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

<u>Submitted by</u>

K. Venkata Giridhar
21H55A0509

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 K.Venkata Giridhar 21H55A0509, 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**
Associate Professor   **Associate Professor and HOD**
Dept. of CSE   **Dept. of CSE**

ACKNOWLEDGEMENT

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

We are grateful to **Dr.G.Ravi Kumar Associate Proffesor** , 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 and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and 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 and 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, CMR Group of Institutions, for his continuous care.

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 and indirectly in completion of this project work.

<div align="right">

K.Venkata Giridhar   21H55A0509
M.Vaishnavi   21H55A0514
N.Yesumani   21H55A0515

</div>

INDEX

List of Figures

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

CHAPTER 1
INTRODUCTION

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.

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 behaviour of men and

thus making them uncomfortable to walk freely in public.
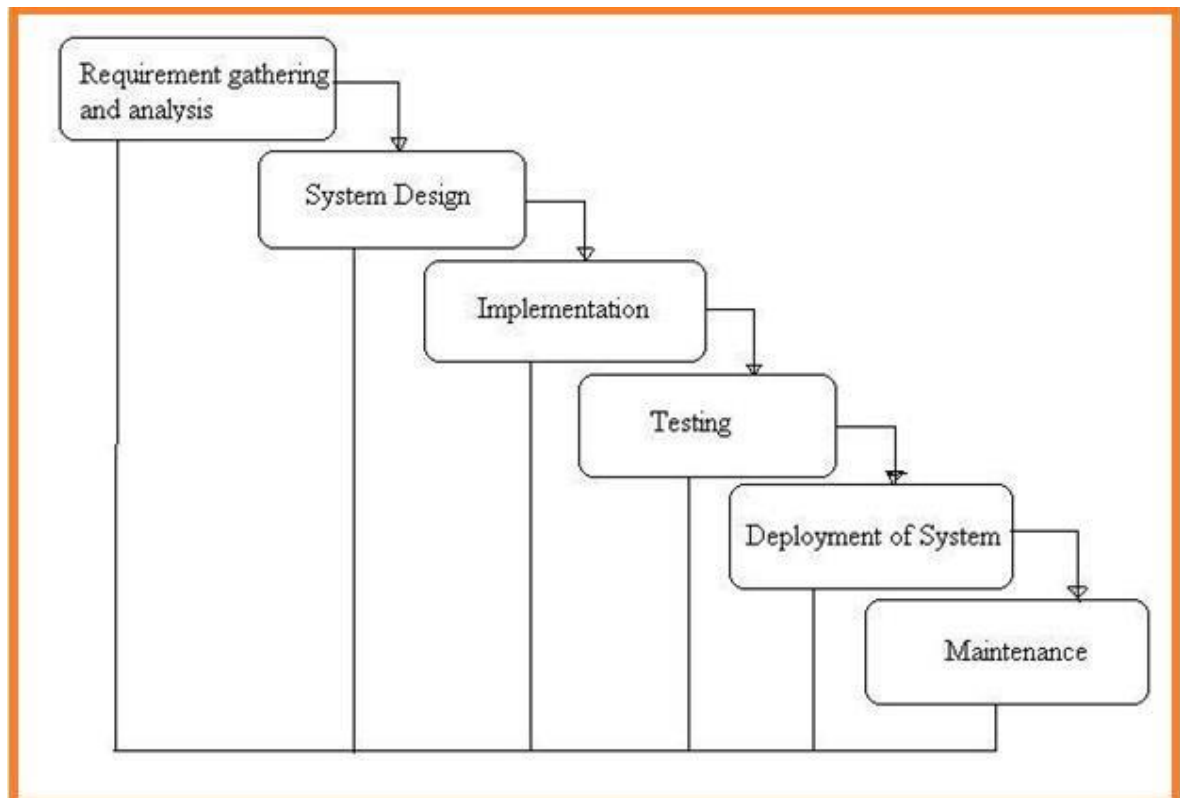
STRUCTURE OF PROJECT (SYSTEM ANALYSIS)

