A Project Report on

Analysis Of Women Safety In Indian Cities

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

In

Computer Science and Engineering

Submitted by

M.Vaishnavi
(21H55A0514)

N.Yesumani
(21H55A0515)

Under the esteemed guidance of

Dr. G. Ravi Kumar

(**Associate Professor**)

Department of Computer Science and Engineering

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

(UGC Autonomous)

*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A^{+} Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2020- 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project Phase I report entitled **"Analysis of Women Safety in Indian cities"** being submitted by, M.Vaishnavi(21H55A0514), N.Yesumani(21H55A0515) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Dr. G. Ravi Kumar	**Dr. Siva Skandha Sanagala**	**EXTERNAL EXAMINER**
Associate Professor	**Associate Professor and HOD**	
Dept. of CSE	**Dept. of CSE**	

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express our heartfelt gratitude to all the people who helped in making this project a grand success.

We are grateful to **Dr. G. Ravi Kumar, Associate Professor**, Department of Computer Science and Engineering for his valuable Technical suggestions and guidance during the execution of this project work.

We would like to thank**, Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering & Technology, who is the major driving forces to complete our project work successfully.

We are very grateful to **Dr. Ghanta Devadasu**, Dean-Academics, CMR College of Engineering & Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering & Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary& Correspondent, CMR Group of Institutions, and **Shri Ch Abhinav Reddy**, CEO, CMR Group of Institutions for their continuous care and support.

Finally, we extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly or indirectly in completion of this project work.

M.Vaishnavi 21H55A0514
N.Yesumani 21H55A0515

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	LIST OF FIGURES	ii
	LIST OF TABLES	iii
	ABSTRACT	iv
1	**INTRODUCTION**	1
	1.1 Introduction	2-3
	1.2 Problem Statement	4
	1.3 Research Objective	4-5
	1.4 Project Scope and Limitations	6
2	**BACKGROUND WORK**	7
	2.1. Analysis of women safety in Indian cities using machine learning on tweets(Sentimental Analysis Algorithm)	8
	2.1.1.Introduction	8
	2.1.2.Merits & Demerits	9-10
	2.1.3.Implementation of women safety using ML	10-14
	2.2. Analysis of women safety in Indian cities using ML on tweets(NLTK)	15
	2.2.1.Introduction	15-16
	2.2.2.Merits,Demerits	17-18
	2.2.3.Implementation of women safety using NLTK	18-21
	2.3. Analysis of women safety in Indian cities using ML on tweets(Linear Regression)	22
	2.3.1.Introduction	22-23
	2.3.2.Merits,Demerits	23
	2.3.3.Implementation of Women safety using ML	24-25
3	**PROPOSED SYSTEM**	**26**
	3.1 Introduction	27
	3.2 Objective	28
	3.2.1 Algorithms Used for Proposed System	28

	3.3. Designing	30
	3.3.1.Architecture Diagram for Proposed System	31
	3.3.2. UML Diagram	32-35
	3.3. Stepwise Implementation and Code	36
4	**RESULTS AND DISCUSSION**	**37-42**
	4.1. Performance metrics	43
5	**CONCLUSION**	44
	5.1 Conclusion and Future Enhancement	45
	REFERENCES	46-47
	Git Hub Link	**48**
	Code	**49-55**
	Journal	
	Certificates	

List of Figures

FIGURE NO.	TITLE	PAGE NO.
2.1	Sentimental Analysis Architecture	11
2.2	Twitter Sentiment Analysis	18
2.3	System Architecture	19
2.4	Machine Learning Algorithms process steps	21
2.5	Twitter Analysis	24
3.1	Classifier of Sentimental Analysis	29
3.2	Architecture for proposed System	31
3.3	Use case Diagram	33
3.4	Sequence Diagram	34
3.5	Class Diagram	35
4.1	User Interface	38
4.2	Uploading Dataset	38
4.3	Data Loading	39
4.4	Reading the Tweets	40
4.5	Tweets cleaning	41
4.6	XG Boosting Algorithm	42
4.7	Women Safety Graph	43

List of Graphs

GRAPH NO.	TITLE	PAGE NO.
1	Women safety graph	43

ABSTRACT

Women and girls have been experiencing a lot of violence and harassment in public places in various cities starting from stalking and leading to abuse harassment or abuse assault. This research paper basically focuses on the role of social media in promoting the safety of women in Indian cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram. This paper also focuses on how a sense of responsibility on part of Indian society can be developed by the common Indian people so that we should focus on the safety of women surrounding them. Tweets on Twitter which usually contain images and text and also written messages and quotes which focus on the safety of women in Indian cities can be used to read a message amongst the Indian Youth Culture and educate people to take strict action and punish those who harass the women. Twitter and other Twitter handles which include hashtag messages that are widely spread across the whole globe sir as a platform for women to express their views about how they feel while we go out for work or travel in a public transport and what is the state of their mind when they are surrounded by unknown men and whether these women feel safe or not.

CHAPTER 1
INTRODUCTION

CHAPTER 1
INTRODUCTION

1.1 Introduction

Women's safety remains a critical issue in Indian cities, characterized by persistent incidents of harassment, assault, and violence in public spaces. Despite legislative measures and societal awareness campaigns, women continue to face significant challenges when navigating urban environments, with concerns ranging from street harassment to more severe forms of gender-based violence. Recognizing the urgency of addressing this pressing issue, there is a growing need for innovative approaches that leverage technology and data-driven methodologies to enhance women's safety.

This endeavors to delve into the intricate dynamics of women's safety in Indian cities through the lens of machine learning. By harnessing the power of advanced computational techniques, this research aims to develop a comprehensive understanding of the factors influencing women's safety, predict potential risks, and devise strategies to mitigate them effectively. Through the application of machine learning algorithms, we seek to analyze vast datasets encompassing various socio-economic, demographic, and spatial parameters to unearth patterns and insights crucial for devising targeted interventions. The safety of women in Indian cities has become a focal point of concern, with incidents of harassment, assault, and violence persisting despite ongoing efforts to address these issues. Social media platforms have become integral to public discourse, providing spaces for individuals to express opinions and share experiences. However, these platforms have also witnessed a troubling rise in abusive commentary and harmful behavior, particularly targeting women.

This employs machine learning techniques to analyze tweets related to women's safety in Indian cities. By leveraging the vast amount of data generated on platforms like Twitter, the analysis aims to uncover patterns, trends, and insights into the challenges faced by women in public spaces. This study represents a crucial step towards leveraging technology and data-driven approaches to address the pressing issue of women's safety in Indian cities. By harnessing the power of machine learning on social media data, we aim to contribute to the creation of safer and more equitable urban spaces for all residents, regardless of gender. Results of the sentimental analysis can be used in many areas like sentiments regarding a particular brand or release of a product, analyzing public opinions on the

government policies, people thoughts on women, etc. In order to perform classification of tweets and analyze the outcome, a lot of study has been done on the data obtained by the twitter. We also review some studies on machine learning in this paper and research on how to perform sentimental analysis using that domain on twitter data.

Social media websites platforms generally used by public to prompt their opinions but these platforms did not help out well to stop the abusive commentary on women. Women and girls have been experiencing a lot of violence and harassment in public places in various cities starting from stalking and leading to abuse harassment or abuse assault. The role of social media in promoting the safety of women in Indian cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram. "In Indian cities, women face significant challenges related to safety, including incidents of harassment, assault, and violence. The objective of this study is to develop a comprehensive, data-driven, and machine learning-based framework to understand the dynamics of women's safety in urban areas. This framework aims to predict and prevent incidents, identify high-risk areas, and improve the overall safety and security of women." Many incidents of violence and harassment against women and girls have occurred in public locations in different cities, starting with stalking and progressing to sexual harassment or sexual assault. Girls are harassed most often for reasons related to safety or a lack of tangible consequences in their lives. Instead of placing limits on women, society should understand the need of protecting them and that women and girls have the same right to safety in the city as men haveI instances of violence and harassment against women and girls in public spaces are alarmingly commonplace across different cities in India. From relentless stalking to egregious acts of sexual assault, women bear the brunt of these abhorrent behaviors. Too often, girls are targeted due to prevailing societal attitudes regarding safety or the glaring absence of tangible consequences for perpetrators.

1.2. Problem Statement

Effort estimation in software development is a critical challenge faced by the software engineering community. Reliable estimation is fundamental for project scheduling, resource allocation, cost estimation, and minimizing the risk of project failures or delays. Despite its importance, a significant number of software projects experience effort or schedule overruns, leading to project failures. Research surveys indicate that inaccurate estimation models are a primary cause of these overruns. The complexity of software projects often leads to vagueness in the early stages, making accurate estimation difficult. Furthermore, each project possesses unique characteristics, making it even harder to estimate the required effort for completion. Existing efforts to address this challenge suggest the development of adaptable estimation models capable of accommodating a wide range of project types. However, the uncertain and variable nature of software projects, coupled with small and often incomplete datasets, intensify the complexity of prediction tasks.

1.3. Research Objective

The primary aim of this study is to employ machine learning techniques for the analysis of tweets concerning women's safety in Indian cities. This research endeavors to achieve several specific objectives. Firstly, it seeks to comprehend the prevailing discourse and sentiments surrounding women's safety on social media platforms, particularly Twitter, within the unique context of urban India. Secondly, the study aims to identify patterns, trends, and recurring themes in tweets pertaining to incidents of harassment, assault, and violence against women in urban areas. Thirdly, it aims to develop machine learning algorithms capable of categorizing tweets based on the nature and severity of safety concerns expressed by users. Furthermore, the research intends to explore the geographic distribution of tweets discussing women's safety within Indian cities. Moreover, the study endeavors to assess the effectiveness of existing interventions or initiatives aimed at addressing women's safety concerns, as reflected in social media discussions.

It also aims to generate predictive insights that can inform proactive measures and policy interventions to enhance women's safety in Indian cities. Lastly, the research will evaluate the potential biases or limitations of using social media data for analyzing women's safety issues and identify strategies for mitigating these challenges.

Ultimately, this study aspires to contribute to the development of data-driven approaches and tools for addressing gender-based violence and promoting women's safety in urban environments. The objective of this research is to employ advanced machine learning methodologies to conduct a comprehensive and nuanced analysis of women's safety in urban areas of India. With a focus on leveraging data-driven approaches, this study aims to delve deeply into the multifaceted aspects of women's safety, considering various socio-economic, infrastructural, and cultural factors that contribute to the prevalence of violence and harassment against .By aggregating and analyzing diverse datasets encompassing crime reports, urban infrastructure data, demographic information, and public perception surveys, this research seeks to identify underlying patterns, trends, and hotspots of safety concerns across different Indian cities. Through a combination of statistical modeling, pattern recognition, and predictive analytics, the study aims to develop a robust framework for assessing the efficacy of existing safety measures and infrastructure interventions. Furthermore, the research endeavors to uncover potential correlations between women's safety and factors such as access to public transportation, urban design, socioeconomic disparities, and cultural norms. By elucidating these relationships, the study intends to provide actionable insights and evidence-based recommendations to stakeholders, including policymakers, urban planners, law enforcement agencies, and community organizations. Ultimately, the goal of this research is to contribute to the development of more effective strategies and interventions aimed at creating safer and more inclusive urban environments for women across India, thereby fostering greater gender equity and social justice.

1.4 Project Scope And Limitations

This project aims to analyze women's safety in Indian cities through the lens of social media data, particularly Twitter, using machine learning techniques. The scope encompasses collecting and preprocessing a large corpus of tweets related to women's safety issues in Indian cities. The analysis will focus on identifying trends, sentiments, and prevalent topics surrounding women's safety concerns, such as harassment, assault, and discrimination. Machine learning models will be developed to classify tweets based on their relevance to different aspects of women's safety, including location-specific incidents, time trends, and severity levels. Additionally, sentiment analysis will be employed to gauge the public perception and emotional responses towards women's safety issues. The project will also explore the potential of geospatial analysis to map the distribution of safety-related tweets across different regions and correlate them with known crime statistics and urban infrastructure data. The findings will be used to generate insights into the dynamics of women's safety concerns in Indian cities and inform the development of targeted interventions and policy recommendations. The project will adhere to ethical guidelines regarding data privacy and consent, ensuring that all data collection and analysis processes are conducted responsibly and transparently. Furthermore, the project will seek to engage with relevant stakeholders, including government agencies, non-profit organizations, and community groups, to ensure the relevance and impact of the research outcomes in addressing real-world challenges related to women's safety in Indian cities.

MERITS

- Analysis of twitter texts collection also includes the name of people and name of women who stand up against abuse harassment and unethical behaviour 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.

DEMERITS

- Applications like twitter and Instagram are used by utmost of the people to come up and put forward their views.
- Few elegances of soppiness that can be distributed like contraption literacy mongrel

CHAPTER 2
BACKGROUND WORK

CHAPTER 2

BACKGROUND WORK

2.1. Analysis ofwomen safety in Indian cities using machine learning on tweets

2.1.1 Introduction

[1] Women and adolescents partake in a lot of violence and importunity in civic spaces in Indian cities early from annoyance and foremost to erotic importunity or bodily crime. This principally emphases on the part of societal mass broadcasting in endorsing the security for women in India with distinct orientation to the part of societal broadcasting internet sites and operations together with Twitter application, Facebook and Instagram. [2] Tweets on Twitter which generally comprehends images and textbook and also printed dispatches and quotations which concentrate on the security of women in Indian metropolises can be castoff to recite a communication between the Indian Youth Culture and instruct people to yield stringent feat and discipline individuals who kill the womenfolk. Twitter and other Twitter handles which comprise mess label dispatches that are extensively feast across the entire globe is a stage for women to prompt one's views around which they sense while they drive out for work or trip in a civic conveyance and the state of particular's observance when individuals are girdled by mysterious menfolk and whether these womenfolk sense harmless or not. 'SENTIMENTAL ANALYSIS CLASSIFICATION' is an approach where tweets are reported and chosen up from Twitter Application. Collection of data, Data Preprocessing, point birth, choose the base structures, discovery of sentiments and bracket of opinions using machine learning procedures or modest controls are the introductory way to achieve sentimental analysis.

[3] Twitter in this contemporary period has surfaced as an eventual micro blogging communal network conforming over other media. As People interconnect and partake their view over hundred million stoners and induce over five hundred laboriously on public medias together with Facebook and million dispatches known as 'Tweets' every day. Twitter social system can be measured as a seamless place to know around people's estimation. For the purpose of executing bracket of tweets and dissect the outgrowth, a proportion of homework has been ended

on the information attained by the application. Few revisions on machine learning in this paper are analyzed and surveyed on how to achieve sentimental analysis by means of information sphere on twitter application data. On social websites, individuals can at liberty prompt what they sense about the Indian policies, civilization and plentiful other studies. also, women can similarly partake their gests if they've challenged any ferocity or sexual importunity and this fetch guiltless individuals composed to position up in contrast to similar occurrences.

Objective

As People interconnect and partake their estimation laboriously on societal websites as well as Facebook application and Twitter application, societal websites can be measured as a flawless stage to learn about individuals view and sentiments concerning different events. There be present several opinion-accustomed material congregation and statistical systems that goal to prize people's estimation concerning dissimilar motifs. Meanwhile Twitter comprises brief textbooks, persons incline to practice diverse disputes and bowdlerization. Prompts are delicate to prize it's soppiness by existing Natural Language Processing systems fluently. Thus, numerous experimenters have used deep literacy and machine literacy ways to prize and booby-trap the opposition of the promotions

2.1.2. Merits

- The fact set that was attained over Twitter is trained and provides accurate information regarding the position of safety of women in economic civilization
- The classification set cleanses the trained model by eradicating spare hashtags like word stoppage, feelings that are directive in making definite non documentary and it is gratified, linked and uninvolved formerly the investigation.
- Study of twitter textbooks assortment also comprises the designation of people

Demerits:

- Applications like twitter and Instagram are used by utmost of the people to come up and put forward their views.

