

# Analysis of Women Safety in Indian Cities Using Twitter Data

**Abstract:** These days women are encountering heaps of savagery like harassment in places in a few urban communities. This beginnings from following which at that point prompts oppressive harassment or additionally called misuse attack. In this paper we principally center around the job of online media which can be utilized to advance the wellbeing of women in India, given the exceptional reference to the interest of numerous web-based media sites or applications, for example, Twitter, Facebook and Instagram stages. This paper likewise centers around building up the obligations among the commoners on the different pieces of Indian urban communities so the wellbeing of women around them is guaranteed. Tweet on the Twitter application contains the instant messages, sound information, video information, pictures, smiley articulations and hash-labels. This tweet substance can be utilized to peruse among individuals and along these lines can instruct them to make severe moves if tweets are oppressive to women and thus can rebuff such individuals if the harassment is made. Applications which incorporate hash-labels, like Twitter and Instagram, can be utilized to spread the messages across the whole globe and cause the women to feel allowed to communicate their perspectives and sentiments. By this we can know the condition of their brain when they go out for work or travel in a public transportation or encompassed by

mysterious men and if it feels they are secure.

**Keywords:** Women, Safety, Sexual Harassment, Hash tag, Sentimental Analysis.

## 1. INTRODUCTION:

On the twitter, clients will impart their insights and point of view in the tweets segment. This tweet can just contain 140 characters, hence making the clients to smaller their messages with the assistance of contractions, slang, shot structures, emojis, and so forth Likewise, numerous individuals express their assessments by utilizing polysemy and mockery too. Thus, twitter language can be named as the unstructured. From the tweet, the supposition behind the message is separated. This extraction is finished by utilizing the wistful investigation system. Aftereffects of the wistful investigation can be utilized in numerous spaces like slants in regards to a specific brand or arrival of an item, breaking down popular conclusions on the public authority approaches, individual's considerations on women, and so forth to perform grouping of tweets and dissect the result, a great deal of study has been done on the information acquired by the twitter. We likewise survey a few investigations on AI in this paper and examination on the most proficient method to perform wistful examination utilizing that area on twitter information. The paper

extension is confined to AI calculation and models.

Gazing at women and passing remarks can be particular kinds of savagery and harassments and these practices, which are unsuitable, are generally ordinary particularly with respect to metropolitan life. Numerous explores that have been led in India shows that women have detailed inappropriate behavior and different practices as expressed previously. Such examinations have likewise shown that in well-known metropolitan urban communities like Delhi, Pune, Chennai and Mumbai, most women feel they are risky when encircled by obscure individuals. Via web-based media, individuals can uninhibitedly communicate what they feel about the Indian legislative issues, society and numerous different contemplations. Additionally, women can likewise share their encounters in the event that they have confronted any viciousness or lewd behavior and this unites blameless individuals to face such episodes. From the investigation of tweets text assortment got by the twitter, it incorporates names of individuals who has irritated the women and furthermore names of women or honest individuals who have remained against such savage demonstrations or untrustworthy conduct of men and hence making them awkward to walk openly out in the open.

The informational collection of the tweet will be utilized to deal with the AI calculations and models. This calculation will perform smoothening the tweet information by disposing of zero qualities. Utilizing Laplace and watchman's

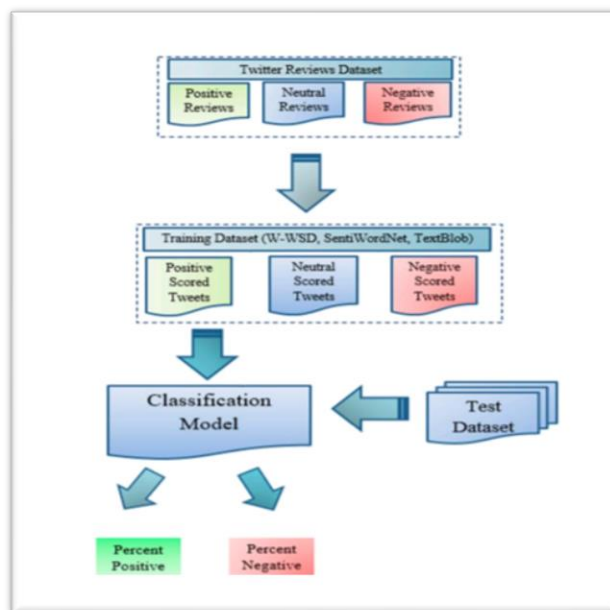
hypothesis, a strategy is created to investigate the tweet information and eliminate repetitive data from the informational index. Gigantic quantities of individuals have been pulled in to online media stage like Twitter, Facebook, Instagram. Individuals express their feelings about society, governmental issues, women, and so forth by means of the instant messages, emojis and work labels out such stages. There are a few strategies for feeling that can be ordered like machine inclining based and vocabulary-based learning.