Fig: 1.1 Project SDLC

- ➢ Project Requisites Accumulating and Analysis
- ➢ Application System Design
- ➢ Practical Implementation
- ➢ Manual Testing of My Application
- ➢ Application Deployment of System
- ➢ Maintenance of the Project

SYSTEM DESIGN

In System Design has divided into three types like GUI Designing, UML Designing with avails in development of project in facile way with different actor and its utilizer case by utilizer case diagram, flow of the project utilizing sequence, Class diagram gives information about different

6

class in the project with methods that have to be utilized in the project if comes to our project our UML Will utilizable in this way The third and post import for the project in system design is Data base design where we endeavor to design data base predicated on the number of modules in our project

IMPLEMENTATION

The Implementation is Phase where we endeavor to give the practical output of the work done in designing stage and most of Coding in Business logic lay coms into action in this stage its main and crucial part of the project

TESTING UNIT TESTING

It is done by the developer itself in every stage of the project and fine-tuning the bug and module predicated additionally done by the developer only here we are going to solve all the runtime errors

MANUAL TESTING

As our Project is academic Leave, we can do any automatic testing so we follow manual testing by endeavor and error methods

1.1. PROBLEM STATEMENT

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 have.

1.2. RESEARCH 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. Expressing views on social media, expressing on micro blogging websites like tweeter is quite common in these days. A lot of people take it to social media to express their views about everything which is going right or wrong in our society and which is happening in day to day life. Woman safety is one of the many things which many people talk and express their views about on social media. Most people talk positive things, pointing out the certain change which is needed in our society that can drive the negativity out of our neighborhood and make women feel safe again. There will be X men and Y women who will tweet about women safety once or twice a day, across the country which can be used as a dataset. Using this dataset, it is quite common to run an analytical algorithm on the extracted data from social media and categorize them in positive and negative aspects.

Agarwal, Apoorv, Fadi Biadsy, and Kathleen R. Mckeown. "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams." Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics.Association for Computational Linguistics, 2009. We present a classifier to predict contextual polarity of subjective phrases in a sentence. Our approach features lexical scoring derived from the Dictionary of Affect in Language (DAL) and extended through WordNet, allowing us to automatically score the vast majority of words in our input avoiding the need for manual labeling. We augment lexical scoring with n-gram analysis to capture the effect of context. We

combine DAL scores with syntactic constituents and then extract ngrams of constituents from all sentences. We also use the polarity of all syntactic constituents within the sentence as features. Our results show significant improvement over a majority class baseline as well as a more difficult baseline

We present a classifier to predict contextual polarity of subjective phrases in a sentence. Our approach features lexical scoring derived from the Dictionary of Affect in Language (DAL) and extended through WordNet, allowing us to automatically score the vast majority of words in our input avoiding the need for manual labeling. We augment lexical scoring with n-gram analysis to capture the effect of context. We combine DAL scores with syntactic constituents and then extract ngrams of constituents from all sentences. We also use the polarity of all syntactic constituents within the sentence as features. Our results show significant improvement over a majority class baseline as well as a more difficult baseline consisting of lexical n-grams. 1 Introduction Sentiment analysis is a much-researched area that deals with identification of positive, negative and neutral opinions in text.

The task has evolved from document level analysis to sentence and phrasal level analysis. Whereas the former is suitable for classifying news (e.g., editorials vs. reports) into positive and negative, the latter is essential for question-answering and recommendation systems. A recommendation system, for example, must be able to recommend restaurants (or movies, books, etc.) based on a variety of features such as food, service or ambience. Any single review sentence may contain both positive and negative opinions, evaluating different features of a restaurant. Consider the following sentence (1) where the writer expresses opposing sentiments towards food and service of a restaurant. In tasks such as this, therefore, it is important that sentiment analysis be done at the phrase level. (1) The Taj

has great food but I found their service to be lacking. Subjective phrases in a sentence are carriers of sentiments in which an experiencer expresses an attitude, often towards a target. These subjective phrases may express neutral or polar attitudes depending on the context of the sentence in which they appear. Context is mainly determined by content and structure of the sentence. For example, in the following sentence (2), the underlined subjective phrase seems to be negative, but in the larger context of the sentence, it is positive.1 (2) The robber entered the store but his efforts were crushed when the police arrived on time. Our task is to predict

contextual polarity of subjective phrases in a sentence. A traditional approach to this problem is to use a prior polarity lexicon of words to first set priors on target phrases and then make use of the syntactic and semantic information in and around the sentence to make the final prediction. As in earlier approaches, we also use a lexicon to set priors, but we explore new uses of a Dictionary of Affect in Language (DAL) (Whissel, 1989) extended using WordNet (Fellbaum,