- Few elegances of soppiness that can be distributed like contraption literacy mongrel and wordbook- grounded literacy
- Due to this there were arithmetical, knowledge- grounded and age wise isolation tactics.

2.1.3 Implementation

- SENTIMENTAL ANALYSIS CLASSIFICATION' is an approach that reports the tweet messages selected up from Twitter Application handed by application itself. Owed at the existence of Twitter application, here exists numerous ways accessible for sentimental investigation of statistics on social websites.
- A Depressed commerce G- graph is engendered through some social graph-model, diminishing the distance among the factual and Dejected commerce charts.
- Realities and their relations in social mass media are linked, and a commerce graph is erected.
- The graph displays the chance regarding different motifs. Since Twitter contains positive, negative and neutral tweets that the short handbooks, people tend to use different words and trained set of data contained condensation.

  These promptions are delicate to prize their sentiments by current NLP systems easily.

a. System Architecture

Every stoner data similar as identifications, original tweet texts, re-tweeted texts and polarity ratio are stored at the back end by admin to cover and accomplishment of study. The sentiment scrutiny is pragmatic at stoner's data acceptable to refuge and authorize if any tweet texts are vituperative to womanhood or any other. Admin implements this study on every stoner tweet texts to bounce women's safety. Sentimental analysis is prescribed on the tweets of stoner that is deposited at back end.

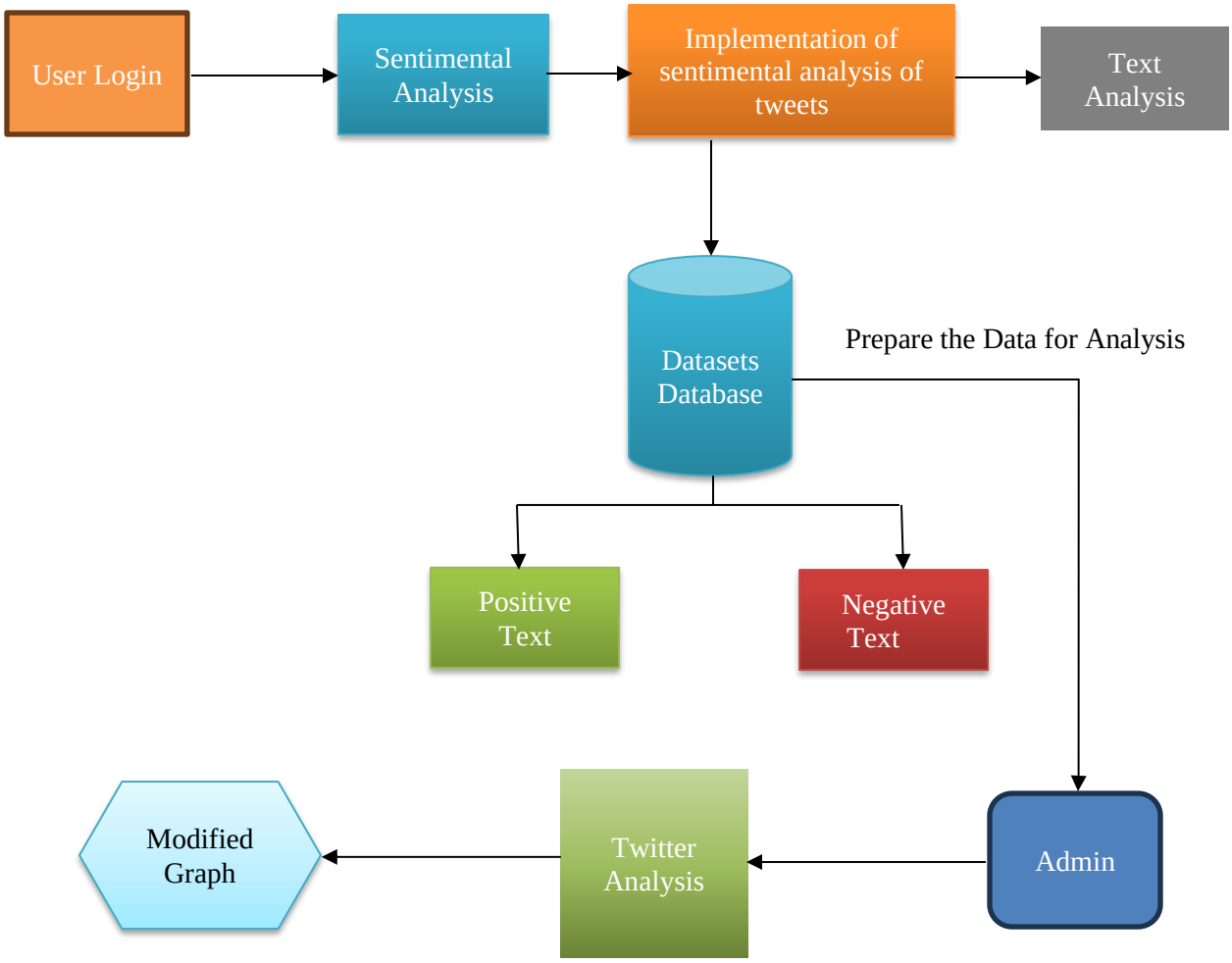


Fig No 2.1 : System Design Architecture

analysis. Similarly, under positive sentimental analysis there are a list of spotless and abusive tweets. End to end through the tweet framework, handler particulars are provided at end of investigation.

b. Algorithm :

SENTIMENTAL ANALYSIS Sentiment analysis is the process of extracting the sentiment behind any sentence or statement. It can be called as a classification technique which is used to

Algorithm 1 Extract Twitter sentiment

```
procedure TWITTER-CONNECTION()
consumer key='XXXXXXXXX'
consumer secret = xxXXXXXX
access token=xxxxxxxx
access token secret = xxxxXXXXXX
self.auth = OAuth Handler(consumer key, consumer secret)
self.auth.set access token(access - token, access token - secret)
self.api = tweepy.API(self.auth)
end procedure
procedure TWEET-CLEANING(t)
tweet = t.remove Stop words
Return tweet
end procedure
procedure TWEET-CLASSIFICATION(t)
t=Tweet-Cleaning(t)
tweet polarityt.sentiment.polarity
tweet - polarity
end procedure
```

Obtain the opinion from tweet. This opinion is useful in formulating a sentiment which can further be used to achieve sentiment classification. Sentiments are personal to the topic and thus we need to decide what kind of specifications is formulated out of it. Person performing the sentimental analysis wants to find the session of units of the tweets using the user interface design prototypical model. The dimension of the sentimental class is an important factor in order to decide the efficiency of the algorithm.

For instance, there can be two class sentimental classification of tweets – Positive and Negative or there can be three class classification – Positive, Negative.and Neutral.

In this paper, we use machine learning approach. Collection of statistics, pre-processing the information, extraction of features, choosing base features, detection of sentiments and classification of sentiments by means of machine learning approaches or simple calculations are the basic steps to execute sentimental analysis.

c. Modules

5 Modules are used to run this project.

- Upload Tweets Trained set of data()
- Read Tweets()
- Tweets Cleaning()
- Run Machine Learning Algorithm()
- Women Safety Graph()

d. Module Description

- Upload Tweets Trained set of data: Upload Tweets Trained set of data Module is Used to upload tweets Trained set of data.

- Read Tweets: Read Tweets Module is used to read tweets from trained set of data.

- Tweets Cleaning: This Module is used to see all tweets comprising distinct ciphers that halt

e. Sentimental Analysis

Sentiment analysis, or opinion mining, is the process of analyzing large volumes of text to determine whether it expresses a positive sentiment, a negative sentiment or a neutral sentiment. With more ways than ever for people to express their feelings online, organizations need powerful tools to monitor what's being said about them and their products and services in near real time. As companies adopt sentiment analysis and begin using it to analyze more conversations and interactions, it will become easier to identify customer friction points at every stage of the customer journey.

Sentiment analysis uses and machine learning (ML) technologies to train computer software to analyze and interpret text in a way similar to humans. The software uses one of two approaches, rule-based or ML—or a combination of the two known as hybrid. Each approach has its strengths and weaknesses; while a rule-based approach can deliver results in near real-time, ML based approaches are more adaptable and can typically handle more complex scenarios.

f. **Machine learning sentiment analysis**

With a machine learning (ML) approach, an algorithm is used to train software to gauge sentiment in a block of text using words that appear in the text as well as the order in which they appear. Developers use sentiment analysis algorithms to teach software how to identify emotion in text similarly to the way humans do. ML models continue to "learn" from the data they are fed, hence the name "machine learning". Here are a few of the most commonly used classification algorithms:

- Linear regression: A statistics algorithm that describes a value (Y) based on a set of features (X).
- Naive Bayes: An algorithm that uses Bayes' theorem to categorize words in a block of text.
- Support vector machines: A fast and efficient classification algorithm used to solve two-group classification problems.
- The twitter information that comprises of tweets and notices every single day, machine literacy set of rules services to unify to accomplish analysis. SPC algorithm, direct arithmetical are a few algorithms that are operative in examining the excessive data that give classification and translate into prompted trained set of data that provides 90% delicacy.
- All the mentioned algorithms are veritably operative and beneficial to assay bulky quantum of data together with the SPC algorithm and direct statistical featured methods that service to later classify the statistics into promotive groups. So, we can execute machine literacy algorithms to attain novelettish analysis and convey further protection to women by scattering the mindfulness.

2.2 Analysis Of Women Safety In Indian Cities Using Machine Learning On Tweet

2.2.1 Introduction

[4] Women and girls have been experiencing a lot of violence and harassment in public places in various cities starting from stalking and leading to abuse harassment or abuse assault. This research paper basically focuses on the role of social media in promoting the safety of women in Indian cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram. This paper also focuses on how a sense of responsibility on part of Indian society can be developed the common Indian people so that we should focus on the safety of women surrounding them. [5] Tweets on Twitter which usually contains images and text and also written messages and quotes which focus on the safety of women in Indian cities can be used to read a message amongst the Indian Youth Culture and educate people to take strict action and punish those who harass the women. [6] Twitter and other Twitter handles which include hash tag messages that are widely spread across the whole globe sir as a platform for women to express their views about how they feel while we go out for work or travel in a public transport and what is the state of their mind when they are surrounded by unknown men and whether these women feel safe or not?.

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 travelling 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 like 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 neighbours 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 behaviour 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 smoothening 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.

Machine Learning:

Machine learning (ML) is a branch of artificial intelligence (AI) and computer science that focuses on the using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy.

How does machine learning work?

UC Berkeley (link resides outside ibm.com) breaks out the learning system of a machine learning algorithm into three main parts.

1. A Decision Process: In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labeled or unlabeled, your algorithm will produce an estimate about a pattern in the data.
2. An Error Function: An error function evaluates the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.
3. A Model Optimization Process: If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this iterative "evaluate and optimize" process, updating weights autonomously until a threshold of accuracy has been met.

Twitter in this modern era has emerged as a ultimate microblogging social network consisting over hundred million users and generate over five hundred million messages known as 'Tweets' every day. Twitter with such a massive audience has magnetized users to emit their perspective and judgemental about every existing issue and topic of internet, therefore twitter is an informative source for all the zones like institutions, companies and organizations. On the twitter, users will share their opinions and perspective in the tweets section. This tweet can only contain 140 characters, thus making the users to compact their messages with the help of abbreviations, slang, shot forms, emoticons, etc. In addition to this, many people express their opinions by using polysemy and sarcasm also.

Hence twitter language can be termed as the unstructured. From the tweet, the sentiment behind the message is extracted. This extraction is done by using the sentimental analysis procedure.

2.2.2 Merits & Demerits

Merits:

- Analysis of twitter texts collection also includes the name of people and name of women who stand up against abuse harassment and unethical behaviour 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.

Demerits:

- Generalization Limitations: Findings derived from social media data may not be generalizable to the broader population or specific geographic regions. Extrapolating insights from social media to real-world women's safety concerns requires careful consideration of context and limitations.

- Validation and Interpretation Challenges: Validating the accuracy of machine learning models trained on social media data for predicting women's safety concerns can be challenging due to the subjective nature of safety perceptions. Interpreting

the model outputs in relation to real-world safety interventions may require additional expertise and qualitative research.

2.2.3 Implementation

Concept to analyze women safety using social networking messages and by applying machine learning algorithms on it. Now-a-days almost all peoples are using social networking sites to express their feelings and if any women feel unsafe in any area then she will express negative words in her post/tweets/messages and by analysing those messages we can detect which area is more unsafe for women's. But women feel that they are unsafe in places like malls, shopping malls on their way to their job location because of the several unknown Eyes body shaming and harassing these women point.

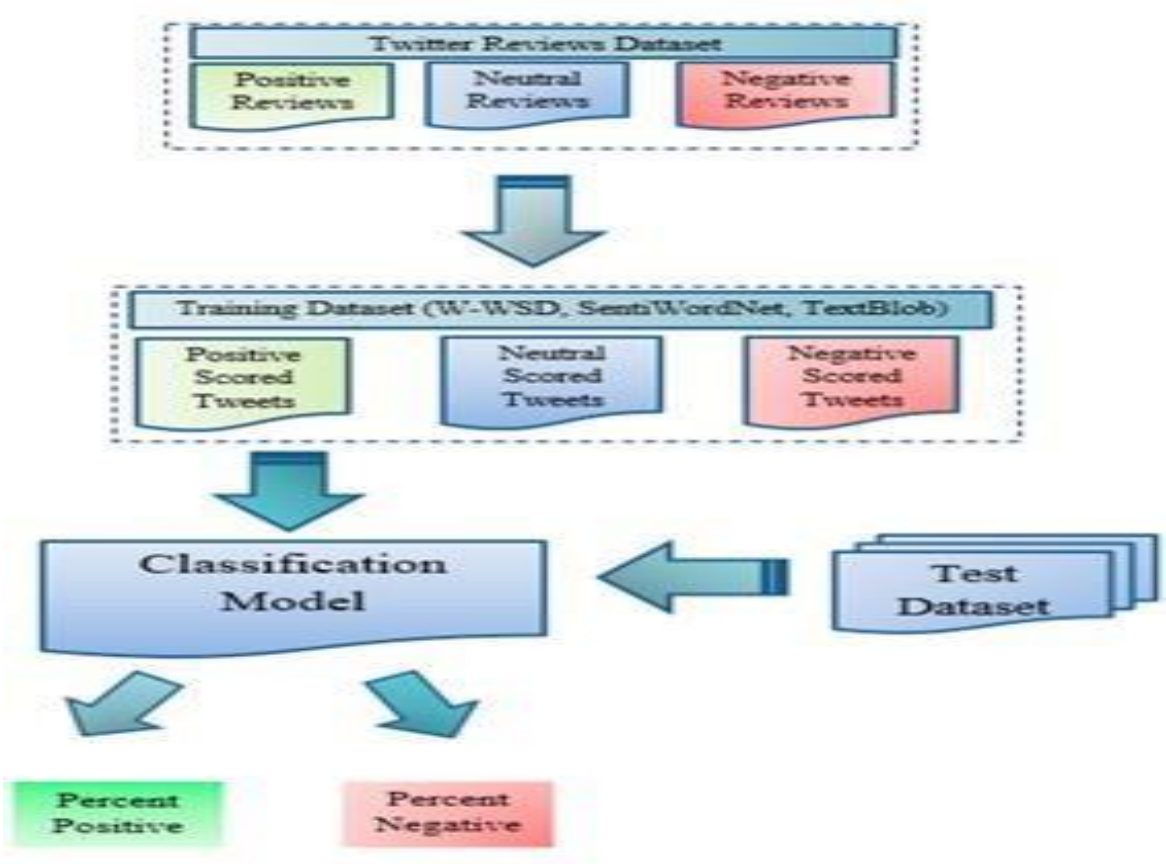


Fig 2.2 : Twitter sentiment analysis

In propose work author using TWEEPY package from python to download tweets from twitter but every time INTERNET will not available to download tweets online so we downloaded MEETOO tweets on women safety and safe inside dataset folder. Application will read this tweets to detect women's sentiments. Author using NLTK (natural language tool kit) to remove special symbols and stop words from tweets and to make them clean. Author using TEXTBLOB corpora package and dictionary to count positive, negative and neutral polarity and the tweets which has polarity value less than 0 will consider as negative as and greater than 0 and less than 0.5 will consider as neutral and polarity greater than 0.5 will consider as positive.

