



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

**Volume 9, Issue 6, June 2021**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.542**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

# Women Safety Analysis Based on Social media

Sk. Shahan<sup>1</sup>, R.Jagadeesh<sup>2</sup>, R. Narendra<sup>3</sup>, V. Tejaswini<sup>4</sup>, Mrs. D. Deepthi<sup>5</sup>

UG students, Dept. of CSE, Vasireddy Venkatadri Institute of Technology, Guntur, India<sup>1,2,3,4</sup>

Asst. Professor, Dept. of CSE, Vasireddy Venkatadri Institute of Technology, Guntur, India<sup>5</sup>

**ABSTRACT:** Women and girls are facing many problems in India by modern society. Even after 33% reservation in social contribution, with increased violence and sexual harassment in public places & workplaces, they live hectic lives. It also sets many unanswered questions to the government and society about women safety and security. This paper focuses on the different forms of violence and threats against the woman and to increase the responsibility of the citizens of India about the safety of the woman and Girls surrounding them with different reference to the social networking sites like Twitter, Facebook and Instagram with the help of Tweets, comments, blogs and posts on the particular incident against woman or girl can be used for the scrutiny. Tweets on Twitter usually contains icons, emojis, messages and quotes that focus on women's and girls' safety in Indian cities and to create awareness among individuals to take severe action and punish those who harass the women. The Twitter platform provides hash-tag messages to express their views about the place's safety while going out for work or travelling in public transport surrounded by unknown persons. The model uses NLTK techniques in machine learning to classify the Tweets. This model would recommend various guidelines and preventive efforts to establish by the government and public awareness to save the woman from the other violence in sexual harassment and alert the different situations.

**KEYWORDS:** women safety, child safety, sexual harassment, twitter sentiment analysis.

## I. INTRODUCTION

India is a Great Country called Punyabhoomi, Veda Bhoomi. Women depicted as Goddess like the slogan "Yatranaryastupujyanteramayantetra,Devatayatraistaastunapujyantesarvaastatrafalaahkriyaah" which means women are honoured where godliness blossoms there, and where women are disgraced, all action, no matter how noble, remain barren. Where the women worshipped there, God resides. But In modern Society, Women are facing many problems like Sexual harassments and passing comments by unknown persons. In India, the survey conducted in most Metropolitan cities shows that 60% of girls and women feel unsafe while going out to work or travelling on public transport. In some situations, girls are facing harassment from their neighbours or an unknown person. Because of these harassments and violence, small girls suffer throughout their lifetime, leading to psychological disturbances. It is the primary responsibility of society to protect women and girls. To achieve this, people can use social media to express their feelings in text messages about unsafe places for girls and women.

Twitter is one of the most prominent social network websites and has been overgrowing. On Twitter, each person can share their thoughts in the form of tweets, through 140 characters which leads to people compacting their opinions by using different terminology, abbreviations, emojis, short forms and so on. Hence it is maintained to term the Twitter comments as undefined. Analysis of Twitter texts collection also includes people's name and the name of women who stand up against sexual harassment and unethical behaviour of men in Indian cities, making them feel uncomfortable to walk. Processed the data set obtained through Twitter about women's status in Indian society through machine learning techniques used to smoothen the data like removing special characters, retweets and categorizing the data set into positive, negative and neutral tweets. A clear and original view of women's safety status in Indian society obtained.

## II. LITERATURE SURVEY

Deepak Kumar, Shivani Aggarwal (2019) [1] proposed a model to classify the tweets as Positive, negative and neutral. It collected tweets using Twitter API and used the SVM algorithm for the classification of Tweets. S. Ramamoorthy, R. Poorvadevi(2019) [2] proposed a model using the K-means Machine Learning algorithm to classify various data sets collected from different social media sites regarding sexual harassments on women and providing measures and precautionary steps to alert the women from difficult circumstances. Bhumika Gupta and Kanika Vishwakarma (2017) [3] proposed a research paper to analyse data and retrieving sentiment that it embodies. They used Bayesian logistic regression and Naive Bayes algorithms to extract the negative and positive comments for tweets' processing and classification.

Ankur Goel, Jyoti Gautam, Sitesh Kumar (2016) [4] presented an approach to classify sentiments from Twitter messages. They used various Machine Learning algorithms to categorize the data into positive and negative classes based on their feelings, such as Baseline, Naive Bayes Classifier, Support Vector Machine etc.

