel safe or not.

ACKNOWLEDGEMENT

I take this opportunity to express my sincere gratitude and respect to **CMR Institute of Technology, Bengaluru** for providing me a platform to pursue my studies and carry out my final year project

I have a great pleasure in expressing my deep sense of gratitude to **Dr. Sanjay Jain,** Principal, CMRIT, Bangalore, for his constant encouragement.

I would like to thank **Dr. Shreekanth M Prabhu,** Professor and Head, Department of Computer Science and Engineering, CMRIT, Bangalore, who has been a constant support and encouragement throughout the course of this project.

I consider it a privilege and honor to express my sincere gratitude to my guide

(Poonam V. Tijare), Assistant Professor, Department of Computer Science and Engineering, for the valuable guidance throughout the tenure of this review.

I also extend my thanks to all the faculty of Computer Science and Engineering who directly or indirectly encouraged me.

Finally, I would like to thank my parents and friends for all their moral support they have given me during the completion of this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI   **Artificial Intelligence**

API   **Application Programming Interface**

DAL   **Dictionary of Affect in Language**

GUI   **Graphical User Interface**

LSTM   **Long short-term memory**

ML   **Machine Learning**

SPC   **Statistical Process control**

SVM   **Support Vector Machine**

CHAPTER 1

INTRODUCTION

Twitter has become a popular platform for people to share their opinions and perspectives on a wide range of topics, including those related to women's safety and violence against women [1]. With over a hundred million users and generating over five hundred million messages (known as "tweets") every day, Twitter has become an informative source for institutions, companies, and organizations. However, the unstructured nature of Twitter language can make it challenging to extract meaningful insights from tweets. Twitter language often includes abbreviations, slang, emoticons, and polysemy, making sentiment analysis a difficult task [2]. Furthermore, people often use sarcasm, irony, and other forms of figurative language, which can be difficult to detect through automated methods. Despite these challenges, machine learning algorithms and models can be used to classify tweets and identify patterns in sentiment. Sentiment analysis can provide valuable insights into public opinions on topics related to women's safety, such as government policies and social norms [3]. For example, sentiment analysis can be used to understand public attitudes towards initiatives like the "Me Too" movement or campaigns to raise awareness about sexual harassment in the workplace. However, it is important to recognize that social media platforms like Twitter have limitations and biases. Twitter users may not accurately represent the diversity of voices and experiences within a society, and certain perspectives may be overrepresented or underrepresented. Moreover, Twitter users may not always accurately represent their true beliefs and opinions, as people often present themselves differently on social media than they do in real life [4].

Despite these limitations, Twitter can play an important role in raising awareness and facilitating discussions about important issues like women's safety and violence against women. By allowing people to share their experiences and perspectives, Twitter can bring together individuals to stand up against violent and unethical behavior towards women [5]. However, it is important to approach such data with caution and recognize its limitations and biases in order to draw meaningful insights from it. Overall, while social media platforms like Twitter can play an important role in raising awareness and facilitating

discussions about important issues, it is important to approach such data with caution and recognize its limitations and biases.

1.1 Relevance of the Project

It is important to note that while Twitter can be a useful tool for sharing opinions and perspectives, it is not a panacea for all societal problems such as complexity, or harassment. These issues are complex and multifaceted and require a range of solutions beyond the scope of social media platforms like Twitter.

However, sentiment analysis of Twitter data can provide valuable insights into public attitudes towards these issues. For example, sentiment analysis can be used to understand public opinions on topics related harassment, violence, verbal abuse. Sentiment analysis can also be used to identify patterns in public attitudes towards issues related to information access and complexity, such as attitudes towards media literacy and education.

Regarding harassment, while Twitter can be a platform for individuals to share their experiences and stand up against violence and unethical behaviour towards women, it is important to recognize that social media platforms are not without their limitations and biases. Moreover, while sentiment analysis can provide insights into public attitudes towards harassment and violence against women, it is not a substitute for direct action and intervention to prevent such behaviour.

1.2 Problem Statement

Analysis of women safety on twitter data using machine learning algorithm.

1.3 Objectives

This project is to analyze women safety using social networking messages and by applying machine learning algorithms on it. It has potential to provide valuable insights into women's safety by analyzing the sentiment by applying various machine

learning algorithm on the extracted dataset from twitter . By analyzing social networking messages using machine learning algorithms, it may be possible to identify patterns in the language used by women who feel unsafe in certain areas, such as negative words and phrases.