1998). We augment this approach with n-gram analysis to capture the effect of context. We present a system for classification of neutral versus positive versus negative and positive versus negative polarity (as is also done by (Wilson et al., 2005)). Our approach is novel in the use of following features: • Lexical scores derived from DAL and extended through WordNet: The Dictionary of Affect has been widely used to aid in interpretation of emotion in speech (Hirschberg 1We assign polarity to phrases based on Wiebe (Wiebe et al., 2005); the polarity of all examples shown here is drawn from annnotations in the MPQA corpus. Clearly the assignment of polarity chosen in this corpus depends on general cultural norms. 24 et al., 2005). It contains numeric scores assigned along axes of pleasantness, activeness and concreteness. We introduce a method for setting numerical priors on words using these three axes, which we refer to as a "scoring scheme" throughout the paper. This scheme has high coverage of the phrases for classification and requires no manual intervention when tagging words with prior polarities. • N-gram Analysis: exploiting automatically derived polarity of syntactic constituents We compute polarity for each syntactic constituent in the input phrase using lexical affect scores for its words and extract n-grams over these constituents. N-grams of syntactic constituents tagged with polarity provide patterns that improve prediction of polarity for the subjective phrase. • Polarity of Surrounding Constituents: We use the computed polarity of syntactic constituents surrounding the phrase we want to classify. These features help to capture the effect of context on the polarity of the subjective phrase. We show that classification of subjective phrases using our approach yields better accuracy than two baselines, a majority class baseline and a more

difficult baseline of lexical n-gram features. We also provide an analysis of how the different component DAL scores contribute to our results through the introduction of a "norm" that combines the component scores, separating polar words that are less subjective (e.g., Christmas , murder) from neutral words that are more subjective (e.g., most, lack). Section 2 presents an overview of previous work, focusing on phrasal level sentiment analysis. Section 3 describes

the corpus and the gold standard we used for our experiments. In section 4, we give a brief description of DAL, discussing its utility and previous uses for emotion and for sentiment analysis. Section 5 presents, in detail, our polarity classification framework. Here we describe our scoring scheme and the features we extract from sentences for classification tasks. Experimental set-up and results are presented in Section 6. We conclude with Section 7 where we also look at future directions for this research. 2 Literature Survey The task of sentiment analysis has evolved from document level analysis (e.g., (Turney., 2002); (Pang and Lee, 2004)) to sentence level analysis (e.g., (Hu and Liu., 2004); (Kim and Hovy., 2004); (Yu and Hatzivassiloglou, 2003)). These researchers first set priors on words using a prior polarity lexicon. When classifying sentiment at the sentence level, other types of clues are also used, including averaging of word polarities or models for learning sentence sentiment. Research on contextual phrasal level sentiment analysis was pioneered by Nasukawa and Yi (2003), who used manually developed patterns to identify sentiment.
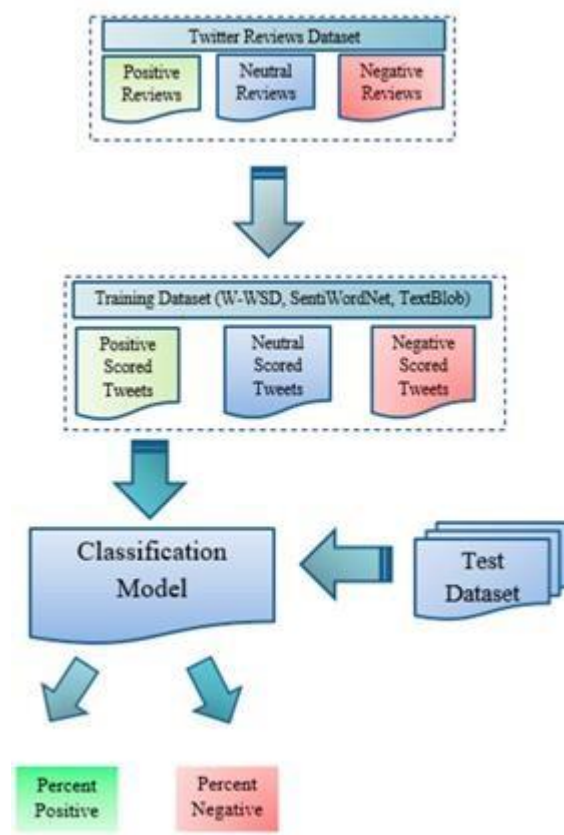