Rouf Ahmad Bhat, Anita Deshpande (2017) [5] presented a paper on "An analytical study on Overview of Sexual Harassment of Women at the workplace in India," which covers various issues like various threats, crimes and sexual harassments on Women.

Varsha Sahayak, Vijaya Shete, Apashabi Pathan (2015) [6] proposed a method to classify different Twitter messages using a distant supervision machine learning algorithm

### III. PROPOSED SOLUTION AND METHODOLOGY

#### *Methodology:*

Women and girls have been experiencing a lot of violence and harassment in the society and this is leading to sexual harassment. Now-a-days all people are using social networking sites like Facebook, twitter, etc., to express their feelings. Twitter is one of such social networking platforms which helps to connect a large number of people and allows to share their thoughts and ideas. Using the twitter, we will get the tweets and based on those tweets we represent a pie chart of women and child safety.

#### A. *Twitter Sentiment Analysis:*

Twitter is the most moving web-based media stage where individuals can communicate their surveys, assessments, and feelings on any item, film or on any issue occurred in the general public. These tweets were recording and examined to mine kin's sentiments identified with recover tweets about disturbances on ladies and young ladies and discover feelings and extremity of tweets. To mine sentiments and extremity in tweets, regardless of whether positive, negative or impartial. Assessment examination is a book characterization strategy that isolates text into various notions.

Slant examination characterizes a technique that associates the mining of perspectives, thoughts, perspectives and sentiments from instant messages, discourse, tweets and other information base sources through Natural Language Processing (NLP). Slant investigation arranges text into classifications like "positive" or "negative" or "unbiased". It is likewise expressed as subjectivity examination, assessment mining, and evaluation extraction.

Slant Analysis incorporates numerous assignments like notion extraction, estimation arrangement, subjectivity characterization and synopsis of conclusions or assessment spam location. It expects to investigate individuals' feelings, considerations, feelings, etc., towards elements such as products, individuals, topics, organizations and services.

The pre-processed dataset has many distinctive properties. In the feature extraction method, we extract the aspects from the processed dataset. Later, this aspect used to compute the positive, negative and neutral polarity in a sentence that is useful for determining the individuals' opinion using models like unigram and bigram. Machine learning techniques require representing the key features of text or documents for processing.

These critical features considered as feature vectors used for the classification task. Some example features that have reported in the literature are:

#### 1. Words and Their Frequencies:

Unigrams, bigrams and n-gram models with their frequency count are considering as features. There has been more research on using word presence rather than frequencies to describe this feature better.

#### 2. Parts Of Speech Tags

Parts of speech like adjectives, adverbs, and some verbs and nouns are good indicators of subjectivity and sentiment. We can generate syntactic dependency patterns by parsing or dependency trees.

#### 3. Opinion Words and Phrases

Apart from specific words, some phrases and idioms which convey sentiments can use as features. e.g. cost someone an arm and a leg.

#### 4. Position of Terms

A term within a text can affect how much the word makes a difference in the text's overall sentiment.

#### 5. Negation

Negation is an essential but difficult feature to interpret. The presence of a negation usually changes the polarity of the opinion.

6. Syntax:

Syntactic patterns like collocations are used as features to learn subjectivity patterns by many of researchers.

B. Pre-processing of the data:

A tweet contains a lot of opinions about the data which are expressed in different ways by different users. The twitter dataset used in this survey work is already labeled into two classes viz. negative and positive polarity and thus the sentiment analysis of the data becomes easy to observe the effect of various features. The raw data having polarity is highly susceptible to inconsistency and redundancy. Preprocessing of tweet include following points,

- Remove all URLs (e.g. www.xyz.com), hash tags (e.g., #topic), targets (@username)
- Correct the spellings; sequence of repeated characters is to be handled
- Replace all the emoticons with their sentiment.
- Remove all punctuations, symbols, numbers
- Remove Stop Words
- Expand Acronyms (we can use an acronym dictionary)
- Remove Non-English Tweets

C. Classification:

Sentiment analysis is a machine learning tool that analyses texts for polarity, into positive, negative and neutral tweets. By training machine learning tools with examples of emotions in the reader, machines automatically learn how to detect sentiment without human input. Many techniques and complex algorithms are using to command and train machines to perform sentiment analysis. There are pros and cons to all of them. But, used together, they can provide exceptional results.

The proposed method of Sentiment Analysis of twitter data consists of two important parts.

Data Extraction pre-processing of extracted data and classification.