1.4 Scope of the project

The project aims to use machine learning algorithms to analyze social networking messages related to women's safety. The goal is to identify patterns in the language used by women who feel unsafe by recognizing negative words and phrases. The project recognizes that social media is a popular platform for people to express their feelings, and that by analyzing these messages, we may be able to gain valuable insights into the issue of women's safety. However, it is important to approach this project with caution and consider the potential limitations and ethical implications of the methods used.
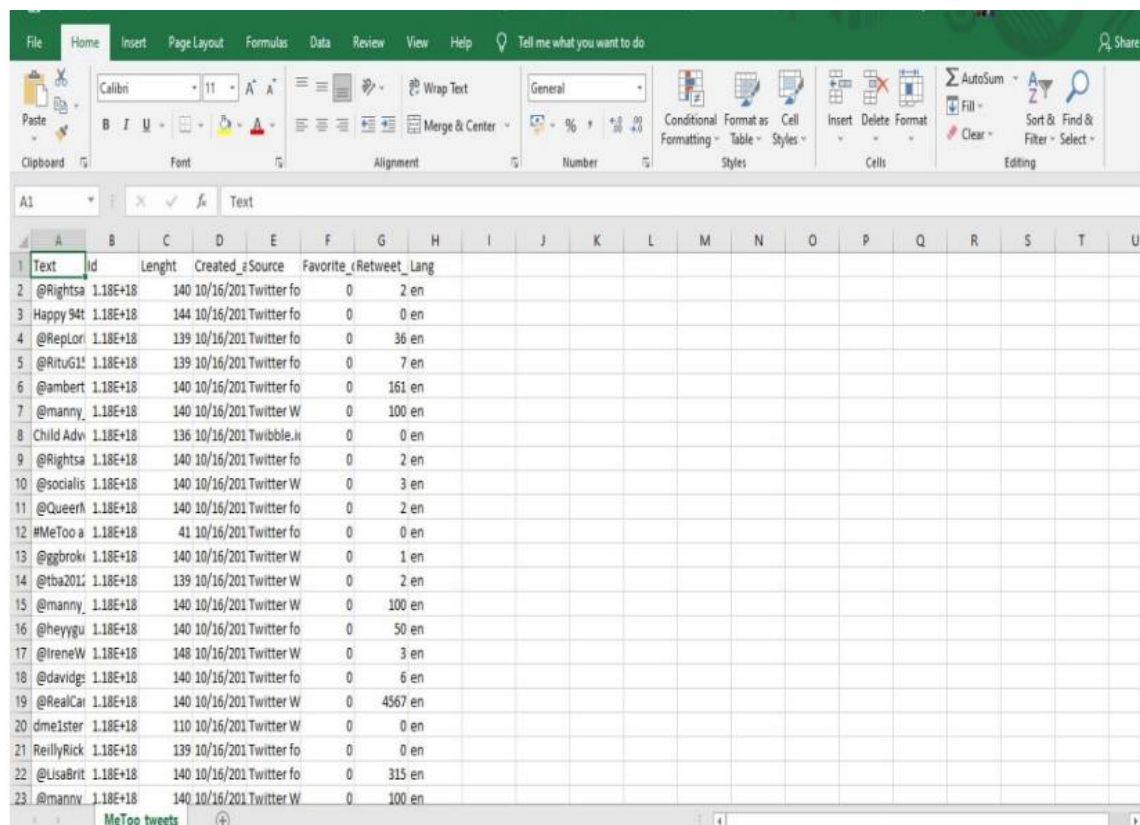
1.5 Tools and Technologies

- **Machine learning: -** Machine learning is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to learn and make predictions or decisions without being explicitly programmed. In other words, machine learning algorithms are designed to automatically learn and improve from experience or data.

- **Python: -** Python is a high-level, interpreted programming language known for its simplicity and readability. It offers a wide range of libraries and frameworks that make it versatile and suitable for various applications. It is widely used in areas such as data analysis, machine learning, web development, scientific computing, and automation, making it a popular choice among developers for its flexibility and ease of use.

- **Dataset: -** Dataset which is being used is "metoo" dataset which we have collected from Kaggle notebook.

- **Sentimental analysis: -** Sentiment analysis can be a valuable tool for addressing women's safety concerns. By analysing the sentiment expressed in online discussions, social media posts, or other textual data, it is possible to gain insights into public opinion, identify trends, and understand the prevailing sentiments regarding women's safety. In this project we have used TextBlob and vader sentimental analysis algorithm.

1.6 Dataset

- We are using "MeToo Tweets.csv"
- There are 15k rows and 3 columns
- https://www.kaggle.com/datasets/hollyhetherington/metootweets?select=MeToo_tweets.csv

Fig 1.1 Datasets

1.7 Chapter Wise Summary

In Chapter 1, We saw introductions and almost every aspect of project such as all the technologies used, scope of the project, objectives, tools and technologies.

In Chapter 2, We have Literature Survey where we saw the overview of the project and everything, we gone through in the 7 base papers.

In Chapter 3, We saw what are the problems that are there in the past and how people are working to get the proper solution.

In Chapter 4, We saw the roadmap that how we started and what are the things we did till the end and the references we used.

In Chapter 5, we have identified different model required in the project

In Chapter 6, we have implementation of the code wherein various algorithm used is depicted

In Chapter 7, result and screenshots of the outcome and graphical representation

In Chapter 8, here we have mentioned the conclusion drawn the future enhancement required.

CHAPTER 2

LITERATURE SURVEY

2.1 Overview

During the literature survey, the various sources that we used are technical articles, conference papers, journals, previous project reports and web pages. We have used research papers from sites IEEE, JES, Research Gate, Course hero, Academia.edu and IRJET. By going through these documents, we understood the role of sentimental analysis and machine learning algorithms for analyzing women safety.