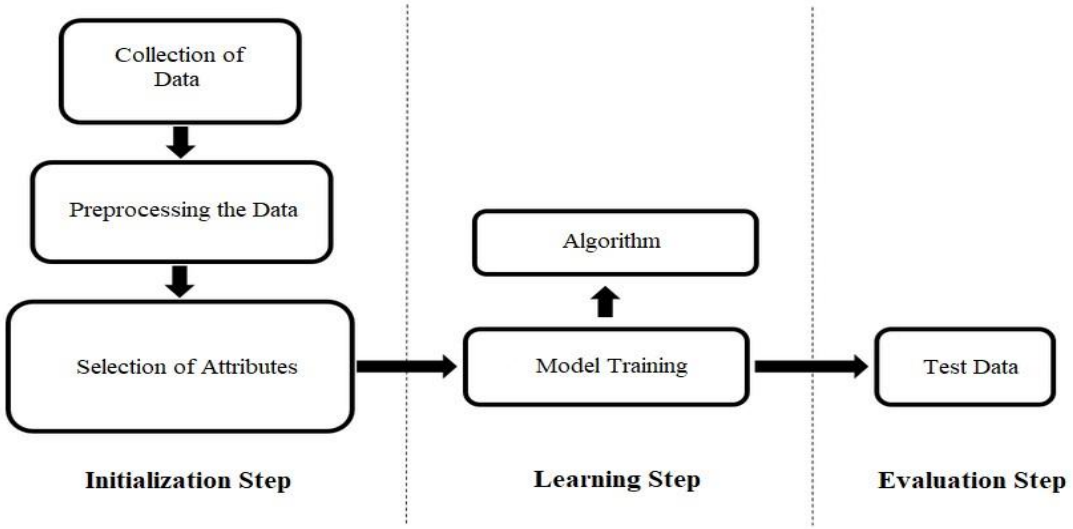


Fig .2.3 : System architecture

MODULES:

- upload dataset: using this module we will upload dataset
- Dataset cleaning: using this module we will find out empty values in the dataset and replace with mean or 0 values.
- Train & Test Split: Using this module we will split dataset into two parts called and training and testing. All machine learning algorithms take 80% dataset to train classifier and 20% dataset is used to test classifier prediction accuracy.

MACHINE LEARNING

A machine learning algorithm is the method by which the AI system conducts its task, generally predicting output values from given input data. The two main processes of machine learning algorithms are classification and regression.

Machine learning (ML) algorithms are broadly categorized as either supervised or unsupervised. Supervised learning algorithms have both input data and desired output data provided for them through labeling, while unsupervised algorithms work with data that is neither classified nor labeled. An unsupervised algorithm might, for example, group unsorted data according to similarities and differences.

However, many ML approaches, including transfer learning and active learning, involve what are more accurately described as semi supervised algorithms. Transfer learning uses knowledge gained from completing one task to help solve a different but related problem, while active learning allows an algorithm to query the user or some other source for more information. Both systems are commonly used in situations where labeled data is scant.

Reinforcement learning, sometimes considered a fourth category, is based on rewarding desired behaviors and/or punishing undesired ones to direct unsupervised machine learning through rewards and penalties.

Supervised learning vs. unsupervised learning:

Independent of these divisions, there are another two kinds of machine learning algorithms: supervised and unsupervised. In supervised learning, you provide a training data set with answers, such as a set of pictures of animals along with the names of the animals. The goal of that training would be a model that could correctly identify a picture (of a kind of animal that was included in the training set) that it had not previously seen.

In unsupervised learning, the algorithm goes through the data itself and tries to come up with meaningful results. The result might be, for example, a set of clusters of data points that could be related within each cluster. That works better when the clusters don’t overlap.

Training and evaluation turn supervised learning algorithms into models by optimizing their parameters to find the set of values that best matches the ground truth of your data. The algorithms often rely on variants of steepest descent for their optimizers, for example stochastic gradient descent (SGD), which is essentially steepest descent performed multiple times from randomized

starting points. Common refinements on SGD add factors that correct the direction of the gradient based on momentum or adjust the learning rate based on progress from one pass through the data (called an epoch) to the next.

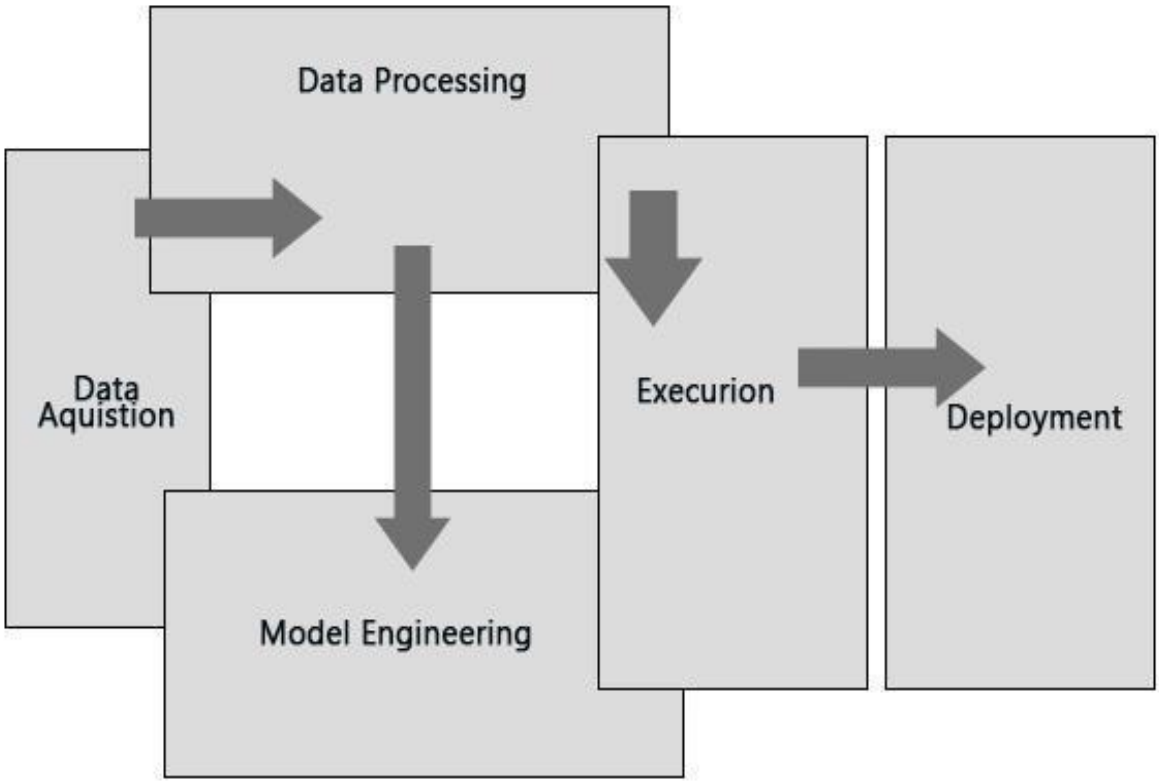


Fig No 2.4 : Machine learning algorithms process steps

There are dozens of machine learning algorithms, ranging in complexity from linear regression and logistic regression to deep neural networks and ensembles (combinations of other models). However, some of the most common algorithms include:

- Linear regression, aka least squares regression (for numeric data)
- Logistic regression (for binary classification)
- Linear discriminant analysis (for multicategory classification)
- Decision trees (for both classification and regression)
- Naïve Bayes (for both classification and regression)
- K-Nearest Neighbors, aka KNN (for both classification and regression)

  Learning Vector Quantization, aka LVQ (for both classification and regression.

2.3 Analysis Of Women Safety In Indian Cities Using Machine Learning On Tweet

2.3.1 Introduction:

[7] In the modern era, Twitter has emerged as a premier microblogging social network with over a hundred million users generating over five hundred million messages, known as 'Tweets,' daily. Twitter's vast audience has attracted users to express their opinions on various internet issues, making it an informative source for institutions, companies, and organizations. Users share their perspectives in tweets, limited to 140 characters, leading to compact messages with abbreviations, slang, emoticons, etc. Additionally, many express opinions using polysemy and sarcasm, rendering Twitter language unstructured. [8] Sentimental analysis extracts the sentiment behind tweets, which finds application in analyzing public opinions on topics such as brand sentiment, government policies, and societal attitudes towards women.

Women and girls face violence and harassment in public places in Indian cities, ranging from stalking to assault. This research focuses on the role of social media, particularly platforms like Twitter, Facebook, and Instagram, in promoting women's safety. [9] It explores how social media can foster a sense of responsibility among Indian society to prioritize women's safety. Tweets on Twitter, containing images, text, and quotes focusing on women's safety in Indian cities, serve as a message to Indian youth culture, advocating for strict action against harassers. Hashtags on Twitter provide a platform for women to express their feelings about safety while traveling or working and their sense of security in public spaces. [10] Staring at women and passing comments can be certain types of violence and harassments and these practices, which are unacceptable, are usually normal especially on the part of urban life. Many researches that have been conducted in India shows that women have reported sexual harassment and other practices as stated above. Such studies have also shown that in popular metropolitan cities like Delhi, Pune, Chennai and Mumbai, most women feel they are unsafe when surrounded by unknown people. On social media, people can freely express what they feel about the Indian politics, society and many other thoughts. Similarly, women can also share their experiences if they have faced any violence or sexual harassment and this brings innocent people together in order to stand up against such incidents. From the analysis of tweets text collection obtained by the

twitter, it includes names of people who has harassed the women and also names of women or innocent people who have stood against such violent acts or unethical behaviour of men and thus making them uncomfortable to walk freely in public.

The data set of the tweet will be used to process the machine learning algorithms and models. This algorithm will perform smoothening the tweet data by eliminating zero values. Using Laplace and porter's theory, a method is developed in order to analyze the tweet data and remove redundant information from the data set. Huge numbers of people have been attracted to social media platform such as Twitter, Facebook, Instagram. People express their sentiments about society, politics, women, etc via the text messages, emoticons and hash-tags through such platforms. There are some methods of sentiment that can be classified like machine leaning based and lexicon based learning.

2.3.2 Merits & Demerits

Merits

- Analysis of twitter texts collection also includes the name of people and name of women who stand up against abuse harassment and unethical behaviour 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**.**

Demerits

- Twitter and Instagram point and most of the people are using it to express their emotions and also their opinions about what they think about the Indian cities and Indian society.
- There are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning.
- Also there are another categorization Janta presented with categories of statistical, knowledge-based and age wise differentiation approaches

2.3.3 Implementation

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 like 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 neighbours 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 abusely 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 abuse 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.

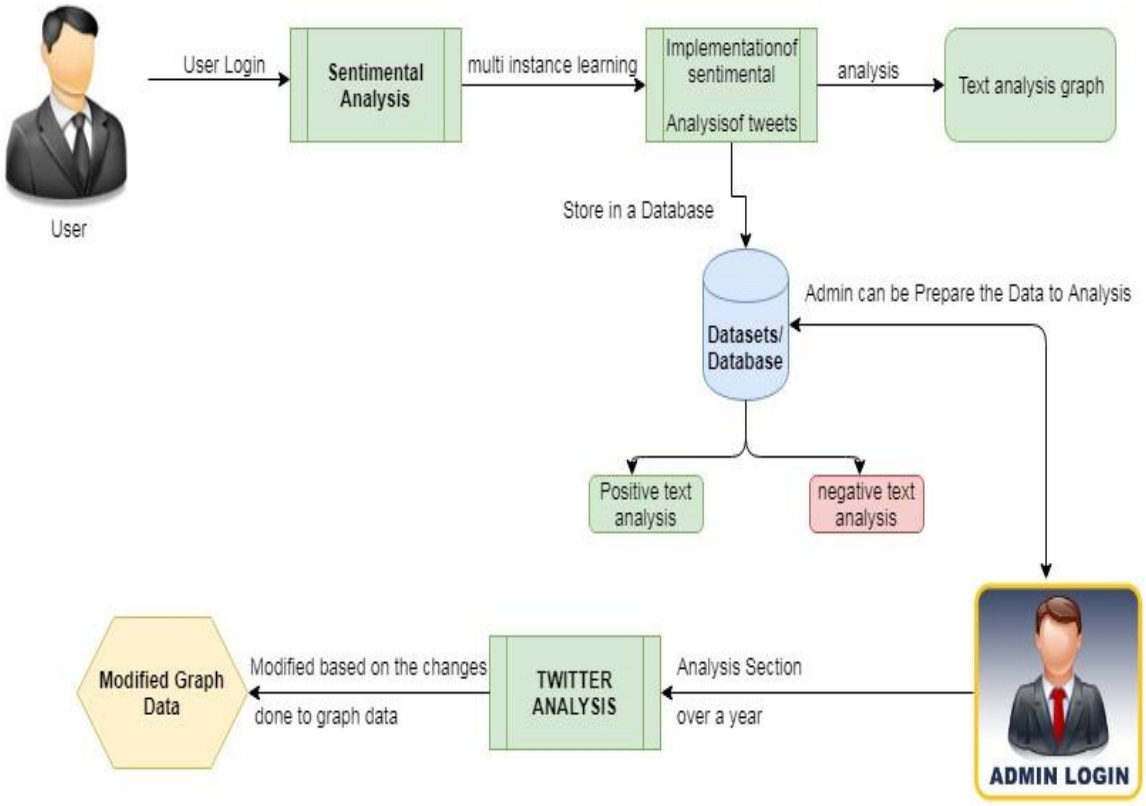


Fig No 2.5 : Twitter Analysis

People communicate and share their opinion actively on social medias including Facebook and Twitter, Social network can be considered as a perfect platform to learn about people's opinion and sentiments regarding different events. There exists several opinion oriented information gathering and analytics systems that aim to extract people's opinion regarding different topics.

IMPLEMENTATION OF SENTIMENTAL ANALYSIS OF TWEETS

Report the tweets picked up from Twitter API provided by Twitter itself. Due to the presence of Twitter API, there are many techniques available for sentimental analysis of data on Social media. In this project a set of available libraries has been used.

If the neutral tweets are significantly high, means that people have a lower interest in the topic and are not willing to haves a positive/negative side on it. This is also important to mention that depends on the data of the experiment we may get different results as people's opinion may change depending on the circumstances for example rape news it becomes the most trending news of the year in 2017. For some queries, the neutral tweets are more than 60% which clearly shows the limitation of the views.