Women Sexual viciousness happens all through the India. Albeit in many states there has been little exploration directed on the issue, accessible data recommend that in certain states around one of every four women may encounter women sexual savagery by a private accomplice [1-3], and dependent upon 33% of young adult young ladies report their underlying sexual experience as being constrained [4-6]. Around 60percent of women are dangerous while going to office or public spots while voyaging, predominantly in Metropolitan urban communities of India. Women should feel free and safe while going to working spots and public spots. Indeed, even little young ladies additionally confronted these harassments while going to educational costs with some of neighbors because of this they are despondent whole lifetime. So, we attempt to investigate this sort of exercises to diminish Women Sexual brutality.

Online media like twitter, Instagram, Facebook are utilized by different individuals for their suppositions called feeling. Immense information accessible in

online media locales. Women are likewise high utilizing Social destinations for imparting their episodes or insights. Feeling focal perspectives via online media content investigation because of the most crucial information that one craving to separate from the web-based media content is the thing that clients quote and what their suppositions are. These are extricating the center undertakings of estimation investigation. Conclusion examinations are a programmed strategy for removing the self-obstinate data.

## 2. SYSTEM ARCHITECTURE



**Fig -1: SYSTEM ARCHITECTURE**

Each client information like certifications, new tweets, re-tweets and tweet score will be put away in the data set for the administrator to screen and play out the examination. The assumption investigation is applied on the client information to screen and affirm if any tweets are oppressive to women. Administrator plays out this

investigation on every single client tweet to give wellbeing to the women. Wistful examination will be carried out on the tweets of client that are put away in the data set. Administrator would now be able to set up the information to play out the examination. The tweets made by each client of the application will be called as the underlying contribution for the opinion examination and consequently they will be the dataset. Alongside this, text examination diagram can likewise be appeared. Administrator will store the channels in the data set. Channels are the watchwords for which the tweet setting will be looked for to announce as oppressive or not. There can be two kinds of channels – positive catchphrase and negative watchword. Positive catchphrases are those words which are oppressive or affront the women using any and all means. Negative catchphrases are the words which are ordinary and won't mishandle the women.

There can be 'n' number of positive and negative watchwords put away in the data set. When the administrator carries out the wistful investigation, each watchword in the data set will be contrasted and every single word in the tweet of the client. On the off chance that any of the positive catchphrase is found in the tweet, that tweet will be named positive wistful examination and these are harmful to women. In the event that negative watchword is found in the tweet, it will be named the negative wistful investigation which isn't oppressive to women. Subsequently, by this stage there will be two kinds of nostalgic investigation made dependent on the channel in the information base. Under sure nostalgic

investigation, there will be a rundown of the multitude of tweets in the application that are oppressive to women. Likewise, under bad wistful investigation there will be a rundown that is spotless and are not oppressive tweets. Alongside the tweet setting, client subtleties will likewise be given at every one of the examination lists.

### 3. TWITTER ANALYSIS

A huge number of clients utilize online media like tweeter to communicate their feelings, assumptions, sentiments, and suppositions for the world to peruse. These tweets can be effectively extricated and can be exposed to an extremity trial of the expressions utilizing profound figuring out how to decide the rating of lady's wellbeing specifically area. A large number of clients utilize web-based media like tweeter to communicate their feelings, slants, sentiments, and assessments for the world to peruse.

These tweets can be handily removed and can be exposed to an extremity trial of the expressions utilizing profound figuring out how to decide the rating of lady's security specifically territory. We carried out API from tweeter to gather every one of the tweets on tweeter which were labeled under lady harassment or lady wellbeing hashtags or contained words like lady harassments and lady security. In the wake of gathering every one of the tweets, we all things considered isolated them in datasets which were then prepared for polarities and isolated by our calculation.

## 4. IMPLEMENTAION OF SENTIMENTAL ANALYSIS OF TWEETS

In this paper, we will investigate the tweets got by the tweeter API and make a bunch of libraries which will be utilized.