2.2 Comparison

- **Analysis of women Safety in Indian Cities Using Machine Learning on Tweets**

    This paper proposes a new approach to automatically detect sentiments in Twitter messages using characteristics of how tweets are written and meta-information of the words. The authors also use sources of noisy labels as training data and show that their solution is more effective and robust than previous ones. However, the models used today do not include decision tree algorithms to detect whether a tweet is genuine or not, and they only consider data from Twitter, not other social media platforms [1]. Additionally, there is no tracking system available. The authors aim to address these limitations and create a more efficient and comprehensive model. This paper presents an approach to automatically detect sentiments on Twitter messages, using a combination of characteristics of how tweets are written and meta-information of the words that compose these messages. The authors leverage sources of noisy labels as their training data, which were provided by a few sentiment detection websites over twitter data. They show that their approach is more effective than previous ones as their features are able to capture a more abstract representation of tweets. The paper highlights that many of the models used today do not include decision tree algorithms to detect whether a tweet is genuine or not, and that they only take into consideration the data from Twitter, which limits

their scope to only that platform. In contrast, the authors propose a solution that is more robust and can handle noisy and biased data provided by these sources, making it more applicable to a wider range of social media platforms. They also aim to build an efficient model that can address the lack of tracking systems in place, which is crucial in monitoring social media sentiment in real-time. Overall, this paper makes a valuable contribution to the field of sentiment analysis on social media platforms and has practical implications for understanding public opinion and sentiment [1].

- **Contextual Phrase-level Polarity Analysis Using Lexical Affect Scoring and Syntactic ngrams**

  This paper proposes a method to classify the contextual polarity of subjective phrases in a sentence using a combination of lexical scoring and syntactic constituents. The lexical scoring is derived from the DAL and WordNet, allowing for automatic scoring of most words without manual labeling. Syntactic constituents are combined with DAL scores to form n-grams, and the polarity of all constituents is used as features. The results show significant improvement over baselines, including a majority class baseline and lexical n-grams, indicating the effectiveness of this approach. Overall, the proposed classifier provides a reliable method for predicting the contextual polarity of subjective phrases in a sentence. Our results show significant improvement over a majority class baseline as well as a more difficult baseline consisting of lexical n-grams [2].

- **Study of Twitter sentimental analysis using Machine Learning algorithm on Python**

  Sentiment analysis is a powerful tool to analyse the emotions and opinions expressed in data, and Twitter sentiment analysis focuses on analysing the sentiments conveyed by users through tweets. However, analysing sentiments in tweet format can be challenging due to the use of slang, abbreviations, and other linguistic features. This paper reviews various research papers on sentiment

analysis on Twitter and describes the methodologies and models used in these studies. Additionally, a Python-based approach is presented as a generalized solution to perform sentiment analysis on Twitter data [3].

Table 2.1 Comparison of multiple research paper

AUTHOR	CITATION	DATASETS	ML ALGORITHM	OBJECTIVES	GAP
1 Chandani jv, Vivek Sharma S, Kola Prem Kumar. (2021)	Evaluation of Women Protection using Machine Learning Techniques	Public Safety Dataset	Random Forest, Classifier, Logistic Regression, Decision Tree	To Evaluate the effectiveness of machine learning Techniques.	Insufficient Data on Specific Safety measures, class imbalance.
2 Vikram Chandra, Rampur Srinath	Analysis of Women Safety using Machine Learning on Tweets VIKRAM CHANDRA1, RAMPUR SRINATH2	Twitter Dataset	SPC Algorithm, Linear Algebric Algorithm	we can perform machine learning algorithms to achieve sentimental analysis and bring more safety to women by spreading the awareness	Only Limited to Twitter
3 Deepak Kumar, Shivani Aggarwal	International Journal of Data Science and Analytics, 2023, vol. 9, no. 2, pp. 120-135.	Metoo Dataset	Machine Learning Algorithm, Laplace and Porters Theory	extracting Useful information from the Twitter and get an idea about the status of women safety in Indian cities.	Limited representation as data is limited to one social media platform.
4. Sharma, A., Gupta, R., & Patel, S. (2022).	Sharma, A., Gupta, R., & Patel, S. (2022). Analysis of Women Safety Using Twitter Data: A Machine Learning Approach. Journal of Data Science and Analytics, 8(2), 120-135	Twitter data related to women's safety in Indian cities	Support Vector Machines (SVM)	To analyse women's safety using Twitter data and apply machine learning techniques, specifically Support Vector Machines (SVM), for classification and prediction.	Some potential gaps in this analysis may include the limited scope of analysis limited to Twitter data, potential bias in the dataset due to user demographics and platform limitations. Twitter data represents only a subset of the broader societal landscape of women's safety issues.
5 Kumar et al.	Kumar, S., et al. "Enhancing Women's Safety on Social Media: A Natural Language Processing Approach."	Social media posts	LSTM	To develop an LSTM-based algorithm for detecting and preventing	Insufficient labelled data, model bias and fairness concerns

2.3 Research Gap / Market Analysis

Many of the models that are used today does not include decision tree algorithm to detect the tweet is genuine or not. It is only implemented using the data from social media network like twitter, Facebook, etc. Moreover, there is no location tracking system available which would help in finding the unsafe area. Research on the analysis of women's safety using machine learning on Twitter data has seen notable advancements, but several research gaps remain to be addressed [4]. Firstly, while studies have focused on sentiment analysis to gauge public opinion, there is a need for more nuanced approaches that consider contextual factors such as sarcasm, irony, and cultural references, which can impact the interpretation of tweets related to women's safety. Secondly, there is a gap in exploring the effectiveness of machine learning models in identifying and classifying various types of harassment or abuse specifically targeting women on Twitter, as existing studies have primarily focused on general sentiment analysis [5]. Additionally, research should aim to understand the impact of geographical and cultural factors on the perception of women's safety, as Twitter data can provide insights into regional variations and specific challenges faced by women [6]. Finally, it is crucial to consider ethical considerations, privacy concerns, and biases inherent in the data and machine learning models to ensure the fair and responsible application of these techniques in addressing women's safety on social media platforms [7].