Throughout about various machine learning algorithms that can help us to organize and analyze the huge amount of Twitter data obtained including millions of tweets and text messages shared every day. These machine learning algorithms are very effective and useful when it comes to analyzing of large amount of data including the SPC algorithm and linear algebraic Factor Model approaches which help to further categorize the data into meaningful groups. Support vector machines is yet another form of machine learning algorithm that is very popular in extracting Useful information from the Twitter and get an idea about the status of women safety in Indian cities.

CHAPTER 3
PROPOSED SYSTEM

CHAPTER 3

PROPOSED SYSTEM

3.1 INTRODUCTION

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

harassment and this brings innocent people together in order to stand up against such incidents. From the analysis of tweets text collection obtained by the twitter, it includes names of people who has harassed the women and also names of women or innocent people who have stood against such violent acts or unethical behavior of men and thus making them uncomfortable to walk freely in publics.

3.2 Objective

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

3.2.1.Algorithms Used for Proposed Model

In the quest for innovation and efficiency, modern projects frequently rely on existing solutions as fundamental building blocks for development. That method doesnt only reminds the expertise and advancements of those who came before us but also nurtures a collaborative ecosystem where ideas can evolve and confront new challenges. In our project, we wholeheartedly embrace this ethos, conscientiously integrating elements from existing solutions to enrich our endeavor. These existing solutions serve as guiding lights, offering insights and frameworks that shape the direction of our project.

Information Extraction:

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

Text Cleaning:

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

NLP(Natural Language Processing) :

Computational methods are exceptionally associated in Ordinary Lingo Taking care of (NLP) to

assessment, get, and make information on human lingos. It incorporates a large values with errands like machine interpretation, named substance assertion, doubt testing, and substanance classification. NLP calculations distill structure and meaning from substance data by applying neural frameworks, quantifiable models, and etymlogical rules. Its livelihoods consolidate a wide amplify of zones, checking chatbots, virtual arsssitstants, data recovery systems, and estimation testing devices, progressing robotized planning and tongue comprehension.

Sentimental Analysis

Once the classifier cleans the dataset, the information is prepared for the nostalgic investigation measure. AI approaches includes in Training the dataset and afterward testing that prepared dataset.

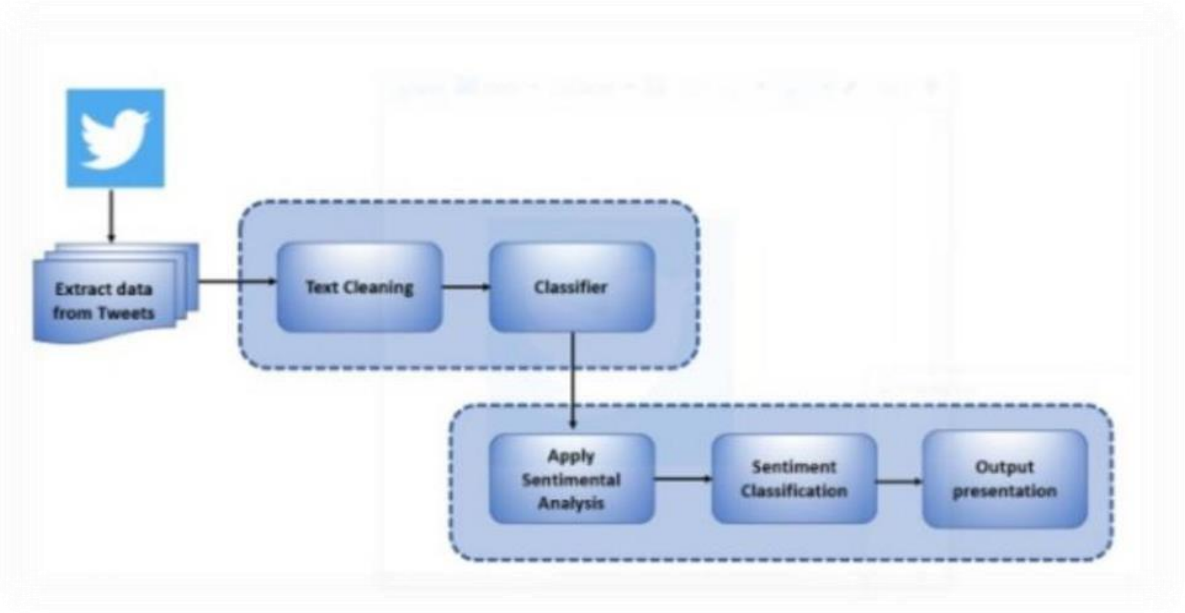


Fig No 3.1 : Classifier of Sentimental Analysis.

Prepared and tried information are helpful for the classifier to play out the calculation. Greatest Entropy, Navies Bayes order, Support Vector Machine and Bayesian Networks are a portion of the calculation which can be utilized to prepare the classifier. Testing information which is utilized to distinguish the effectiveness of the slant classifier.

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

XGBoosting Algorithm:

Extraordinary Angle Boosting, or XGBoost for brief, could be a effective machine learning strategy that exceeds expectations at taking care of both organized and unstructured information. It is based on the slope boosting design and finds utilize in gathering learning procedures. The adaptability of the approach makes it fitting for dealing with expansive datasets, and regularization procedures are coordinates into the calculation to anticipate overfitting issues. Iteratively progressing expectation precision, XGBoost makes utilize of an gathering of choice trees. Due to its flexibility and steadfastness, it has been broadly utilized for a assortment of ML applications, counting relapse, positioning examination, and classification.

3.3 Designing

Software design sits at the technical kernel of the software engineering process and is applied regardless of the development paradigm and area of application. Design is the first step in the development phase for any engineered product or system. The designer's goal is to produce a model or representation of an entity that will later be built. Beginning, once system requirements have been specified and analyzed, system design is the first of the three technical activities - design, code and test that is required to build and verify software.

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

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

3.3.1.ARCHITECTURE DIAGRAM FOR PROPOSED MODEL

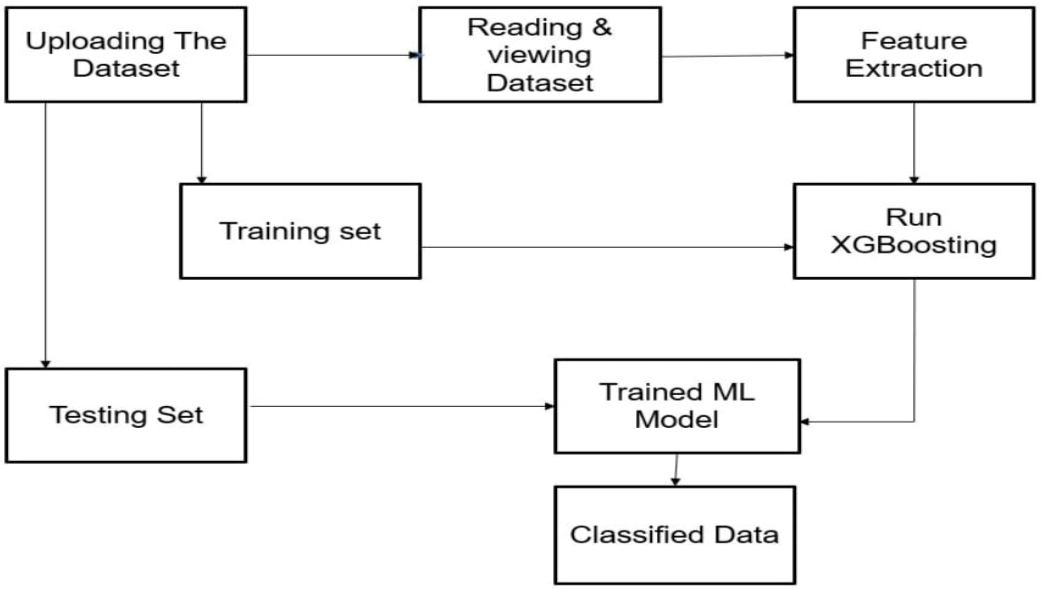


Fig No 3.2 : Architecture For Proposed System

Step1: Pre-processing.

1) *Step2:* Features extraction utilizing counting vectorization, word level TF-IDF and n-gram level TF-IDF. Then Converting the Text into Vectors.

$$TF(t) = \frac{\text{(Number of times term t appears in a document)}}{\text{(Total number of terms in the document)}}$$

$$IDF(T) = \log \frac{\text{Total number of docs}}{\text{Number of docs with term t in i}}$$

Step3: Fit the training dataset, it learns the model

Step4: XGBoosting classifier is utilized to predict the values as the XGBoosting initializtion is done, the accurate calculation of the model will be done inlcuding all the three vectorization techniques.

Dataset

- The dataset we are using is “MeToo Tweets”
- There are 15k rows and 3 columns.
- This dataset includes many fields about each tweet like Text, Favorite_count and Retweet_ count.
- https://www.kaggle.com/hollyhetherington/metootweets?select=MeToo_tweet s.csv

3.3.2 UML Diagrams

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

Use Case Diagram

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

These internal and external agents are known as actors. Use case diagrams consist of actors, use cases and their relationships. The diagram is used to model the system/subsystem of an application. A single use case diagram captures a particular functionality of a system.

Hence to model the entire system, a number of use case diagrams are used.

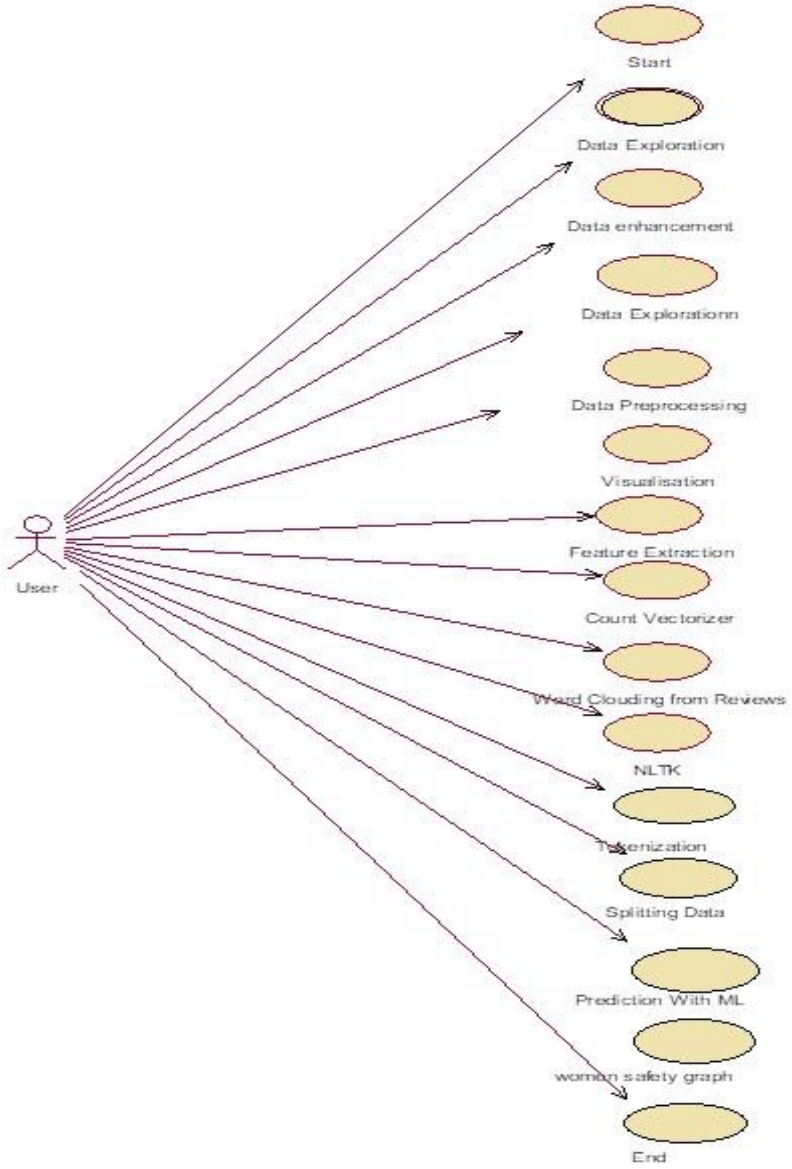


Fig No 3.3 : Use Case Diagram

Sequence Diagram

Sequence Diagrams represent the objects participating the interaction horizontally and time vertically. A Use Case is a kind of behavioural classifier that represents a declaration of an offered behaviour. Each use case specifies some behaviour, possibly in including variants that the subject can perform in collaboration with one or more actors. Use cases define the offered behaviour of the subject without reference to its internal structure. These behaviours, involving interactions between the actor and the subject, may result in changes to the state of the subject and communications with its environment. A use case can include possible variations of its basic behaviour, including exceptional behaviour and error handling.

Fig 3.4 : Sequence Diagram

Class Diagram

The class diagram is the main building block of object-oriented modelling. It is used for general conceptual modelling of the systematic of the application, and for detailed modelling translating the models into programming code. Class diagrams can also be used for data modelling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.

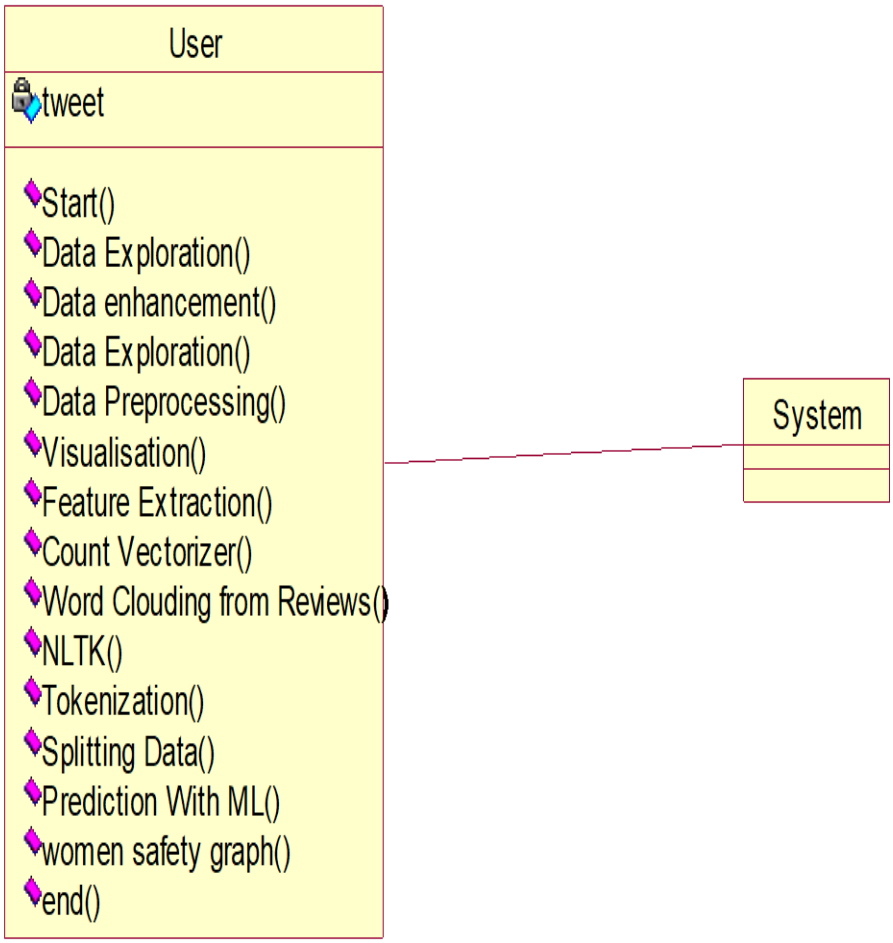


Fig 3.5 : Class Diagram

3.3 Stepwise Implementation

- TWEEPY package from python to download tweets from twitter but every time INTERNET will not available to download tweets online so we downloaded MEETOO tweets on women safety and safe inside dataset folder.
- Application will read these tweets to detect women's sentiments.Here we are using NLTK (Natural language tool kit) to remove special symbols and stop words from tweets and to make them clean.
- Uploading dataset
- Data Loading
- In the above step upload MeeToo_Tweets.csv record and After that press on'Open' button to stack dataset in above activity each line represents single document, will be able scroll down
- At the above screen we will view every single doc, includes special texts, halt words and to minimize those texts tap on 'Tweets Cleaning button
- At above screen we will see all extraordinary images and Halt words and expels through the tweets and as it were clean wordsare there and presently press on "Run ML algorithm button to foresee assumptions .
- See all extraordinary images and halt words expefrom tweets and as it were clean words are there and presently press at 'Run ML Algorithm' button to foresee assumptions.
- In above activity 0.74 multiplied by 100 which gives 74% that means 74% of audience are talking negatively and area is not safe and only 22 and 3% of audience are talking positively.