Fig.1.2: Twitter sentiment analysis.

Their approach had high precision, but low recall. Wilson et al., (2005) also explore contextual phrasal level sentiment analysis, using a machine learning approach that is closer to the one we present. Both of these researchers also follow the traditional approach and first set priors on words using a prior polarity lexicon. Wilson et al. (2005) use a lexicon of over 8000 subjectivity clues, gathered from three sources ((Riloff and Wiebe, 2003); (Hatzivassiloglou and McKeown, 1997) and The General Inquirer2). Words that were not tagged as positive or negative were manually labeled. Yi et al. (2003) acquired words from GI, DAL and WordNet. From DAL, only words whose pleasantness score is one standard deviation away from the mean were used. Nasukawa as well as other researchers (Kamps and Marx, 2002)) also manually tag words with prior polarities. All of these researchers use categorical tags for prior lexical polarity; in contrast, we use quantitative scores, making it possible to use them in computation of scores for the full phrase. While Wilson et al. (2005) aim at phrasal level analysis, their system actually only gives "each clue instance its own label" [p. 350]. Their gold standard is also at the clue level and assigns a value based on the clue's appearance in different expressions (e.g., if a clue appears in a mixture of negative and neutral

expressions, its class is negative). They note that they do not determine subjective expression boundaries and for this reason, they classify at the word level. This approach is quite different from ours, as we compute the polarity of the full phrase. The average length of the subjective phrases in the corpus was 2.7 words, with a standard deviation of 2.3. Like Wilson et al. 2 http://www.wjh.harvard.edu/ inquirer 25 (2005) we do not attempt to determine the boundary of subjective expressions; we use the labeled boundaries in the corpus. 3 Corpus We used the Multi-Perspective QuestionAnswering (MPQA version 1.2) Opinion corpus (Wiebe et al., 2005) for our experiments. We extracted a total of 17,243 subjective phrases annotated for contextual polarity from the corpus of 535 documents (11,114 sentences). These subjective phrases are either "direct subjective" or "expressive subjective". "Direct subjective" expressions are explicit mentions of a private state (Quirk et al., 1985) and are much easier to classify. "Expressive subjective" phrases are indirect or implicit mentions of private states and therefore are harder to classify. Approximately one third of the phrases we extracted were direct subjective with non-neutral expressive intensity whereas the rest of the phrases were expressive subjective. In terms of polarity, there were 2779 positive, 6471 negative and 7993 neutral expressions. Our Gold Standard is the manual annotation tag given to phrases in the corpus. 4 DAL DAL is an English

language dictionary built to measure emotional meaning of texts. The samples employed to build the dictionary were gathered from different sources such as interviews, adolescents' descriptions of their emotions and university students' essays. Thus, the 8742 word dictionary is broad and avoids bias from any one particular source. Each word is given three kinds of scores (pleasantness – also called evaluation, ee, activeness, aa and imagery, ii) on a scale of 1 (low) to 3 (high). Pleasantness is a measure of polarity. For example, in Table 1, affection is given a pleasantness score of 2.77 which is closer to 3.0 and is thus a highly positive word. Likewise, activeness is a measure of the activation or arousal level of a word, which is apparent from the activeness scores of slug and energetic in the table. The third score, imagery, is a measure of the ease with which a word forms a mental picture. For example, affect cannot be imagined easily and therefore has a score closer to 1, as opposed to flower which is a very concrete and therefore has an imagery score of 3. A notable feature of the dictionary is that it has different scores for various inflectional forms of a word (affect and affection) and thus, morphological parsing, and the possibility of resulting errors, is avoided. Moreover, Cowie et al., (2001) showed that the three scores are uncorrelated; this implies that each of the three scores provide complementary information