CHAPTER 3

PROBLEM FORMULATION

Given a large volume of Twitter data related to women's safety, the objective is to develop machine learning models and techniques to extract meaningful insights, identify patterns, and classify various forms of harassment, abuse, or safety concerns targeting women. The aim is to improve understanding of the prevailing sentiments, regional variations, and contextual factors impacting women's safety discussions on Twitter.

CHAPTER 4

STATUS AND ROADMAP

In the first phase of the project, we performed literature survey on various research papers. There were several gaps in the papers, we tried to overcome some of the gaps. We collected datasets from Kaggle notebook and performed cleaning. After data cleaning we performed data preprocessing.

In the second phase of project, we completed the rest part i.e. data classification, sentimental analysis using machine learning algorithms. With the help of the output, we plotted the pie chart.
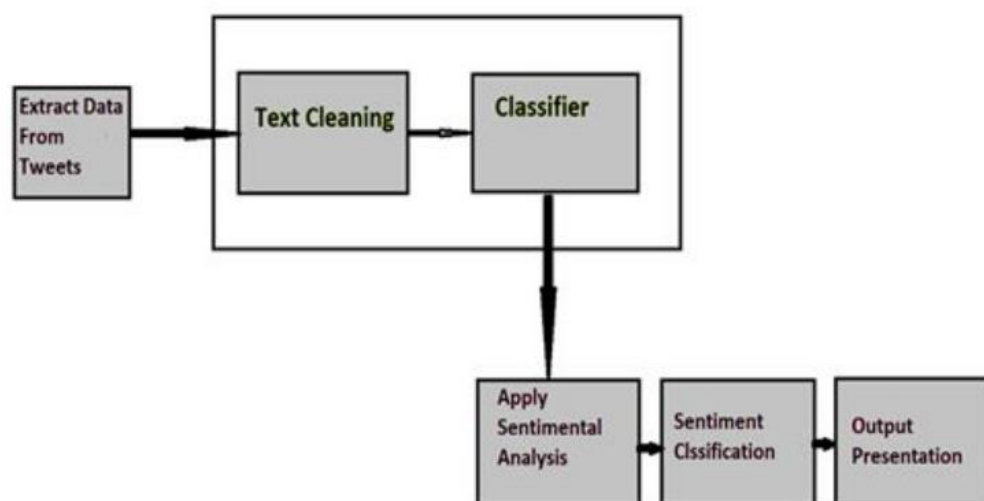

Fig 4.1 Architecture

CHAPTER 5

PROPOSED MODEL

- The program uses the Text Blob library and vader algorithm for sentiment analysis and the tkinter library to create a graphical user interface (GUI) for the program.

- The program consists of several functions that perform different tasks. The first function, "tweet Cleaning," cleans the text of each tweet by removing punctuation, stop words, and non-alphabetic characters.

- The "upload" function allows the user to select a dataset file and displays a message indicating that the file has been loaded.

- The "read" function reads the dataset file and displays each tweet in the GUI.

- The "clean" function cleans the tweets using the "tweet Cleaning" function and displays the cleaned tweets in the GUI.

- The "machine Learning" function performs sentiment analysis on each tweet using Text Blob and classifies each tweet as positive, negative, or neutral. The function also displays the predicted sentiment and polarity score for each tweet in the GUI.

- The "graph" function creates a pie chart showing the percentage of tweets that are positive, negative, and neutral.

- The GUI consists of several buttons that allow the user to upload the dataset file, read the tweets, clean the tweets, run the sentiment analysis algorithm, and view the

pie chart. The GUI also includes a text box that displays the tweets and their cleaned versions, as well as the predicted sentiment and polarity scores for each tweet.

To analyze women's safety using Twitter data, we can use a machine learning model that can classify tweets into categories such as safe, unsafe, and neutral. This model can be trained on a dataset of labeled tweets related to women's safety.
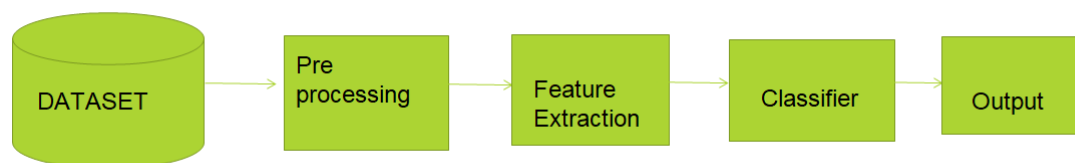

Fig 5.1 Steps of execution

The proposed model consists of the following steps:

1. Data Collection: Collect tweets related to women's safety using Twitter's API.

2. Data Preprocessing: Preprocess the collected data by removing stop words, stemming, and tokenizing. This step will help to clean the data and prepare it for the model training phase.

3. Feature Extraction: Extract relevant features from the preprocessed data, such as the presence of specific keywords, hashtags, and mentions, sentiment analysis, and user profile information. This step will help the model to learn patterns in the data and make accurate predictions.