CHAPTER 4
RESULTS AND DISCUSSION

4 . RESULTS AND DISCUSSION

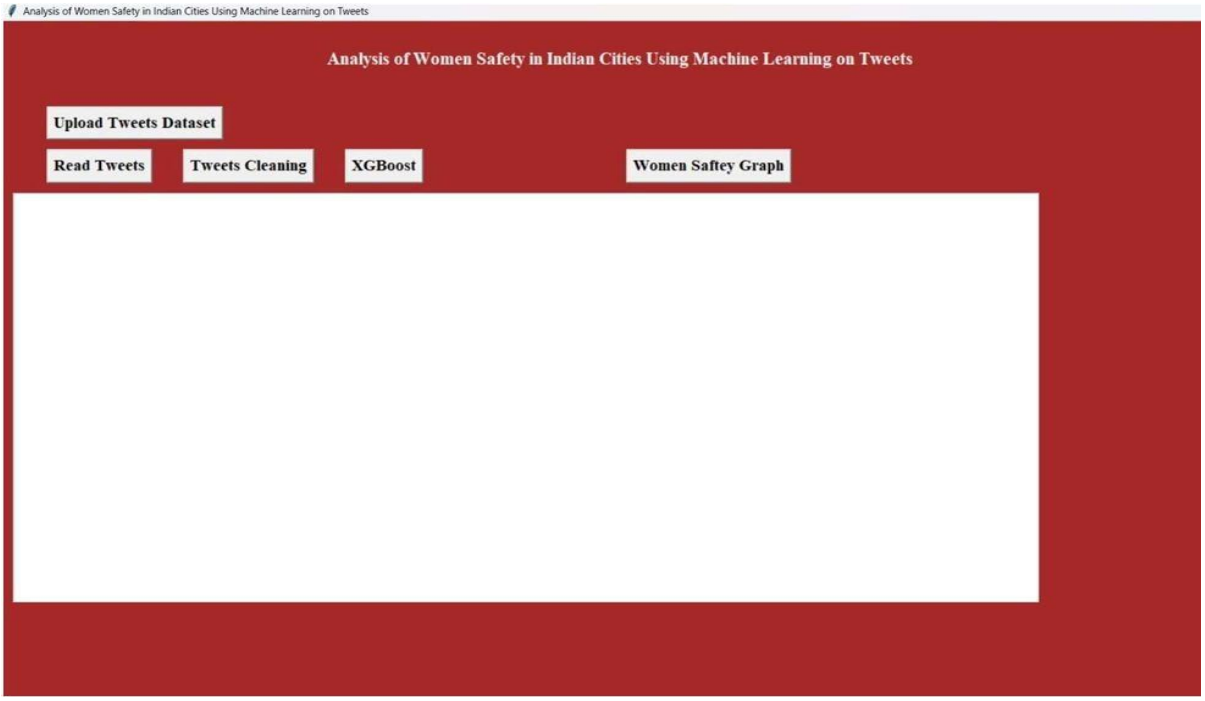


Fig No 4.1 : User Interface

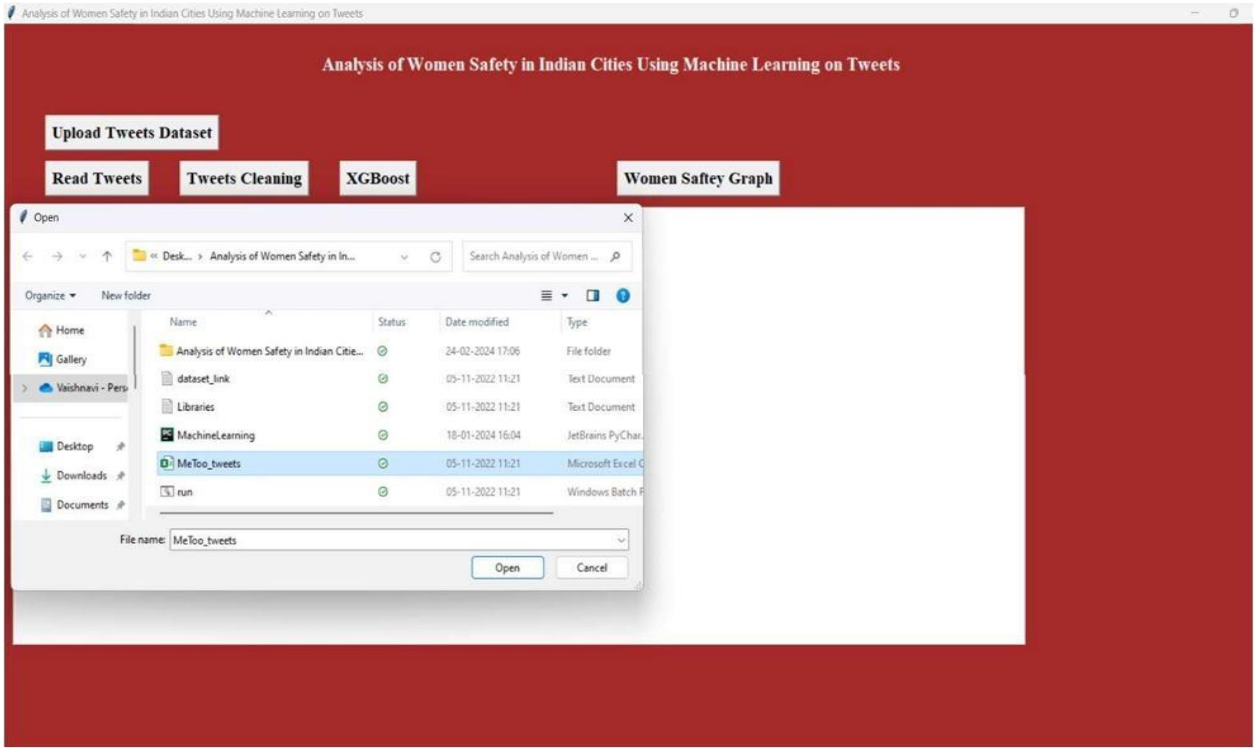


Fig 4.2 : Uploading dataset

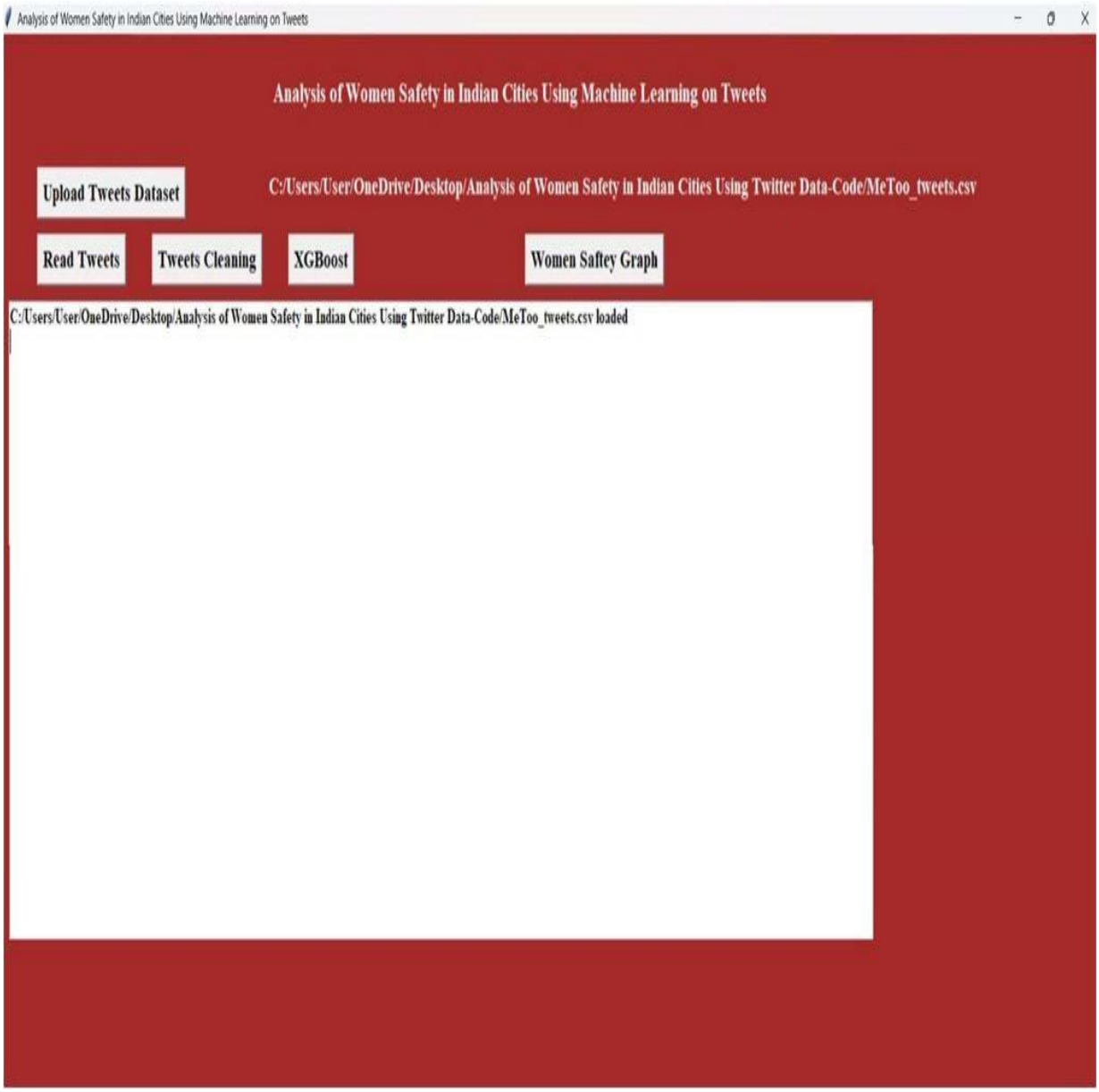


Fig 4.3 : Data Loading

In the above step upload MeeToo_Tweets.csv record and After that press on'Open' button to stack dataset in above activity each line represents single document, will be able scroll down

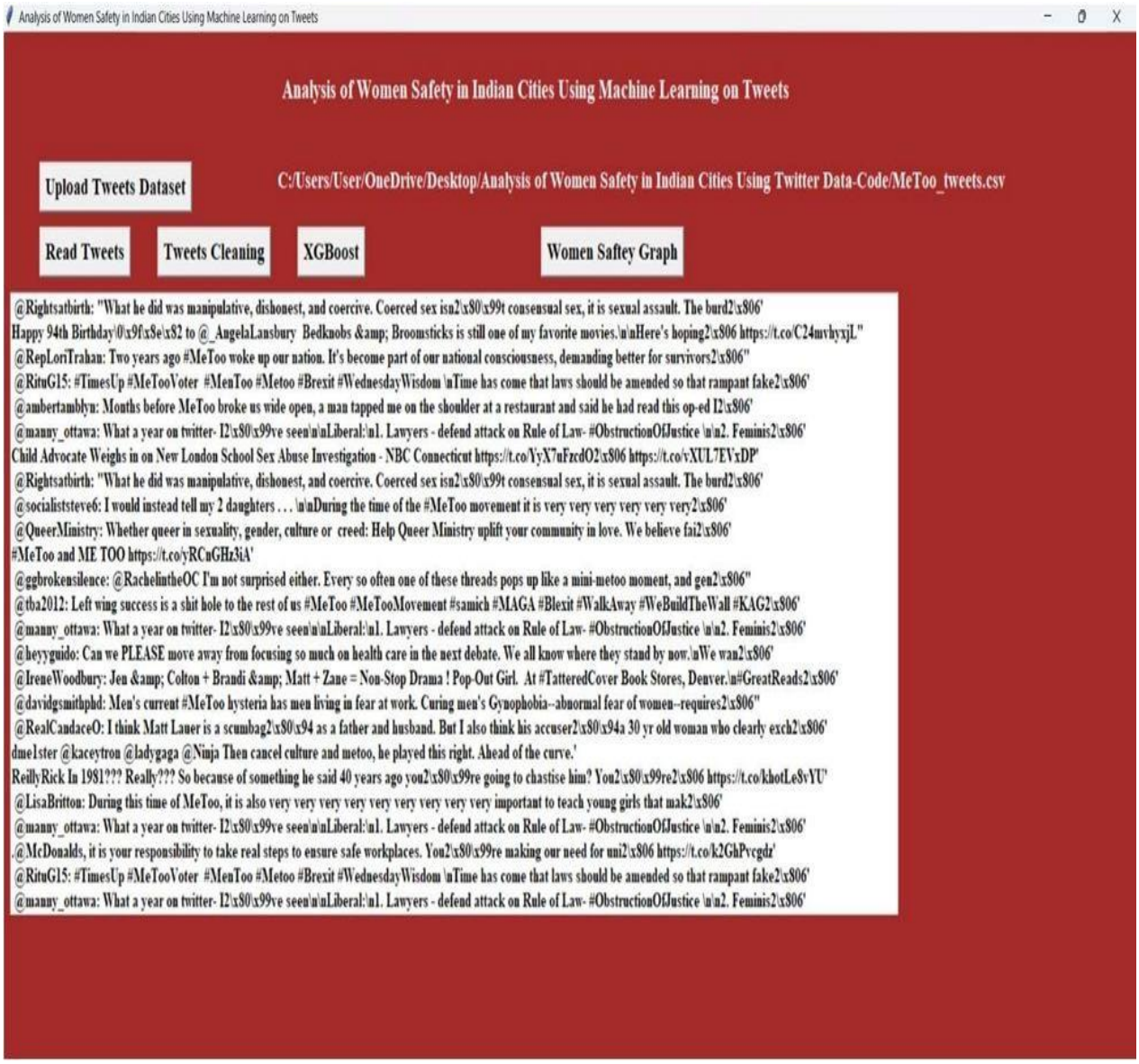


Fig 4.4 : Reading the Tweets

At the above screen we will view every single doc, includes special texts, halt words and to minimize those texts tap on 'Tweets Cleaning button

Fig 4.5 : Tweets Cleaning

At above screen we will see all extraordinary images and Halt words and expels through the tweets and as it were clean wordsare there and presently press on “Run ML algorithum button to foresee assumptions .