1.3 PROJECT SCOPE AND LIMITATIONS

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

he project's scope encompasses a holistic approach to address women's safety in Indian cities. By analyzing various aspects, from data to infrastructure, legal frameworks, socio-cultural factors, and technology, the project aims to provide actionable recommendations and implement pilot programs that can contribute to safer urban environments for women. Collaboration with stakeholders and continuous evaluation will be essential for the project's success.

This revolves around leveraging machine learning and data-driven approaches to enhance women's safety in Indian cities. By predicting safety concerns, identifying hotspots, and developing real-time monitoring systems, the project aims to provide proactive and actionable

insights for creating safer urban environments. Collaboration with stakeholders, privacy considerations, and continuous evaluation are vital components of the project's success.

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 and wordbook- grounded literacy.

CHAPTER 2

CHAPTER 2
BACKGROUND WORK

EXISTING METHOD 1:

2.1 INTRODUCTION:

Particular kinds of provocation and Savagery are extremely forceful including gazing and passing remarks and those inadmissible practices are generally considered a regular piece of the metropolitan lifestyles. There have been a few investigations that have been led in urban areas across India and ladies report comparative sort of lewd behavior and passing off remarks by other obscure individuals. The assessment that became led during maximum Metropolitan city organizations of India which incorporates Delhi, Mumbai, and Pune, it became validated that 60 % of the women experience risky while going out to paintings or at the same time as going out withinside the open car. women have the right to the town because of this that they may flow freely every time they want whether or not or now no longer or now no longer it`s too an educational Institute or each different vicinity ladies need to head. however, women revel in that they`re unstable in places like branch shops, shopping for branch shops on their manner to their activity vicinity because of the several unwell known body shaming and harassing the one's women element safety or loss of concrete consequences in the existence of women is the primary motive of harassment of ladies. There are times the same time as the harassment of women became finished via way of means of the usage of their neighbors at the same time as they have been on the way to school or there was a loss of protection that created a feeling of worry withinside the minds of small women who at some point of their life go through because of that one instance that passed off of their lives wherein they had been forced to do something unacceptable or became sexually burdened thru one in every one of their neighbors or any other unfamous character. Most secure towns method women safety from an attitude toward women's rights to the effect the metropolis without the worry of violence or sexual harassment. It is a social responsibility to hide women's desire for protection, rather than imposing restrictions on women as a whole, and women and women have the same rights as men covered in big cities. Twitter text collection analysis also calls individual female names to fight sexual harassment and unethical behavior by male agents in Indian cities included. Information on women's perceptions of safety in Indian society by using a series of records obtained on Twitter, omitting

the value 0, and processing with Laplace through systematic insights into the algorithm. Organized. To win. Applying Porter's ideas is a form of evaluation. It collects records, retweets, and redundant statistics from the retrieved datasets to provide a smooth and reliable overview of the recognition of Women's Protection in Indian society. Extinguish the fire

2.2 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 and designation of women who viewpoint in contrast to abuse importunity. iv. With the algorithms used, further security is handed.

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.

2.3 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 graphmodel, 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

17

easily.