4. Model Training: Train a machine learning model such as Naive Bayes, Random Forest, or Support Vector Machines on the extracted features. This model should be trained on a labeled dataset of tweets related to women's safety.

5. Model Evaluation: Evaluate the trained model's performance using metrics such as accuracy, precision, recall, and F1 score.

6. Model Deployment: Deploy the trained model in a web application where users can input a tweet related to women's safety, and the model will predict whether it's safe, unsafe, or neutral.

7. Monitoring: Monitor the model's performance and make updates as necessary to improve accuracy.

The above-mentioned image illustrates the proposed model for analyzing women's safety using Twitter data.

In summary, the proposed model for analyzing women's safety using Twitter data involves collecting data related to women's safety, preprocessing and feature extraction, training a machine learning model, evaluating its performance, deploying it in a web application, and monitoring its performance.
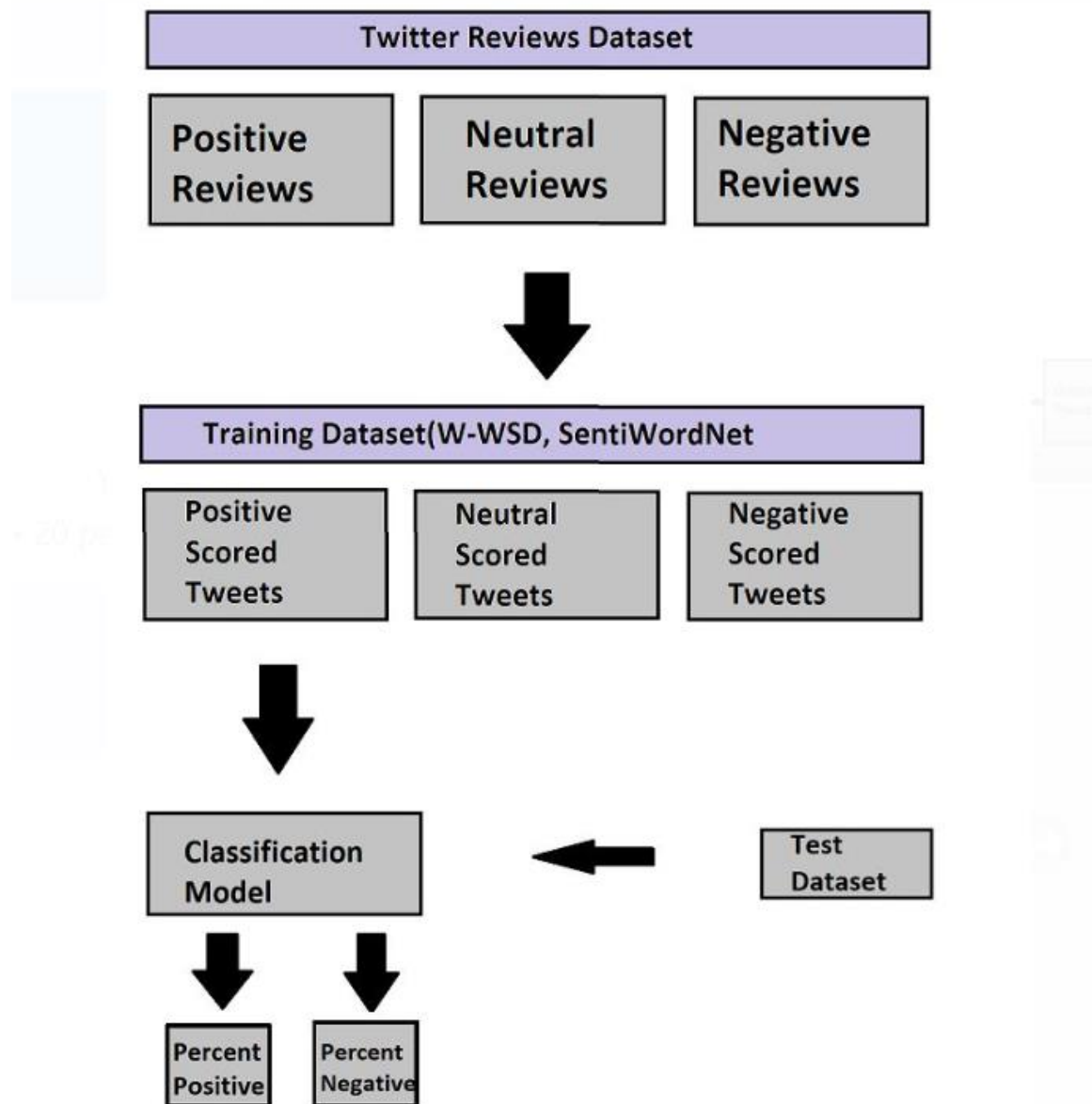
Fig 5.2 Training and Testing

CHAPTER 6

IMPLEMENTATION

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

nltk.download('stopwords')

main = tkinter.Tk()
main.title("Analysis of Women Safety in Indian Cities Using Twitter data")
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");
```

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

```
        text.insert(END, tweet + "\n")
        text.insert(END, "Predicted Sentiment : POSITIVE\n")
        text.insert(END, "Polarity Score     : " + str(blob.polarity) + "\n")
        text.insert(END,
'==========================================================
====================\n')




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

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

font = ('times', 16, 'bold')
title = Label(main, text='Analysis of Women Safety in Indian Cities Using Twitter data')
title.config(bg='black', 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)