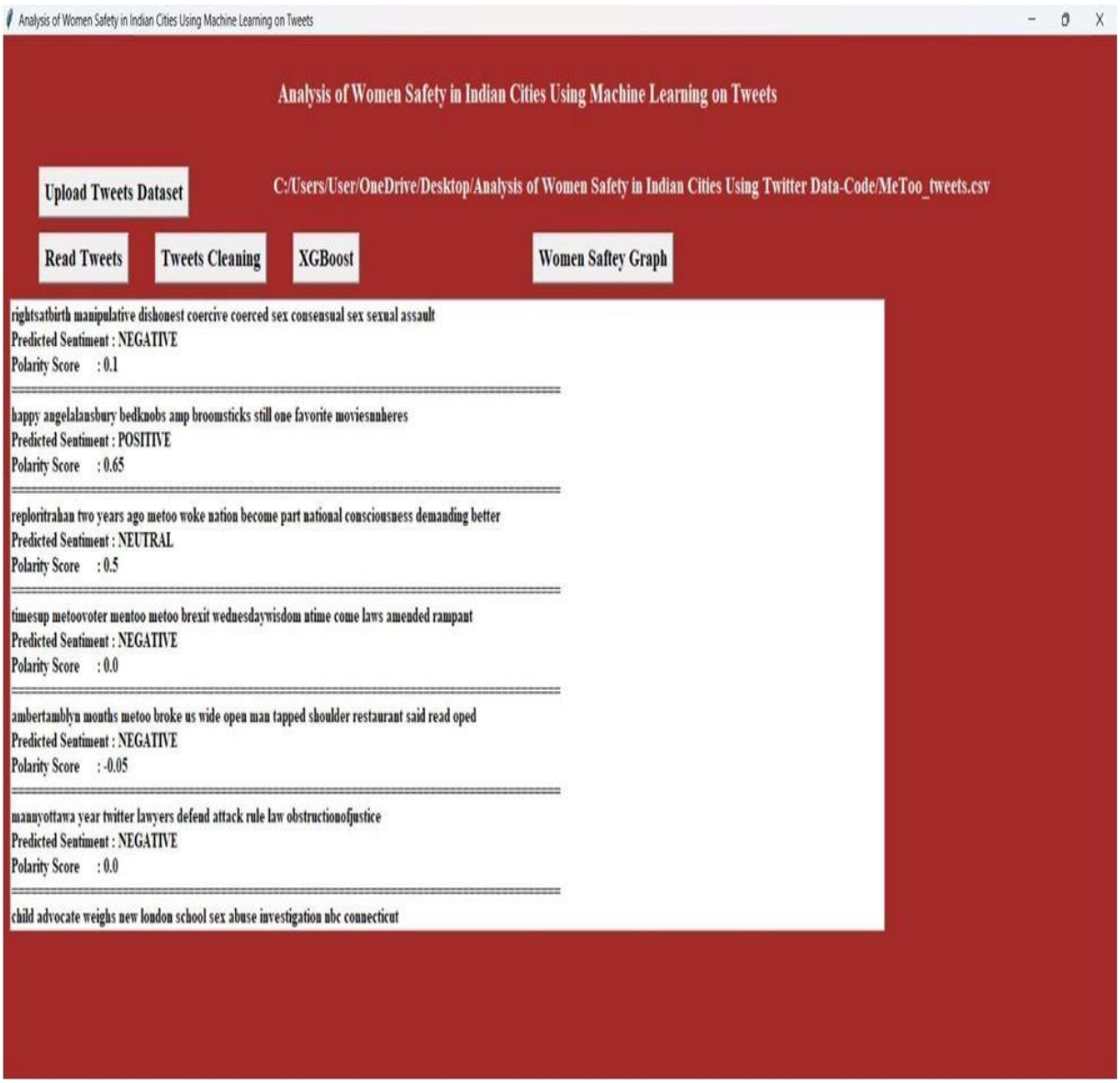


Fig 4.6 : XG Boosting Algorithm

4.1 Performance Measure:

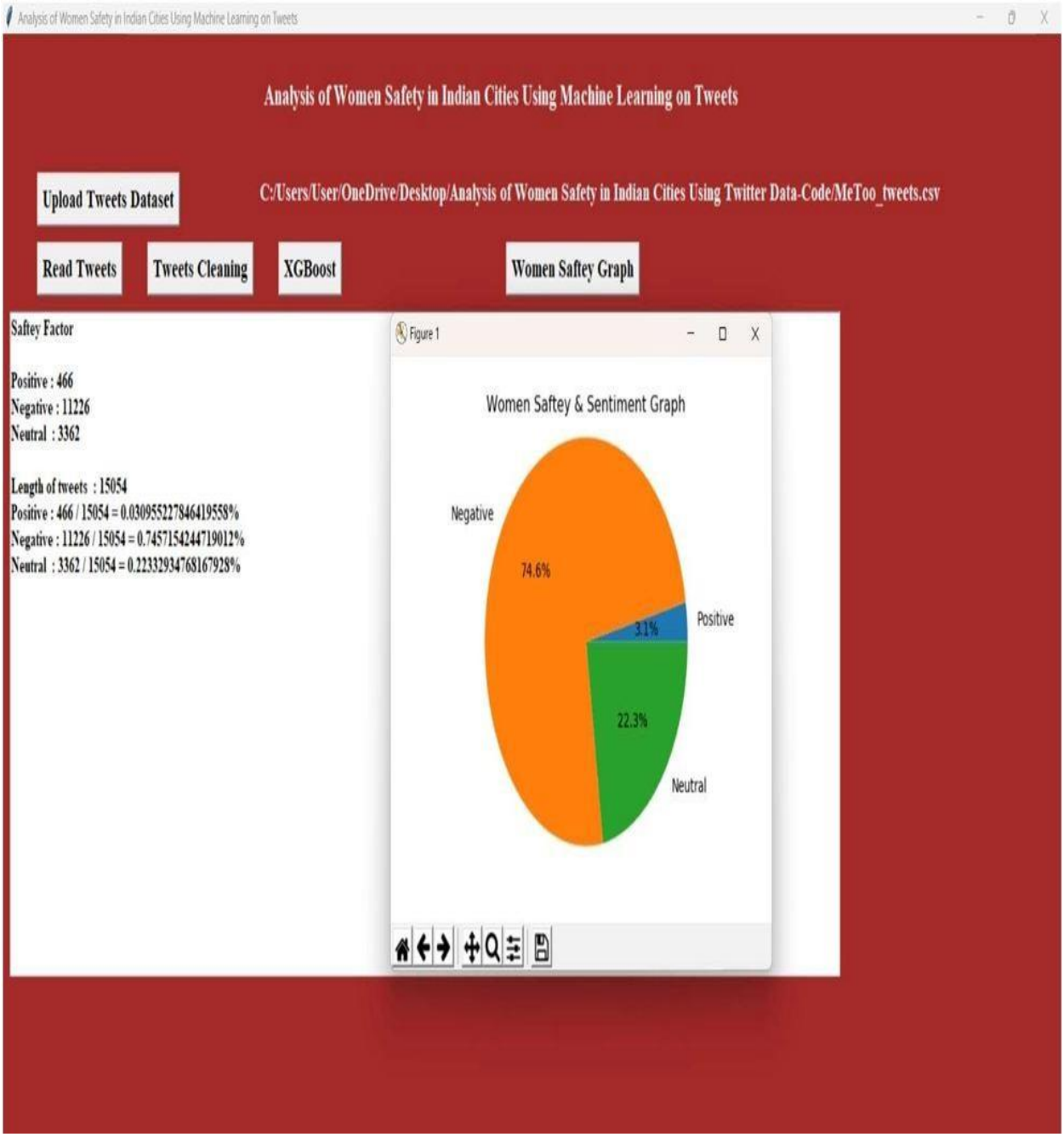


Fig 4.7 : Women Safety Graph

In above activity 0.74 multiplied by 100 which gives 74% that means 74% of audience are talking negatively and area is not safe and only 22 and 3% of audience are talking positively.

CHAPTER 5
CONCLUSION

CHAPTER 5

CONCLUSION

Conclusion:

Throughout the various algorithms have been discussed about machine learning which can help in analyzing huge amount of data accumulated via tweeter to help determine the safety of women in the society. The machine learning algorithms used are very effective and work efficiently on various platforms when it comes to handling the large amountof data from social media platforms. These algorithms can really help make a dent in women safety and extracting information and create various datasets to work with. As an outcome, this will present a global picture of various crimes against women and demonstrate how the scenario's goal and motivation are structured. We can identify risky areas in metropolitan cities and increase safety in those areas as a result of our project, allowing women to feel free to go wherever they want. As a result, we present recommendations in order to reduce the possibility of unsecured problems appearing in society, which will undoubtedly aid in. lowering the risk of these devious actions affecting society, and by rigorously adhering to the rules, we will be able to reach a greater level of security

5.1 Future Enhancement

When it comes to handling enormous amounts of data from social media sites, the machine learning algorithms utilised are quite effective and perform well on a variety of platforms. These algorithms have the potential to make a significant difference in women's safety by extracting data and creating a diverse range of datasets to work with. In the near future, we intend to work on this project more and tweak it to make it even more efficient. As well as we add more social medias like instgram, facebook etc.

REFERENCES

REFERENCES

1. Kumar, D., & Aggarwal, S. (2019, February). Analysis of women safety in indian cities using machine learning on tweets. In *2019 Amity International Conference on Artificial Intelligence (AICAI)* (pp. 159-162). IEEE.
2. TEJASHWINI, M., VENNA LAKSHMI PRASANNA, SUBBARA PRAVALLIKA, SHAIK AFROZ, RACHAGOLLA RAGHUVARAN YADAV, and MADIGA DHANUNJAYA. "Analysis of Women Safety in Indian cities using Machine Learning on Tweets."
3. Chandra, V. and Srinath, R., 2020. Analysis of Women Safety using Machine Learning on Tweets. ,*(IRJET)*.
4. Rani, K. Pushpa, T. Raja Rajeswari, N. Thulasi Chitra, and B. Devananda Rao. "Analysis of women safety in Indian cities using machine learning." In *AIP Conference Proceedings*, vol. 2492, no. 1. AIP Publishing, 2023.
5. Kavitha, N., N. Srinivasa Rao, and Srinivasa Rao Madala. "Applying Machine Learning Techniques To Analyze The Women Safety." *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO* (2021): 1289-1294.
6. Mamatha, Mrs M. "ANALYSIS OF WOMEN SAFETY IN INDIAN CITIES USING MACHINE LEARNING ON TWEETS."
7. Pradeep, Goppisetty Sai, Parasa Satya Krishna, Pasangulapati Anil, and Gowrisettti Rohit. *Analyzing the Security of Womens in Indian Cities by Using Machine Learning*. No. 12678. EasyChair, 2024.
8. Chandrakala, N., K. Mamatha, G. Anitha, and M. Rohith. "ANALYSIS OF WOMEN SAFETY IN INDIAN CITIES ON TWEETS USING MACHINE LEARNING." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 10, no. 3 (2019): 1211-1220.
9. Reddy, A. Yashwanth, et al. "The Role of Social Media In Promoting The Safety Of Women in Indian Cities." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 14.03 (2023): 1342-1351.
10. Bhavani, Mrs B. Ganga, et al. "Women's Safety Analysis on social media using Machine Learning." *Journal of Algebraic Statistics* 13.2 (2022): 2758-2767.

GIT HUB LINK

https://github.com/Yesu19/Analysis-of-women-safety-in-indian-cities.git4

CODE:

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

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

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

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

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


def read():
    text.delete('1.0', END)
    tweets_list.clear()
    train = pd.read_csv(filename,encoding='iso-8859-1')
    for i in range(len(train)):
        tweet = train.get_value(i, 'Text')
        tweets_list.append(tweet)
        text.insert(END,tweet+"\n")
    text.insert(END,"\n\nTotal tweets found in dataset is : "+str(len(tweets_list))+"\n\n\n")
def clean():
    text.delete('1.0', END)
    clean_list.clear()
    for i in range(len(tweets_list)):
        tweet = tweets_list[i]
        tweet = tweet.strip("\n")
        tweet = tweet.strip()
        tweet = tweetCleaning(tweet.lower())
        clean_list.append(tweet)
        text.insert(END,tweet+"\n")
    text.insert(END,"\n\nTotal tweets found in dataset is : "+str(len(clean_list))+"\n\n\n")
def machineLearning():
    text.delete('1.0', END)
    global pos, neu, neg
```

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

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

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

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

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

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

font = ('times', 16, 'bold')
title = Label(main, text='Analysis of Women Safety in Indian Cities Using Machine Learning
on Tweets')
title.config(bg='brown'g='white')
```

```
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)

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

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

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

mlButton = Button(main, text="XGBoost", 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.configure(yscrollcommand=scroll.set)
text.place(x=10,y=200)
text.config(font=font1)



main.config(bg='brown')
main.mainloop()
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from xgboost import XGBRegressor
import joblib
import pandas as pd
# Load the dataset
data = pd.read_csv('MeToo_tweets.csv')

# Selecting the 'Source' and 'Retweet_count' columns as features and target respectively
X = data[['Source']]
y = data['Retweet_count']

# Encoding the 'Source' column as it is categorical
label_encoder = LabelEncoder()
X_encoded = label_encoder.fit_transform(X['Source']).reshape(-1, 1)
# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_encoded, y, test_size=0.2,
random_state=42)

# Define the XGBoost regressor with reduced complexity
xgb_regressor = XGBRegressor(n_estimators=50, max_depth=3, learning_rate=0
```

```
# Train the XGBoost regressor on the training data
xgb_regressor.fit(X_train, y_train)

# Save the trained model to disk
model_filename = 'xgb_meToo_tweets_model.joblib'
joblib.dump(xgb_regressor, model_filename)

print(f"Model saved successfully at {model_filename}")
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error

# Predictions
y_train_pred = xgb_regressor.predict(X_train)
y_test_pred = xgb_regressor.predict(X_test)

# Calculate MSE for both training and testing sets
mse_train = mean_squared_error(y_train, y_train_pred)
mse_test = mean_squared_error(y_test, y_test_pred)

# Plotting the MSE values
plt.figure(figsize=(10, 6))
plt.bar(['Train MSE', 'Test MSE'], [mse_train, mse_test], color=['blue', 'green'])
plt.title('Model Performance: Mean Squared Error')
plt.ylabel('Mean Squared Error')
plt.show()
```

iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59199

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Analysis of Women Safety in Indian Cities Using Machine Learning

Dr. G. Ravi Kumar[1], K. Venkata Giridhar[2], M. Vaishnavi[3], N. Yesumani[4]
[1]*Associate Professor, Dept of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*
[2, 3, 4]*UG Student, Dept of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*

Abstract: ***In later times, ladies have been confronting an disturbing rise in viciousness, counting badgering, totally different cities. This frequently starts with stalking and can raise into injurious ambush. In this dialog, we are centering with in the area of social mediastages like Twitters, FB, and Insta at advancing the security of ladies in India. These stages give a chance at raise mindfulness and teachindividuals approximately taking unequivocal activities against damaging behavior. By utilizing hashtags and spreading messages all inclusive, ladies are enabled to openly express their conclusions and feelings. This empowers us to pick up experiences into their state ofintellect while ladies out for working or travel, utilizing open transportation, or within the nearness of new people. It makes a difference us decide whether they feel secure or not in these circumstances.These times, badgering and other shapes of savagery against ladies are common in numerous cities! Stalking is the primary step towards injurious badgering, which is now and then alluded to as manhandle attack? With specific reference to the utilization of various social media stages, counting Twitter, Facebook, and Instagram, our essential center in this consider at with in the task with social media in advancing women's security in India. This exposition moreover emphasizes the require for common individuals to assume more duty in several Indian city neighborhoods in arrange to ensure the security of ladies in their quick region. Inside the Twitter app, a tweet comprises content messages, photographs, recordings, sound records, emoticons, and hashtags! These regions, require more consideration in regards with women's security, and communalengagement is crucial for improvement.***
Keywords: Twitter, Civic, Facebook, Instagram, Sentimental Analysis, Data Preprocessing.