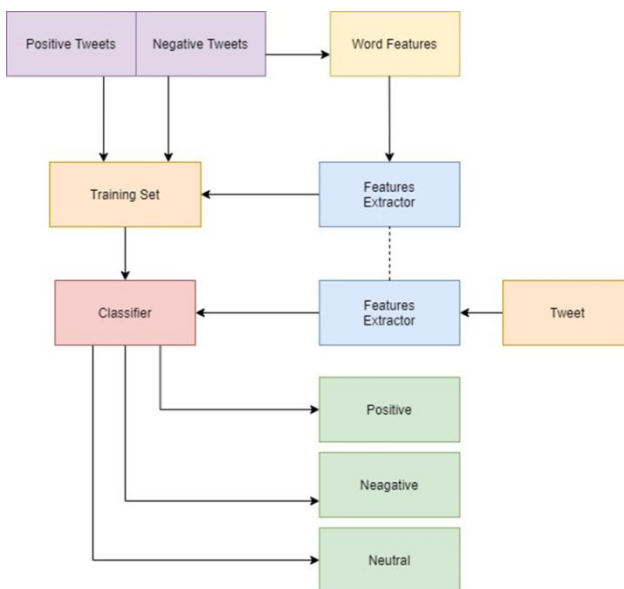


Fig1: twitter sentiment analysis architecture

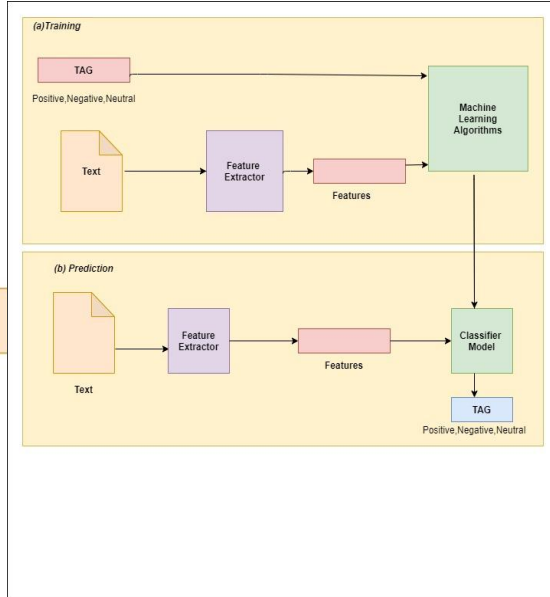


Fig 2: model of sentiment analysis

1. NAIVE BAYE'S ALGORITHM

The Naïve Bayes is one of the most improved classifications (classifier) methods. First, to perform a variety, we must select the features from the data set. The classifiers will process all the tweets in the data sets. A Naive Bayes classifier accepts that the presence of a particular element in a class is irrelevant to any other component.

A Naïve bayes algorithm is very easy to build up and mainly used for a large set of data. It provides a way of calculating  $p(c|x)$  from  $p(c)$ ,  $p(x)$  and  $p(x|c)$ .

Here is called the posterior probability and it is given by the formula

$$p(c|x) = \frac{p(x|c)p(c)}{p(x)}$$

Where

P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes)

p(c) is the prior probability of the class

p(x|c) is the likelihood which is the probability of predictor given class

P(x) is the prior probability of predictor

## 2. LINEAR REGRESSION

Linear regression is a statistical algorithm recycled to forecast a Y value, given X features. Suppose machine learning, the data sets investigated to show a relationship—the connections placed along the X/Y axis, with a straight line executing them to forecast farther relationships.

Linear regression determines how the X input (words and phrases) relates to the Y output (polarity). It will calculate where words and phrases fall on a scale of polarity from “really positive” to “really negative” and everywhere in between.