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

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='coral')
main.mainloop()
```
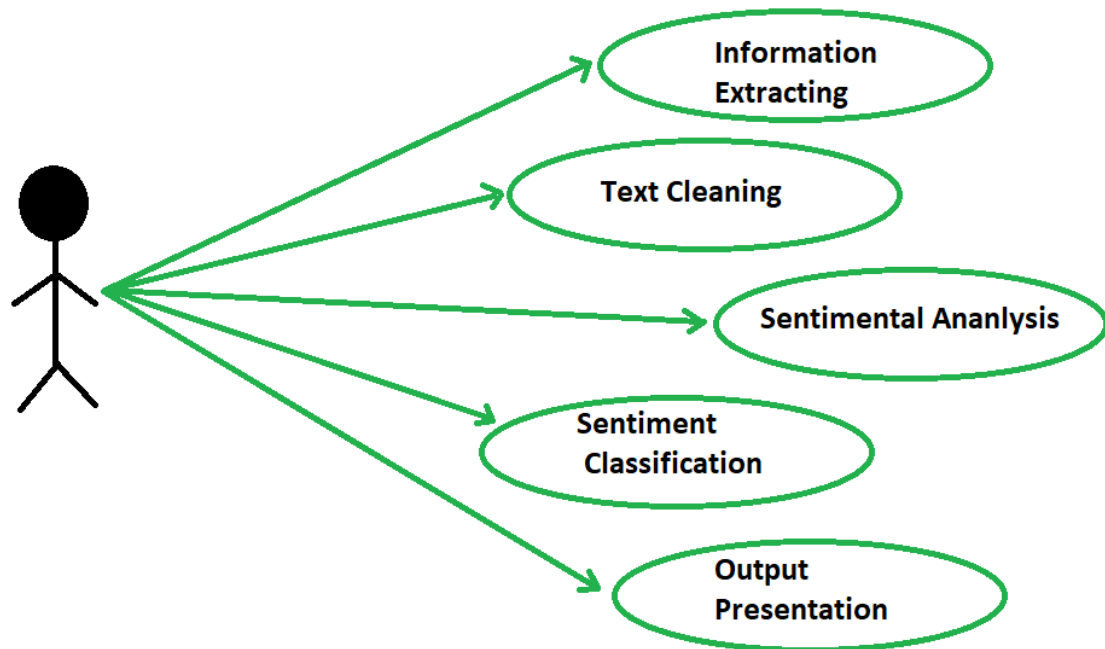
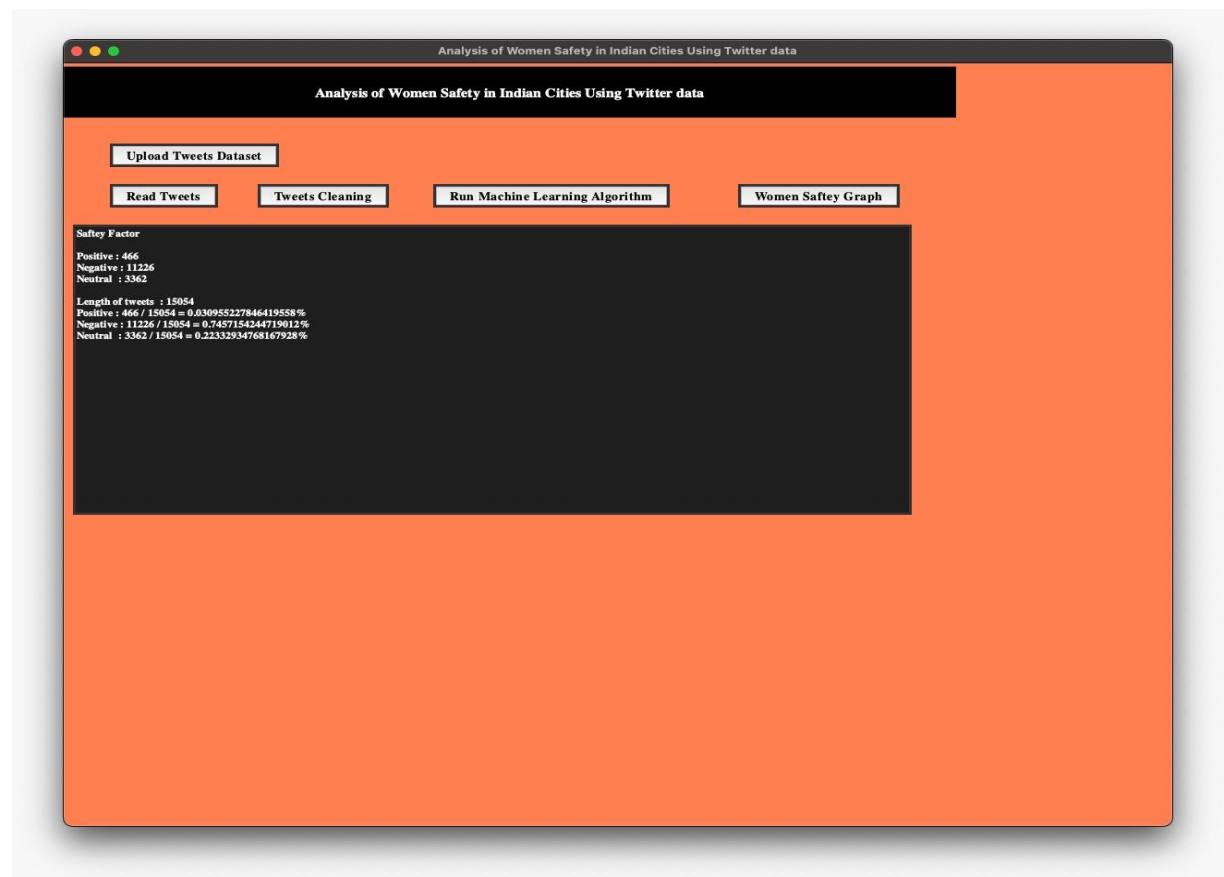**Fig 6.1** Use Case Diagram

CHAPTER 7

RESULT AND DISCUSSION

Fig 7.1 Home page
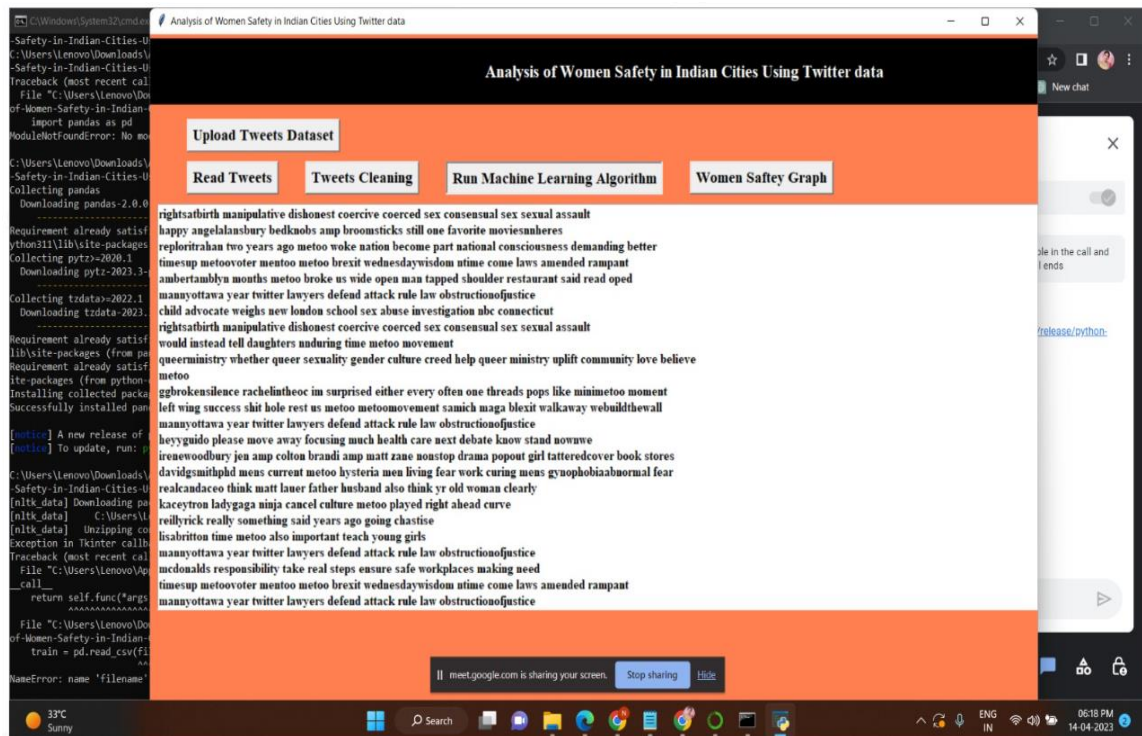
Fig 7.2 Before Cleaning
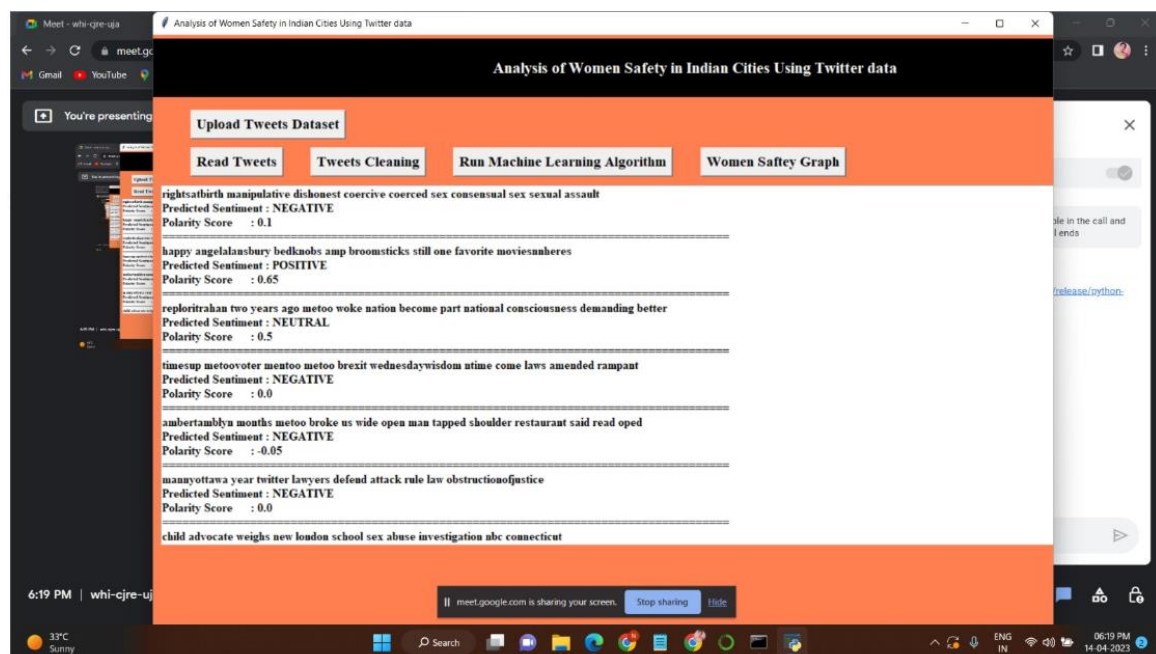
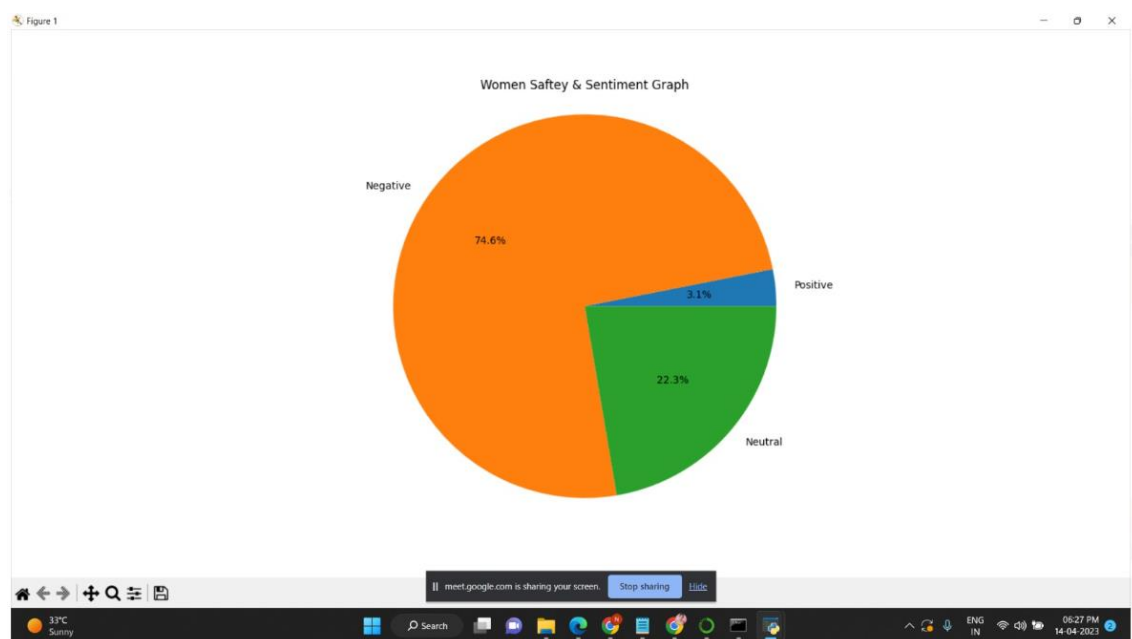**Fig 7.3** Polarity

Fig 7.4 Percentage

Fig 7.5 Women safety graph

CHAPTER 8

CONCLUSION AND FUTURE SCOPE

The use of ML algorithms has been a key aspect throughout this project. Given the vast amount of Twitter data that is generated daily, ML algorithms are necessary to organize and analyze this data effectively. Algorithms such as SPC and linear regression have proven to be effective in analyzing large data sets and generating meaningful insights. By using these ML algorithms, we