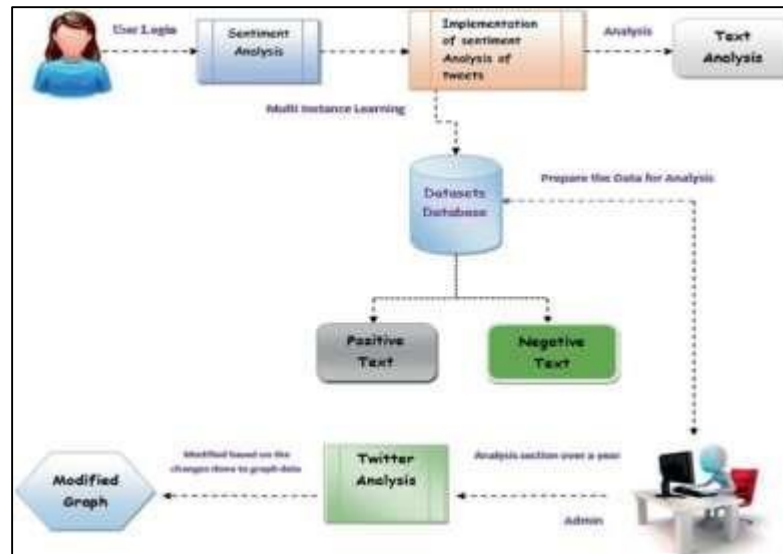These promptions are delicate to prize the

Fig 2.1: Architecture view of Sentimental Analysis for women safety

The analysis can be accomplished by admin now. The hash tags made by each stoner, procedure is entitled as, original involvement for the sentiment study and henceforth and drive the trained set of data. End to end, textbook scrutiny diagram is exposed. Admin hoards the pollutants at the backend database. The important words for which the tweet texts atmosphere is examined to assert as vituperative Re filters. There are two types of tweet hashtags – positive tags and negative tags. Those words which are vituperative or discourteous the women are positive words.

The keywords which are abusive and that misuse women are negative words. Tweet tags are classified as positive noveletish analysis and vituperative to women, if any of the positive tag is set up in the tweet, it'll be confidential and the negative sentimental analysis are vituperative to women, if negative keyword is established up in the tweet . There are a list of all the tweets in the application that are offensive to women under negative sentimental 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.

Algorithm

SENTIMENTAL ANALYSIS Sentiment analysis is the process of extracting the sentiment

18

behind any sentence or statement. It can be called as a classification technique which is used to 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. Whereas lexicon learning based approach uses the vocabulary and scoring method to detect opinions. 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.

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()

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 abusing words and to remove individual's tweet texts.

- ➢ Run Machine Learning Algorithm: This Module is used to prognosticate opinions from tweets where respective tweet having tweet textbook and also demonstrating tweetsentiments with opposition slash.

Women Safety Graph: Women Safety Graph module is used to contract the results and by seeing that product, stoner can fluently comprehend whether zone is safe or not. If area is safe, people will prompt positive or

Implementation

To run this project, install python software and also run below commands to install packages.

Pip install Textblob

Pip install matplot

Pip install pandas == 0.25.3

After installing above packages double click on

'download_nltk' train and also a window will appear and also click on 'download' button from that window and stay for 10 twinkles to allow operation to download all TEXTBLOB corpora packages. Once all packages download also that window turn to green color to indicate download process complete and also you can close that window.

Testing

The method of testing is where the assessment data is set and is cast-off for testing the segments collectively and latterly the confirmation assumed for the turfs. Also the system testing receipts the area which makes definite that all factors of the scheme assets function as a component. The assessment statistics must be preferred similar that it conceded over all probable circumstances. The methods of testing followed in this project are:

➢ SYSTEM TESTING- It is demarcated as testing of a whole and entirely unified software package product.

➢ MODULE TESTING- Module testing defines software testing type, which checks discrete subprograms, measures, classes, or procedures in a package.

➢ ACCEPTANCE TESTING- The testing sector where all the modules undergo various methods of testing and are finally accepted.

EXISTING MENTHOD 2

2.4 INTRODUCTION

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. 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 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.

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 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 class of entities of the tweets using the programming 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. Approaches of sentimental analysis can be broadly differentiated into two types – machine learning based and lexicon learning based.

 Machine learning approach includes the process of extraction of features, programming model training using dataset of features. Whereas lexicon learning based approach uses the vocabulary and scoring method to detect opinions. In this paper, we use machine learning approach.