**Algorithm 1** Extract Twitter sentiment

```

1: procedure TWITTER-CONNECTION()
2:   consumer - key = 'xxxxxxxx'
3:   consumer - secret = 'xxxxxxxx'
4:   access - token = 'xxxxxxxx'
5:   access - token - secret = 'xxxxxxxx'
6:   self.auth = OAuthHandler(consumer - key, consumer -
7:   self.auth.set - access - token(access - token, access - to
8:   self.api = tweepy.API(self.auth)
9: end procedure
10:
11: procedure TWEET-CLEANING(t)
12:   tweet = t.remove - Stop - words
13:   Return tweet
14: end procedure
15:
16: procedure TWEET-CLASSIFICATION(t)
17:   t = Tweet - Cleaning(t)
18:   tweet - polarity = t.sentiment.polarity
19:   tweet - polarity
20: end procedure
21:
22: procedure GET-TWEETS(q, count)
23:   fetches - tweets = self.api.search(q = query, count = co
24:   Return fetches - tweets
25: end procedure

```

Allow us to check bit by bit calculation utilized in the examination of tweets on tweeter.

1. Beginning with downloading the wistful word reference.
2. At that point download the twitter testing informational indexes and add them as a contribution to the program.
3. Clean tweets by eliminating the stop words and commotion like redundant letters.
4. Tokenize each word and assign solidarity to the words in the dataset and feed it to the program.

5. For each word, contrast it and positive assessments and negative slants word reference and afterward increase positive tally or negative tally of the general expression.

6. At last, in view of the positive check and negative tally, we can get result rate about estimation to choose the extremity which is classified in Positive, Negative and Neutral.

This calculation will bring the tweets which will be identified with lady wellbeing, and dependent on the words and expressions utilized in the tweets the calculation will make a library of dataset subsequent to erasing the zero words which will eventually check the extremity dependent on the negative or positive opinions of the words and expressions.

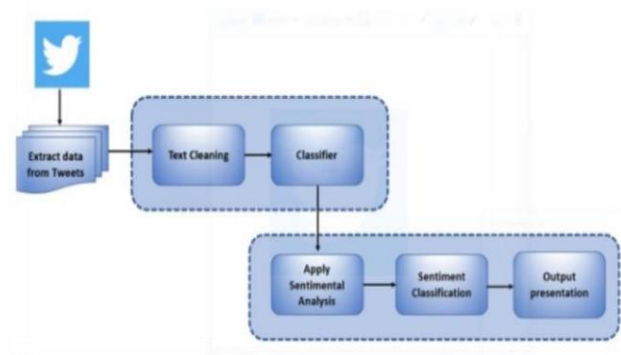
## 5. SENTIMENTAL ANALYSIS

Assumption investigation is the way toward extricating the assessment behind any sentence or articulation. It tends to be called as a grouping procedure which is utilized to acquire the assessment from tweet.

This assessment is valuable in figuring an assumption which can additionally be utilized to accomplish slant characterization. Notions are close to home to the theme and accordingly we need to choose what sort of determinations is defined out of it. Individual playing out the wistful examination needs to discover the class of elements of the tweets utilizing the programming model. The component of the nostalgic class is a significant factor to choose the productivity of the calculation. For example, there can be two class wistful

characterization of tweets – Positive and Negative or there can be three class grouping – Positive, Negative and Neutral. Approaches of nostalgic examination can be comprehensively separated into two kinds – AI based and vocabulary learning based. AI approach incorporates the cycle of extraction of highlights, programming model preparing utilizing dataset of highlights. While dictionary learning based methodology utilizes the jargon and scoring strategy to identify sentiments. In this paper, we use AI approach. Assortment of information, pre-preparing the information, extraction of highlights, picking base highlights, recognition of assessments and characterization of notions utilizing AI approaches or straightforward calculations are the fundamental strides to perform nostalgic examination.

### 5.1 Analysis of Sentimental Data



**Fig -2: Process of Analysis**

The way toward acquiring the estimations of tweet incorporates five stages:

1) Data extraction: First step associated with examination of estimation is the assortment of data from the informal organization site

like twitter. This aides in separating the tweet message however this message additionally incorporates additional information like tweets likes, aversions and remarks.

2) Text Cleaning: Once the information is extricated from the twitter source as the datasets, this data must be passed to the classifier. The classifier cleans the dataset by eliminating excess information like stop words, emojis to ensure that non-text-based substance is distinguished and taken out before the investigation.