I. INTRODCUTION

In These Advanced times, bullying and other shapes of brutality against women are common in various cities and all over the world. Stalking is the essential step towards harmful bullying, which is presently and after that implied to as abuse assault. With particular reference to the utilization of differentsocial media stages, checking Twitter Tweets, Facebook and Instagram posts, our fundamental center in this consider is on the work of social media Information in progressing women's security in India. This piece in addition emphasizes the require for common people to expect more obligation in a few Indian city neighborhoods in orchestrate to guarantee the security of women in their speedy locale. Interior the Twitter app, The Tweets comprises of substance messages, photos, recordings, sound records, emoticons, and hashtags. These locales, require more thought in terms of women's security, and communal engagement is vital for enhancement. Opposite to common convictions, women have the proper to feel secure, though utilizing social media stages and have to be be secured from Cyber threats. The government's part it to stay in most extreme centrality in actualizing strict laws to ensure the online securitywith lady at these times. Furthermore, the cooperation of non-profit organizations is imperative to create social mindfulness on these issues and offer back to casualties of online viciousness and badgering towards woman.The progression of social media has driven to both positive and negative impacts in society, especially concerning women's safety. In conclusion, collaboration among government, non profit organizations, and people is vital in making a more secure environment for ladies in India through analyzing the information like messages through Twitter organize with the assistance of a few computerized strategies like wistful investigation, Normal Dialect Preparing at that point Actualizing and processing them with the assistance of Different algorithms like For case XGBoosting Machine Learning Calculation.To Extent quick preparing time and Tall exactness, it is signify to select calculation.

II. LITERATURE SURVEY

1) Apoorva Agarwal, Fadi Biadsy and KathleenR. Mckeown: Within the cutting edge day, Twitter has risen as a communal microblogging arrange that can match other shapesof media. As individuals connected, they watch hundreds of millions of stoners and create over 5 hundred carefully on open media stages like FB and millions of "Tweets" each day. One way to degree the Twitter social framework is as a smooth way to memorize almost people's suppositions.

2) Barbosa, L., & Feng, J:In this paper, we propose a 2- step opinion examination classification strategy for Twitter, which to begin with classifies messages as subjective and objective, and encourage recognizes the subjective tweets as positive or negative. To diminish the labeling exertion in making these classifiers, instep of utilizing physically explained information to compose the preparing information, as normal administered learning approaches, we use sources of boisterous names as our preparing information. These loud names were given by many assumption location websites over twitter information. To superior utilize these sources, we confirm the potential esteem of utilizing and combining them, giving an investigation of the given names, look at diverse methodologies of combining these sources in arrange to get the finest result; and, propose a more vigorous include set that captures a a lot of theoretical showing of tweet, composed by meta-information related to words and particular characteristic with what tweet can be composed. By utilizing it, we point to handle way better: the issue of need of data on tweets, making a difference on the generalization prepare of the classification calculations; and the loud and one-sidednames given by those websites.

3) Gupta, B. ,Negi, M. ,Vishwakarma, K. ,Rawat, G., Badhani, P.: Twitter has developed as a major micro blogging site, having over 100 million clients creating over 500 million tweets each day. With such huge gathering of people, Twitter has reliably pulled in clients to communicate their conclusions and point of view almost any issue, brand, company orelse various different point of intrigued. Due tothis reason, Twitter is utilized as an enlightening source bynumerous organizations, educate and companies. On Twitter, clients are permitted to share their conclusions within the shape of tweets, utilizing as it were 140 characters. This leads to individuals compacting their explanations by using slang, shortened forms, emoticons, brief shapes etc. In conjunction with this, individuals pass on their conclusions by utilizingmockery and polysemy.

4) Mamgain,N., Mehta,E., Mital,A., & Bhat, G: In today's world, conclusions and reviews accessible to us are one of the most basic components in defining our sees and impacting the success of a brand, item or benefit. With the coming and growth of social media within the world, partners regularly take to expressing their conclusions on prevalent social media, to be specific Twitter. Whereas Twitter information is greatly instructive, it presents a challenge for investigation since of itshumongous and disorganized nature. This paper may be a intensive exertion to plunge into novels space of perform the sentimental analysis of audience suppositions as respecting to the beat collage in Indian. Other than taking extra preprocessing measures just like the extension of Netlingo & expulsion of copy information, a probabilitys show based on the Bayes hypothesis was utilized for spelling adjustment, which is neglected in other investigate ponders. This paper moreover highlightning a difference of the comes about gotten by misusing the taking after machine learning calculations: Naïve Bayes and Bolster Vector Machine and an Fake Neural Organize show: Multilayer Perceptron. Besides, a differentiate has beendisplayed between four diverse parts of SVM: RBF, straight, polynomial and sigmoid.

5) Sahayak, V., Shete, V., & Pathan, A: Twitter site is the popular small scale blogging administrations where client can study and post messages which are 148 characters in length. Twitter messages are also called as Tweets. We are going utilize these tweets as crude information. We'll utilize a strategy that consequently extricates tweets into positive, negative or unbiased assumptions. By utilizing the assumption investigation the client can know the criticism almost the item or administrations some time recently making a buy. Thecompany can utilize assumption investigation to know the supposition of clients approximately their items, with it we are able to analyze client fulfillment and concurring to with they are able to process the item. Estimation examination has gotten to be one of well known inquire about zone in computational phonetics, since of the blast of assumption data from social web locales (i.e., Twitter and Facebook), online gatherings, and blogs as in paper.

6) Charniak, E., & Johnson, M: Twitter could be a stage broadly utilized by individuals to precise their suppositions and show opinions on diverse events. Estimation examination which is approaching with analyzing information & recover idea that with epitomizes. Twitter estimation investigation which an advantage with assumption investigation on information through Twitters(tweeting), with arrange to extricate opinions passed on by the client. Within the previous times, as investigate with the work was reliably developed. The reasons below at this can be the challenging arrange with the tweet which make the handling troublesome.

7) Kim, S. M., & Hovy, E: Distinguishing assumptions (the full of feeling parts of conclusions) may be a challenging issue. We display a framework that, given a theme, naturallyfinds the individuals who hold conclusions around that theme and the assumption of each supposition. The framework contains a module for deciding word estimation and another for combining opinions inside a sentence. We explore with different models of classifying and combining assumption at word and sentence levels, with promising comes about.

III. RELATED WORKS

In the quest for innovation and efficiency, modern projects frequently rely on existing solutions as fundamental building blocks for development. That method doesnt only reminds the expertise and advancements of those who came before us but also nurtures a collaborative ecosystem where ideas can evolve and confront new challenges. In our project, we wholeheartedly embrace this ethos, conscientiously integrating elements from existing solutions to enrich our endeavor. These existing solutions serve as guiding lights, offering insights and frameworks that shape the direction of our project.

A. NLP(Natural Language Processing)

Computational methods are exceptionally associated in Ordinary Lingo Taking care of (NLP) to assessment, get, and make information on human lingos. It incorporates a large values with errands like machine interpretation, named substance assertion, doubt testing, and substanance classification. NLP calculations distill structure and meaning from substance data by applying neural frameworks, quantifiable models, and etymlogical rules. Its livelihoods consolidate a wide amplify of zones, checking chatbots, virtual arsssitstants, data recovery systems, and estimation testing devices, progressing robotized planning and tongue comprehension!

B. A. Sentimental Analysis

The tweets picked up from Twitter API given by Twitter itself. With the nearness of Twitter API, its has number of methods available for wistful examination of information on social media. A few accessible libraries were utilized in this venture. Here's how to extricate feelings from a tweet:

1) Beginning with downloading the thoughtful word reference
2) In it the position download the Twitter testing enlightening collections and include them as a commitment to the program.
3) Clean tweets by dispensing with the halt words and repetitive letters.
4) Tokenize each word and apportion solidarity to the words within the dataset and bolster it to the program.
5) For each word, differentiate it and positive conclusions and negative sentiments word reference and a short time later increment the great tally or critical number of the common expression.
6) At long last, given the positive check and negative tally, the result rate around feeling to conclude the limit which is orchestrated in Great, Critical, and Fair-minded. Engineers have done diverse wistful investigations .

C. XGBoosting Algoritum

Extraordinary Angle Boosting, or XGBoost for brief, could be a effective machine learning strategy that exceeds expectations at taking care of both organized and unstructured information. It is based on the slope boosting design and finds utilize in gathering learning procedures. The adaptability of the approach makes it fitting for dealing with expansive datasets, and regularization procedures are coordinates into the calculation to anticipate overfitting issues. Iteratively progressing expectation precision, XGBoost makes utilize of an gathering of choice trees. Due to its flexibility and steadfastness, it has been broadly utilized for a assortment of ML applications, counting relapse, positioning examination, and classification.

1) *Step1:* Pre-processing.
2) *Step2:* Features extraction utilizing counting vectorization, word level TF-IDF and n-gram level TF-IDF. Then Converting the Text into Vectors.

$$TF(t) = \frac{\text{(Number of times term t appears in a document)}}{\text{(Total number of terms in the documents)}}$$

$$IDF(T) = \log\left(\frac{\text{Total number of docs}}{\text{Number of docs with term t in it}}\right)$$

3) *Step3:* Fit the training dataset, it learns the model
4) *Step4:* XGBoosting classifier is utilized to predict the values as the XGBoosting initializtion is done, the accurate calculation of the model will be done inlcuding all the three vectorization techniques.

D. Algorithms Comparsion Table

ALGORITHUM	ACCURACY	SCALABILITY	ROBUSTNESS	INSIGHTS
Logistic Regression	78%	High	Moder-ate	Identification of key factors influencing women's safety, such as lighting, police presence, transportation options.
Random Forest	85%	High	High	Ensemble method offers improved accuracy by combining multiple decision trees. Identification of complex relation between safety features and actionable insight.
k-Nearest Neighbors	82%	Low	High	Instance-based learning method identifying local safety patterns. Recommendations for immediate safety improvements in proximity to high-risk locations.
Gradient Boosting	87%	High	High	Ensemble method similar to XGBoost, offering improved accuracy by boosting weak learners.Insights into areas requiring targeted safety interventions for maximum impact.

Table-1:Algorithm Comparison Table

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

IV. METHODOLOGY

Within the Estimation Investigation the taking after steps are major to distinguish the positive, negative or unbiased of the twitter post.

They are:

1) Collecting the Dataset: The information is assembled from the twitter utilizing API. Application program interface (API) is utilized to accumulate the information. Twitter site could be a source which comprises of clients tweets.

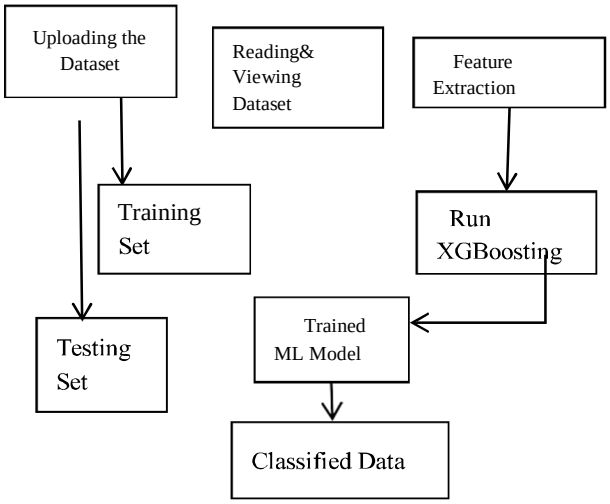


Fig 1 : System Architecture

2) *Pre-Processing the Data-set:* In Information Pre- processing evacuating boisterous, disconnected information, inconsistent and incomplete info from dataset. For the most part, in twitter we have to be evacuate URLs, extraordinary characters, Re-tweets, hash labels.

3) *Feature Extraction:* In this work, we utilized Sack of Words to extricate highlights from content archives. After extraction, these highlights used at training the ML calculations. It makes a language of the clear large number of novel words happening in all the reports within the arrangement set. Sack of words highlights containing term frequencies of each word in each record, i.e. the number ofevent and not sequence or arrange of words things. This will be done by Tally Vectorizer strategy in Python.

4) *Classification:* A classification issue is associated within the occasion that the abandon variable may be a title or category,in our work “Positive” or “Negative”.

V. PROPOSED SYSTEM

The method known as "Sentimental Analysis CLASSIFICATION" reports the tweets that are chosen andgiven by the Twitter program. According to the existence of the Twitter app, there are a few strategies accessible for emotive investigation of social media information.

1) A Downturn in Exchange The crevice between the genuine and sad commerce charts is decreased when a social organize demonstrate makes a G-graph.

2) A commerce chart is developed, and substances andtheir connections in social mass media are associated.

3) The chart appears the likelihood with regard to different topics. Individuals tend to utilize diverse terms and a prepared collection of information that includes condensation since Twitter contains tweets that are unbiased, negative, and positive in expansion to the brief handbooks.

4) Agreeing to advanced NLP, these prompts are touchysufficient to esteem their sentiments.

This expansion proposes a strategy for evaluating women'ssecurity through the examination of social media organizing messages, particularly tweets almost women's security issues gotten through the utilize of the MEETOO hashtag.

The Characteristic Tongue Gadget Pack (NLTK) and the Python programming tongue will be utilized to assemble and plan tweets in arrange to extricate pivotal information. The estimation and substance of the tweets will be assessedutilizing the XG Boost calculation, which is well-known forits capacity to handle both organized and unstructured information.
Extraordinary Angle Boosting, or XGBoost for brief, will bea effective machine learning strategy that exceeds expectations at taking care of both organized and unstructured information. It is related up on slope boosting design and finds utilize in gathering learning procedures. The adaptability of the approach makes it fitting for dealing with expansive datasets, and regularization procedures are coordinates into the calculation to anticipate overfitting issues. Iteratively progressing expectation precision, XGBoost makes utilize of an gathering of choice trees. Dueto its flexibility and steadfastness, it can be used for a assortment of ML applications, counting relapse, positioningexamination, and classification.