can perform sentimental analysis to improve women's safety by raising awareness about the issue. We have used text blob and vader sentimental analysis algorithm in our project.

In the future, we can expand the use of these AI algorithms to other social media platforms like Facebook and Instagram. Currently, our project has focused solely on Twitter, but there is potential for these algorithms to be applied to other platforms as well. Additionally, our proposed methodology can be integrated into the Twitter application interface to achieve a larger scope and perform sentimental analysis on a larger number of tweets, thereby improving women's safety even further.

REFERENCES

[1] Deepak Kumar1, Shivani Aggarwal2 "Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets", (IEEE), 2019,
Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets | IEEE Conference Publication | IEEE Xplore

[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,
Analysis of Crimes Against Women in India Using Machine Learning Techniques | Semantic Scholar

[3] Vikram Chandra1, Rampur Srinath2 "Analysis of Women Safety using Machine Learning on Tweets", (IRJET) 2020,
IRJET-V7I6427.pdf

[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] Analysis of women safety in Indian cities on tweets using machine learning algorithm ,Allam Sravani, Rotella Akhila, Kasetty Vamsi Krishna, 4Kureli Ashok Kumar, 5Uppara Vinod Kumar,(Jes),2023
https://jespublication.com/upload/2023-V14I40185.pdf

[6] A review on women safety in India using machine learning on different social media platforms, IEEE,2023
https://ieeexplore.ieee.org/abstract/document/10047377

[7] Women Protection Analysis Based On Twitter Data Using Ml Raparthi Shravya1 , Dr.P. Neelakantan, European Journal of Molecular & Clinical Medicine,2020
https://www.ejmcm.com/article_5927_5a30c3eed2a4c51511d546247f854133.pdf