ANALYSIS OF WOMEN SAFETY IN INDIAN CITIES USING MACHINE LEARNING ON TWEETS

Dr. V. Anantha Krishna[1], N. Shravani[2], B. Preethi[3], K. Shreya[4]

Abstract

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. 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.

Keywords: Twitter, Civic, Facebook, Instagram, Sentimental Analysis, Data Preprocessing

[1]Professor, Computer Science and Engineering, Sridevi Women's Engineering College B.Tech IVYear Hyderabad, India
[2]Computer Science and Engineering, Sridevi Women's Engineering College, B.Tech IVYear Hyderabad, India
[3]Computer Science and Engineering, Sridevi Women's Engineering College, B.Tech IVYear Hyderabad, India
[4]Computer Science and Engineering, Sridevi Women's Engineering College, B.Tech IVYear, Hyderabad, India

Email: [1]krishnaananthav@gmail.com, [2]nalubandushravani@gmail.com, [3]preethibodha@gmail.com

1.   Introduction

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.

Existing System

Women can also partake their conduct and citations which concentrate on the safety of women in exploits if they have confronted any viciousness or erotic assaults. Strict action can be taken and discipline individuals who kill the womankind.   Twitter application along with the account holders in Existing system include mess marker dispatches the considerably feast on social websites sites people can at liberty prompt transversely the entire globe and it's a stage for women to put forward their view points. Women can partake to prompt their views about how they feel when they have face abuse importunity out for effort or trip in a public conveyance or where we would have fight back against the abuse state of their mind when they're girdled by importunity that was assessed on them.

Disadvantages of Existing System

i.   Applications like twitter and Instagram are used by utmost of the people to come up and put forward their views.
ii.   Few elegances of soppiness that can be distributed like contraption literacy mongrel and wordbook- grounded literacy.
iii.   Due to this there were arithmetical, knowledge- grounded and age wise isolation tactics.

Problem Statement

Facebook and Instagram were the social websites platforms generally used by public to prompt their opinions but these platforms didn't help out well to stop the abusive commentary on women.

Proposed System

'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.

i.   A Depressed commerce G- graph is engendered through some social graph-model, diminishing the distance among the factual and Dejected commerce charts.
ii.   Realities and their relations in social mass media are linked, and a commerce graph is erected.
iii.   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.
iv.   These promptions are delicate to prize their sentiments by current NLP systems easily.

Advantages of Proposed System:

i.   The fact set that was attained over Twitter is trained and provides accurate information regarding the position of safety of women in economic civilization.
ii.   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.
iii.   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.

System Design 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.
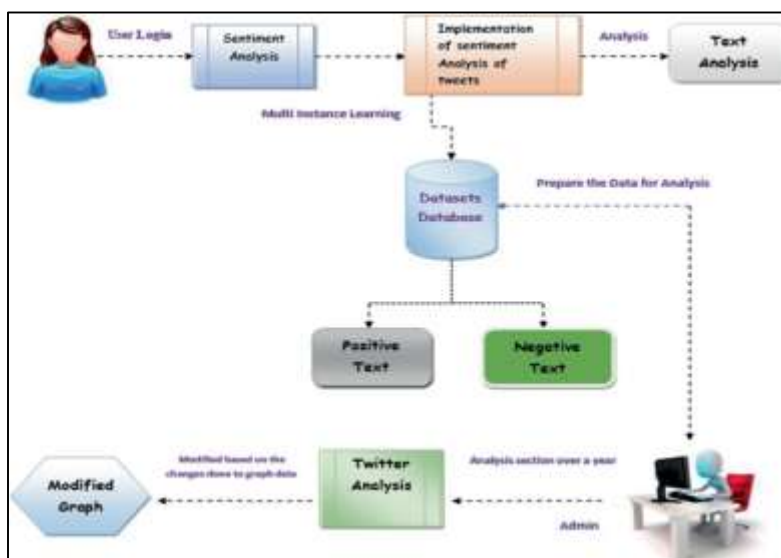
Figure 1. Architectural view of Sentimental Analysis for women safety

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 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.

Figure 2. Sentimental analysis algorithm

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

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

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

3. Tweets Cleaning: This Module is used to see all tweets comprising distinct ciphers that halt abusing words and to remove individual's tweet texts.

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

5. 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 neutral tweets, if not innocent, people will bandy negative tweets.

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.