VI. RESULTS AND DISCUSSION

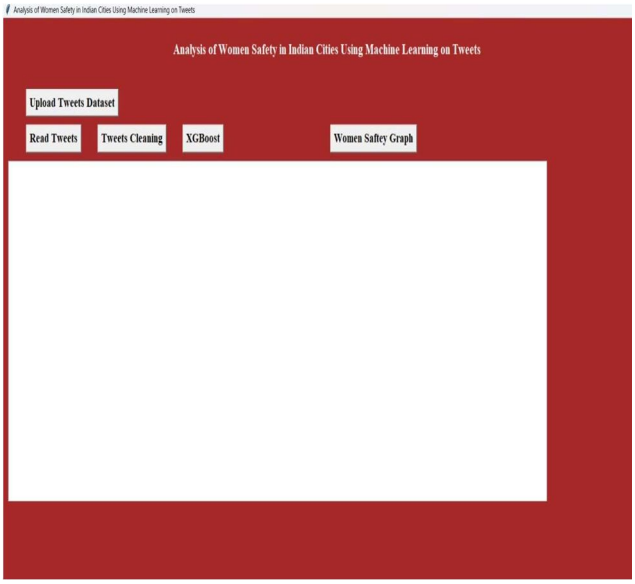


Fig 2: Home Screen

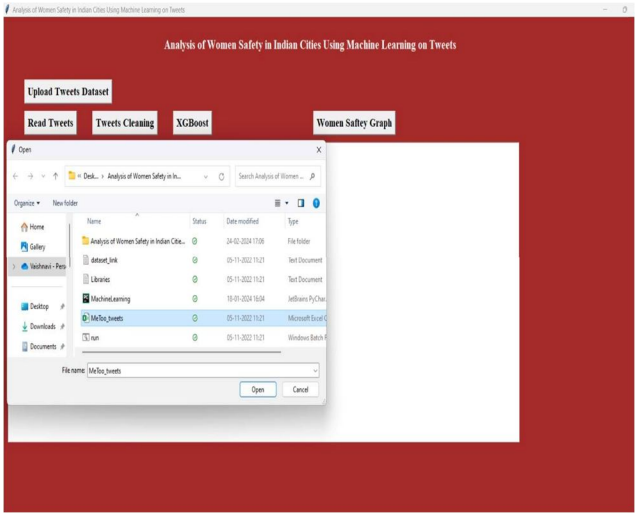


Fig 3: Uploading dataset

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

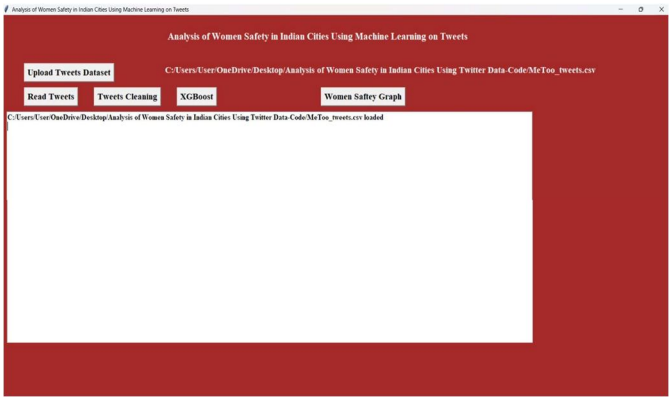


Fig 4: Data Loading

In the above step upload MeeToo_Tweets.csv record and After that press on'Open' button to stack dataset in above activity each line represents single document, will be ablescroll down

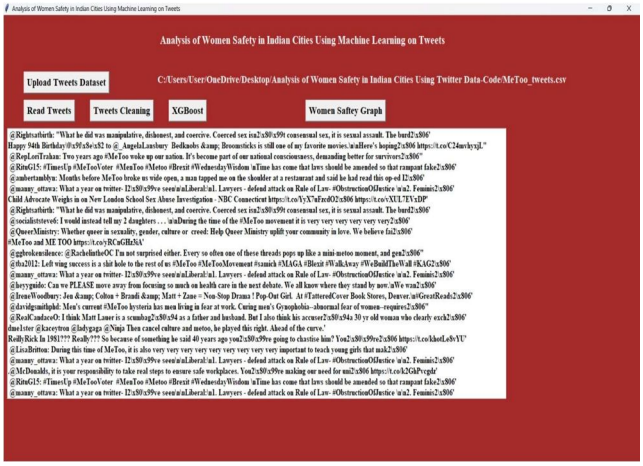


Fig 5: Reading the Tweets

At the above screen we will view every single doc, includes special texts, halt words and to minimize those texts tap on 'Tweets Cleaning button

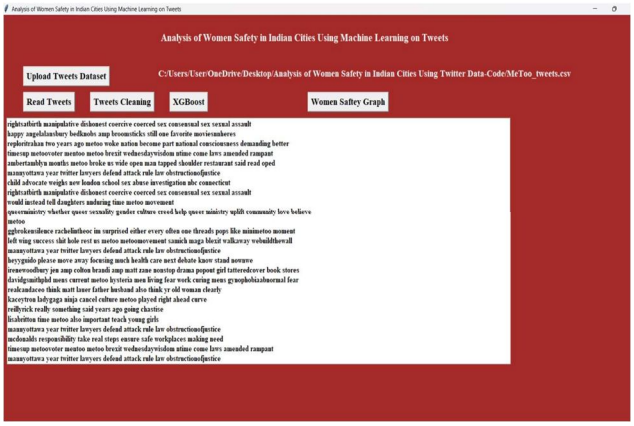


Fig 6: Tweets Cleaning

At above screen we will see all extraordinary images and Halt words and expels through the tweets and as it wereclean wordsare there and presently press on "Run MLalgorithum button to foresee assumptions

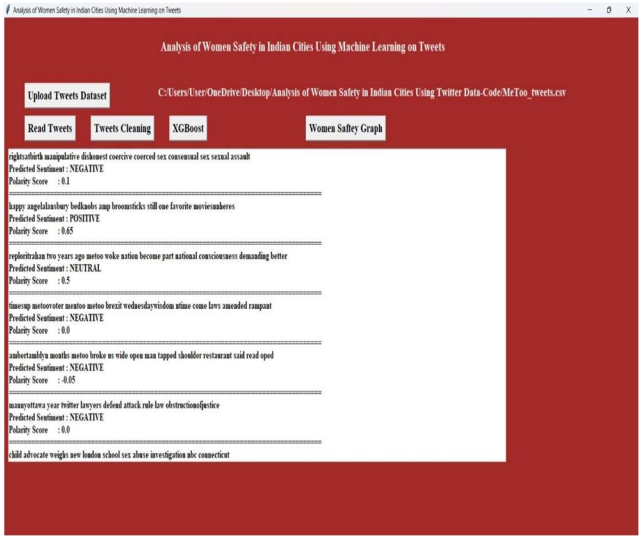


Fig 7: XG Boosting Algorithm

See all extraordinary images and halt words expefrom tweets and as it were clean words are there and presently press at 'Run ML Algorithm' button to foresee assumptions.

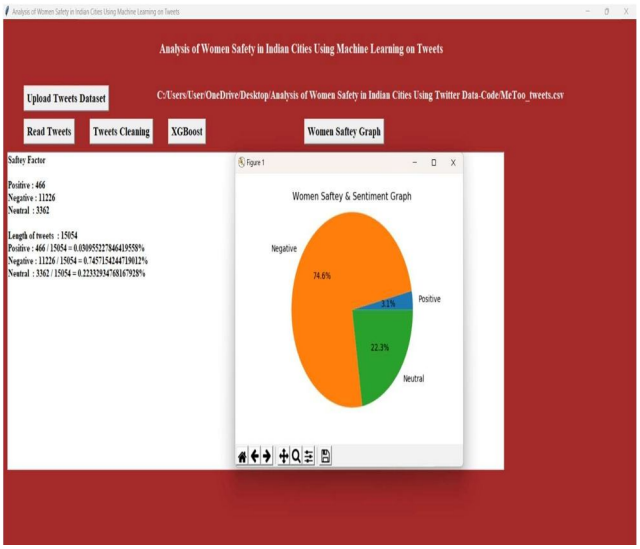


Fig 8: Women Safety Graph

In above activity 0.74 multiplied by 100 which gives 74%that means 74% of audience are talking negatively and area is not safe and only 22 and 3% of audience are talking positively.

VII. CONCLUSION

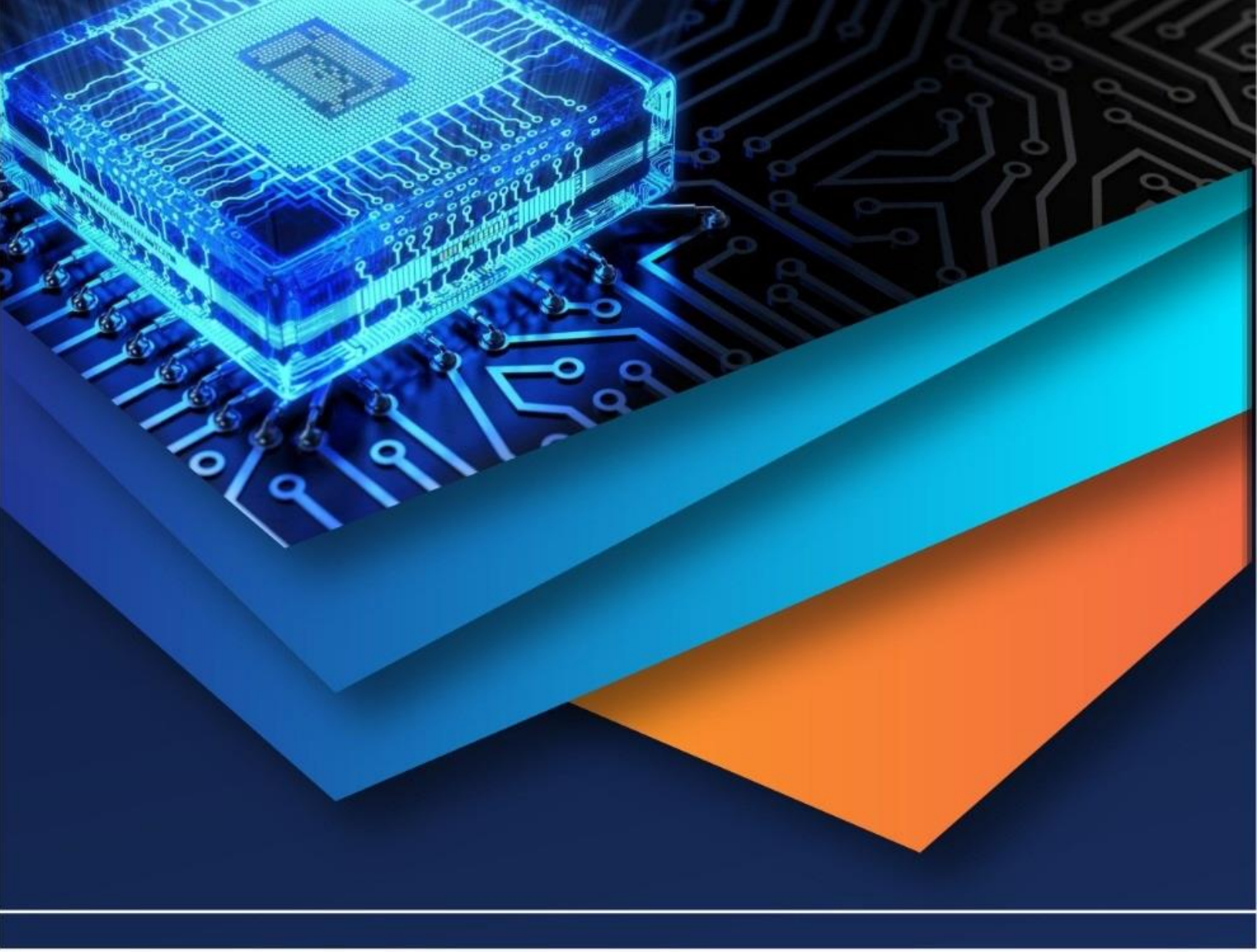
All through the term sheet we utilize examined almost various ML calculations which could offer assistance to organizing and analyzing the colossal sum of Tweeting information gotten counting huge number of Texts and content messages shared each cycle.

This ML calculations are exceptionally compelling and valuable as it arrives to analysis with expansive sum of info include with SPC calculation and direct logarithmic Calculate Show approaches which offer assistance to assist categorize the information into significant bunches.

Back VM with however another shape at ML calculation that's exceptionally prevalent in extricating Valuable data through the Tweeting and receive a thought around at stats of ladies security in India.

REFERENCES

[1] Apoorv Agarwal, Fadi Biadsy,and Kathleen R. Mckeown. "Contextual phrase-level polarity analysis usinglexical affect scoring and syntactic n-grams." Proceedings of 12th Conference of European Chapter of the Association for Computational Linguistics. Association for the ComputationalLinguistics, 2009.
[2] Barbosa, L., & Feng, J. (2010, August). Robusted sentimental detection on twitter from biased and noisy data. InColing 2010: Posters (pp. 36-44).
[3] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G.,Badhani, P., & Tech, B. (2017). Study for Twitter sentiment analysis using ML algorithms on Python. International Journalof Computer Applications, 165(9), 29-34.
[4] Mamgain,N., Mehta, E., Mittal, A., & Bhatt, G. (2016, March). Sentimental analyzing of top colleges in India utilizing Twitter data. In 2016 international conference on computational techniques in information and communication technologies (ICCTICT) (pp. 525-530). IEEE.
[5] Sahayak, V., Shete, V., & Pathan, A. (2015). Sentimentanalysis on twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.
[6] Charniak, E., & Johnson, M. (2005, June). Coarse-to- fine n-best parsing and maxent discriminative reranking. In Proceedings of the 43rd Annual Meeting of the Association forComputational Linguistics (ACL'05) (pp. 173-180).
[7] Kim, S. M., & Hovy, E. (2004). Deterninating the sentiments of opinions. In COLING 2004: Proceedings of the 20th International Conference at Computational Linguistics (pp. 1367-1373).
[8] Bermingham, A., & Smeaton, A. F. (2010, October). Classifying sentiment in microblogs: is brevity an advantage?.In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 1833-1836).
[9] Klein, D., & Manning, C. D. (2003, July). Accurately unlexicalized parsing. In Proceeding with the 41st annually meeting of the association for computational linguistics (pp. 423-430).
[10] Gamon, M. (2004). Sentimental classifications oncustomers feedback data: noise data, largely feature vectors, and the role of linguistics of analysis. In COLING 2004:Proceeding of 20th international conference on computationallinguistics (pp. 841-847).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS Researcher ID: N-9681-2016

10.22214/IJRASET doi cross ref

TOGETHER WE REACH THE GOAL SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET59199, entitled

Analysis of Women Safety in Indian Cities Using Machine Learning

by

Dr. G. Ravi Kumar

after review is found suitable and has been published in Volume 12, Issue III, March 2024

in

International Journal for Research in Applied Science & Engineering Technology

(International Peer Reviewed and Refereed Journal)

Good luck for your future endeavors

Editor in Chief, iJRASET

ISSN No. : 2321-9653

IJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS Researcher ID: N-9681-2016

10.22214/IJRASET

TOGETHER WE REACH THE GOAL

SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET59199, entitled

Analysis of Women Safety in Indian Cities Using Machine Learning

by

K. Venkata Giridhar

after review is found suitable and has been published in Volume 12, Issue III, March 2024

in

International Journal for Research in Applied Science & Engineering Technology

(International Peer Reviewed and Refereed Journal)

Good luck for your future endeavors

Editor in Chief, iJRASET

ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS Researcher ID: N-9681-2016

10.22214/IJRASET doi cross ref

TOGETHER WE REACH THE GOAL SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET59199, entitled

Analysis of Women Safety in Indian Cities Using Machine Learning

by

M. Vaishnavi

after review is found suitable and has been published in Volume 12, Issue III, March 2024

in

International Journal for Research in Applied Science & Engineering Technology

(International Peer Reviewed and Refereed Journal)

Good luck for your future endeavors

Editor in Chief, iJRASET

ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS Researcher ID: N-9681-2016

10.22214/IJRASET doi cross ref

TOGETHER WE REACH THE GOAL SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET59199, entitled

Analysis of Women Safety in Indian Cities Using Machine Learning

by

N Yesumani

after review is found suitable and has been published in Volume 12, Issue III, March 2024

in

International Journal for Research in Applied Science & Engineering Technology

(International Peer Reviewed and Refereed Journal)

Good luck for your future endeavors

Editor in Chief, iJRASET