3) Sentiment Analysis: After the classifier cleans the dataset, the information is prepared for the wistful examination measure. AI and Lexicon based learning and Hybrid learning are a portion of the methodologies of nostalgic examination. There are additionally some different methodologies like Nero Linguistic Programming and Natural Language Processing. Preparing the dataset and afterward testing that prepared dataset includes in AI approach. Preparing information and Testing information are valuable for the classifier to play out the calculation. Greatest Entropy, Naives Bayes grouping, Bayesian Networks and Network Support Vector Machine are a portion of the calculation which can be utilized to prepare the classifier. Testing information is utilized to recognize the proficiency of the assessment classifier. If there should arise an occurrence of Lexicon based inclining, preparing dataset isn't utilized. This methodology utilizes an underlying word reference in which words related with slants of human are available. The third

methodology, which is the Hybrid learning, consolidates both machine inclining approach and dictionary learning approach to improve the exhibition of classifier.

4) Sentiment Classification: At this progression, the dataset is prepared for the characterization. Every single sentence of the tweet will be analyzed and assessment will be shaped as needs be for subjectivity. Emotional articulation sentences are held and those of target articulation sentences are dismissed. Methods like Unigrams, Negation, Lemmas, etc. are utilized at various degrees of wistful investigation. Slants can be recognized extensively into two gatherings – Positive and Negative. Now of wistful examination, every one of the abstract sentences which will be held are ordered into acceptable, awful or like, aversion or positive and negative.

5) Output Presentation: To produce valuable and significant data out of the crude information, wistful investigation assumes fundamental part. When the calculation is finished, the result of the examination can be pictured by making various kinds of diagrams. Structured presentations, Time arrangement and Pie outlines are a portion of the models which can be utilized to show the yield. To gauge the estimation of the tweets as far as Positive and Negative, Bar diagrams can be utilized. Likewise, to gauge as far as preferences, loathes, normal length of tweet for a specific period, Time arrangement can be utilized. To get the underlying wellspring of the tweet, pie diagrams can be utilized.

## 6. CONCLUSION AND FUTURE WORK

AI calculation has been examined all through the venture. For the twitter information that incorporates a huge number of tweet and messages each day, AI calculation assists with getting sorted out and perform investigation. SPC calculation, direct mathematical are a portion of the calculations which are powerful in dissecting the huge information that give arrangement and convert into significant datasets. Subsequently we can perform AI calculations to accomplish nostalgic examination and carry more security to women by spreading the mindfulness.

For the future improvement, we can reach out to apply these AI calculations on various web-based media stages like Facebook and Instagram additionally since in our undertaking just twitter is thought of. Present philosophy which is proposed can be incorporated with the twitter application interface to arrive at bigger degree and apply wistful investigation on huge number of tweets to give more wellbeing.

## REFERENCES

- [1] Deepak Kumar<sup>1</sup>, Shivani Aggarwal<sup>2</sup> "Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets", (IEEE), 2019.
- [2] Teja KR, Kumar KA, Praveen GS, Harini DN. Analysis of Crimes Against Women in India Using Machine Learning Techniques. In Communication Software and Networks 2021 (pp. 499-510). Springer, Singapore.
- [3] Vikram Chandra<sup>1</sup>, Rampur Srinath<sup>2</sup> "Analysis of Women Safety using Machine Learning on Tweets", (IRJET) 2020.
- [4] P. Anbumani and R. Dhanapal, "Review on Privacy Preservation Methods in Data Mining Based on Fuzzy Based Techniques," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN),
- [5] Reyes-Menendez A, Saura JR, Alvarez-Alons C. Understanding# World Environment Day user opinions in Twitter: A topic-based sentiment analysis approach. International journal of environmental research and public health. 2018 Nov;15(11):2537.
- [6] Gupta B, Negi M, Vishwakarma K, Rawat G & Badhani P (2017). "Research of Twitter sentiment evaluation making use of system mastering algorithms on Python." International Journal of Computer System Applications, one hundred sixty-five(nine) 0975-8887.
- [7] Mamgain N, Mehta E, Mittal A & Bhatt G (2016, March). "Belief evaluation of top schools in India making use of Twitter data." In Computational Strategies, in Details and Interaction Technologies (ICCTICT).
- [8] Bakshi RK, Kaur N, Kaur R, Kaur G. Opinion mining and sentiment analysis. In 2016 3rd international conference on computing for sustainable global development (INDIACom) 2016 Mar 16 (pp. 452-455). IEEE.