22

Collection of data, pre-processing the data, extraction of features, choosing base features, detection of sentiments and classification of sentiments using machine learning approaches or simple computations are the basic steps to perform sentimental analysis.
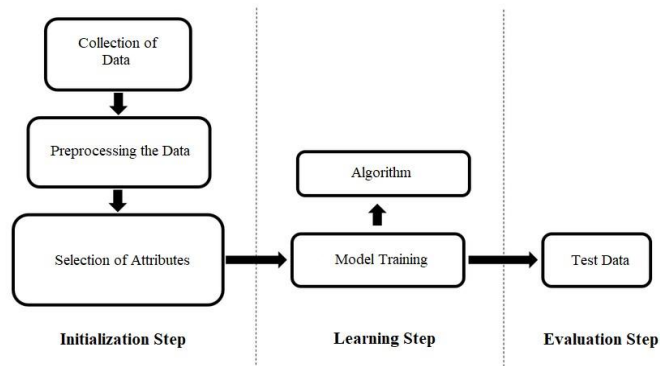
Analysis of Sentimental Data

Fig -2.2: Process of Analysis

The process of obtaining the sentiments of tweet includes five steps:

2.5 MERITS & DEMERITS

Data extraction: First step involved in analysis of sentiment is the collection of information from the social network website like twitter. This helps in extracting the tweet message but this message also includes extra data like tweets likes, dislikes and comments.

Text Cleaning: Once the data is extracted from the twitter source as the datasets, this information has to be passed to the classifier. The classifier cleans the dataset by removing redundant data like stop words, emoticons in order to make sure that non textual content is identified and removed before the analysis.

Sentiment Analysis: After the classifier cleans the dataset, the data is ready for the sentimental analysis process. Machine learning and Lexicon based learning and Hybrid learning are some of the approaches of sentimental analysis. There are also some other approaches such as Nero Linguistic Programming and Natural Language Processing. Training the dataset and then testing that trained dataset involves in machine learning approach. Training data and Testing

data are useful for the classifier to perform the algorithm. Maximum Entropy, Naives Bayes classification, Bayesian Networks and Network Support Vector Machine are some of the algorithm which can be used to train the classifier. Testing data is used to identify the efficiency of the sentiment classifier.

In case of Lexicon based leaning, training dataset is not used. This approach uses a built-in dictionary in which words associated with sentiments of human are present. The third approach, which is the Hybrid learning, combines both machine leaning approach and lexicon learning approach in order to improve the performance of classifier.

Sentiment Classification: At this step, the dataset is ready for the classification. Each and every sentence of the tweet will be examined and opinion will be formed accordingly for subjectivity. Subjective expression sentences are retained and those of objective expression sentences are rejected. Techniques like Unigrams, Negation, Lemmas and so on are used at different levels of sentimental analysis. Sentiments can be distinguished broadly into two groups – Positive and Negative. At this point of sentimental analysis, each of the subjective sentences which will be retained are classified into good, bad or like, dislike or positive and negative.

Output Presentation: To generate useful and meaningful information out of the raw data, sentimental analysis plays vital role. Once the algorithm is completed, the outcome of the analysis can be visualized by creating different types of graphs. Bar graphs, Time series and Pie charts are some of the examples which can be used to display the output. To measure the sentiment of the tweets in terms of Positive and Negative, Bar graphs can be used. Similarly, to measure in terms of likes, dislikes, average length of tweet for a certain period, Time series can be used. To obtain the initial source of the tweet, pie charts can be used.
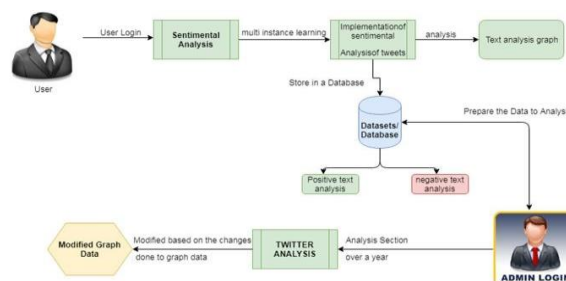
System Architecture