## 3. SUPPORT VECTOR MACHINE (SVM)

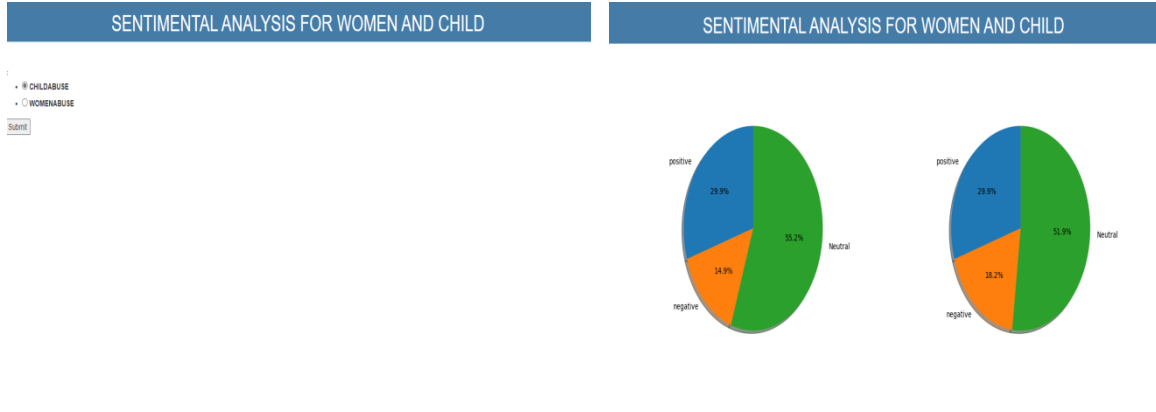
SVM is very efficient method for regression, classification and pattern recognition. It is known as a good classifier because of its high generalization performance without need prior knowledge and input space is very high. SVM is based on the concept of decision planes that defined decision boundary and point that form the decision boundary between the classes called support vector treat as parameter. The main aim of SVM is to find the best classification function to distinguish between two classes in the training data.

Support Vector Machine is a Machine Learning tool used for classification that is based on Supervised Learning which classifies points to one of two disjoint half-spaces. Support Vector Machine is a new classification method for both linear and non-linear data. Linear data can easily separate by two classes whereas non-linear data are not easily distinguished between classes. SVM will separate data between two hyperplanes. The main concept of SVM is found out best classifier between two classes.

### *Proposed solution:*

For the classification of twitter sentiments for women safety we are proposing solution that includes, installing the Tweepy package. This will access the twitter API with python. In orderto fetch the tweets through twitter API, one needs to login to the twitter account using the consumer key, consumer secret, access token and access token secret.This will authorize the authentication credentials using the authentication handler(OAuthHandler).After authenticating the credentials, we will retrieve the tweets using the twitter API. Using the re (regular expression) module we will remove all the links and special characters (symbols) that are included in those tweets. By using textblob we will classify the tweets based on sentiments. This textblob will return the polarity. Polarity is used to get the probability of the positive, negative and neutral tweets. Based on this polarity the percentage of the tweets will be calculated and the output will be shown as pie chart representation.Using the flask API, we build a web page. In that web page we are providing two options i.e., child abuse and women abuse. Based on those options we can choose accordingly.

#### IV. SIMULATION RESULTS



a. sentiment analysis for women and child

b. representation of women and child analysis

#### V. CONCLUSION AND FUTURE WORK

This paper discussed different machine learning techniques for sentiment analysis on vast data obtained from Twitter on Women Safety. In our proposed system, we used NLTK techniques such as textblob to classify Positive, Negative and Neutral Tweets. Even we can use Maximum Entropy, SPC and linear algebraic algorithm to achieve sentiment analysis to circulate awareness towards Women and Girl Safety. We can extend to apply these machine learning techniques on different social media platforms like Facebook and Instagram to achieve more safety and security for future enhancement.

#### REFERENCES

1. Deepak Kumar, Shivani. "Analysis of Women Safety in Indian Cities using Machine Learning on Tweets" Amity International Conference on Artificial Intelligence (AICAI),2019.
2. S. Ramamoorthy, R. Poorvadevi. "Safety Measures Against Women Violence in India using Sentimental Analysis". International Journal of Innovative Technology and Exploring Engineering (IJITEE), volume8,2278-3075, April 2019.
3. Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., & Badhani, P. (2017). "Study of Twitter sentiment analysis using machine learning algorithms on Python". International Journal of Computer Applications, 165(9), 0975-8887.
4. Ankur Goel, Jyoti Gautam, Sitesh Kumar. "Real Time Sentiment Analysis of Tweets Using Naive Bayes". 2nd International conference on Next Generation Computing Technologies (NGCT-2016) Dehradun, India 14-16 October 2016.
5. Rouf Ahmad Bhat, Anita Deshpande. "An Overview of Sexual Harassment of Women at Workplace in India: An Analytical Study". International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 7, July 2017.
6. Sahayak V., Shete V., Pathan, A. (2015). "Sentiment analysis on twitter data". International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.



**INNO**  **SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 7.542**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details