Figure 3. Download nltk file

Click on "run.bat" train to track design. The window displays the chance of either positive or negative tweet set and neutral tweet set from the trained set of data given along with its opposition score.
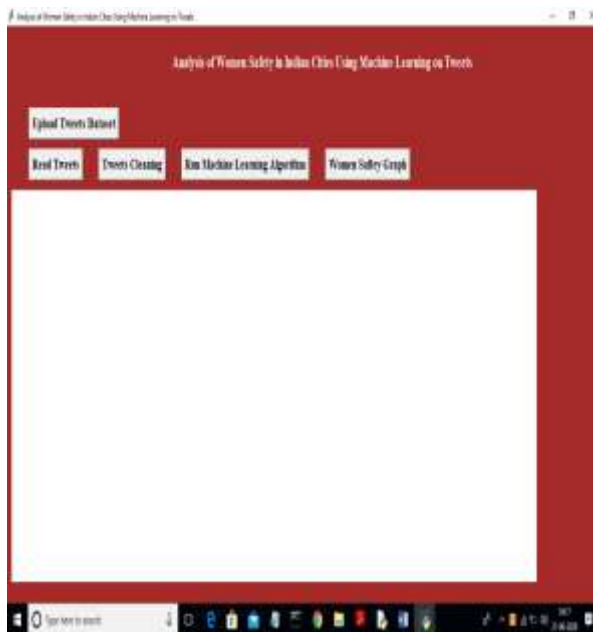
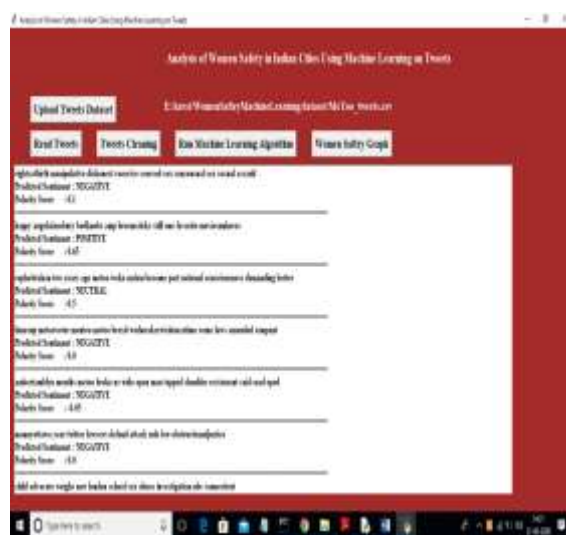**2.   Result**

Figure 4. Output window
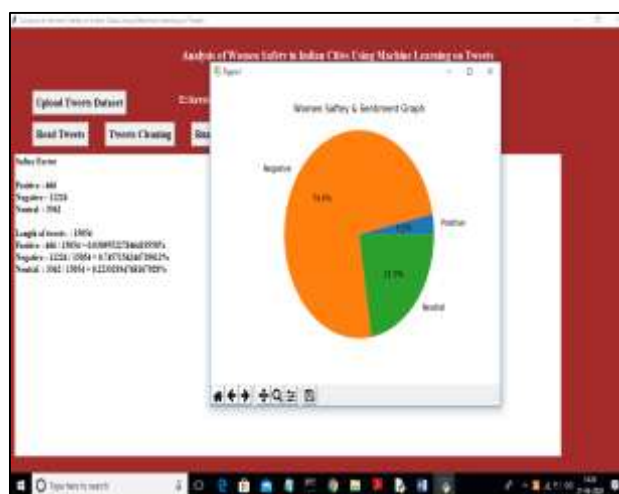

Figure 5. Output window

Figure 6. Women Safety Graph

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.

Future Scope

For the imminent enrichment, one can encompass to smear these algorithms on diverse societal websites like Facebook application and Instagram application. Subsequently, scheme contracts with tweets, present philosophy which is projected can be unified with the application request interface to influence higher possibility and spread on the algorithm used on loads of tweets that delivers added protection.

3.   Conclusion

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.

4.   References

Apoorva Agarwal, Fade Biadsy, and Kathleen R. McKeown. "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams."

Luciano Barbosa and Jenolan Feng. "Robust sentiment detection on twitter from biased and noisy data."

Soo-Min Kim and Eduard Hovey. "Determining the sentiment of opinions."

Dan Klein and Christopher D. Manning. "Accurate unlexicalized parsing."

Eugene Charmian and Mark Johnson. "Coarse-to-fine nbest parsing and Maxent discriminative reranking."