Fig -2.3: System Architecture

24

2.6 IMPLEMENTATION

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

There can be 'n' number of positive and negative keywords stored in the database. When the admin implements the sentimental analysis, every keyword in the database will be compared with each and every word in the tweet of the user. If any one of the positive keyword is found in the tweet, that tweet will be classified as positive sentimental analysis and these are abusive to women. If negative keyword is found in the tweet, it will be classified as the negative sentimental analysis which is not abusive to women. Hence, by this stage there will be two types of sentimental analysis made based on the filter in the database. Under positive sentimental analysis, there will be a list of all the tweets in the application that are abusive to women. Similarly, under negative sentimental analysis there will be a list that is clean and are not abusive tweets. Along with the tweet context, user details will also be provided at each of the analysis list.

PROPOSED SYSTEM

Women have the right to the city which means that they can gofreely 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 isthe 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 ofwomen 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 CityIn this project.

we will propose an idea for evaluating women's safety using social networking messages and identifying which areas are safe and which areas are risky for women using machine learning algorithms. We'll be using Python and NLTK (Natural Language Tool Kit) to download tweets concerning women's safety via MEETOO.We then use the XG Boost method, a famous Machine Learning algorithm, to evaluate the tweets and develop a women safety graph as an output. This project might potentially be put as an add-on to social networking apps so that females can identify whether conversing with a certain individual is safe or unsafe for them. By this project, we can identify risky places in metropolitan cities and increase safety within these areas as a result of our research, allowing women to feel safe to go wherever they want..

Fig 2.4: Sequence Diagram

27

ADVANTAGES:

➤ 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

CHAPTER 3

CHAPTER 3

RESULT

RESULT

Fig .3.1 Download nltk file

Fig :3.2 Output Window
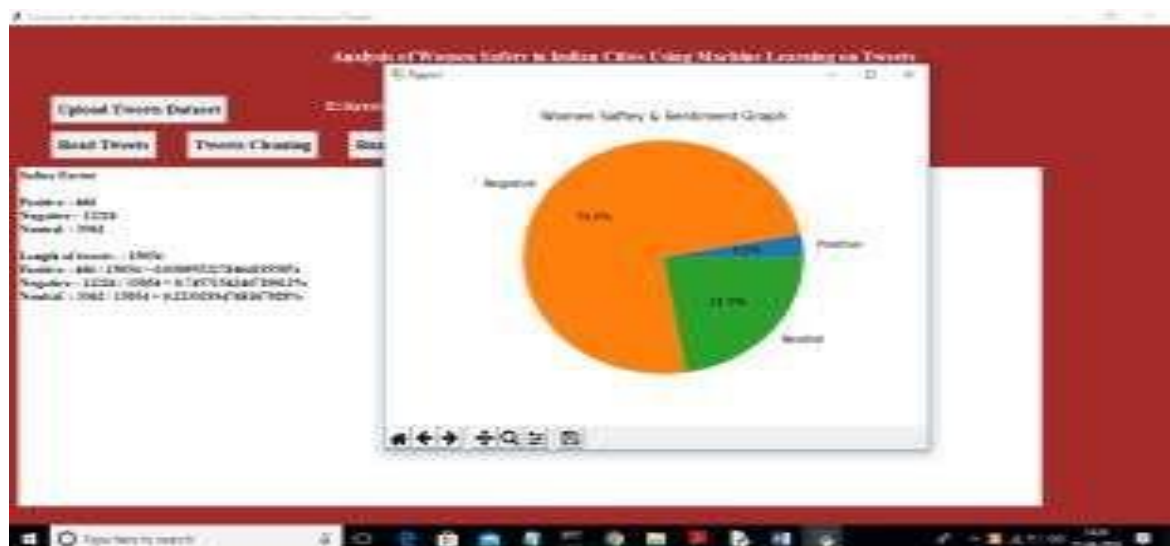
Fig :3.3 Output window

Fig : 3.4 Women safety graph

CHAPTER 4

CHAPTER 4
CONCLUSION

CONCLUSION

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

REFERENCES

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

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

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

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

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

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

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

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