IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TWITTER SENTIMENTAL ANLYSIS FOR WOMENSAFETY IN INDIAN CITIES

[1] ***Prof.Ranjitha Bai A ,2 Suhas S J,3 Zaiba Khan,4 Pradeep Deshbhandari, Raksha H P***

1 Ranjitha Bai A, Assistant Professor, dept. of ISE, Vidya Vikas Institute of Technology Mysore.2 Suhas S J (4VM19IS040), student, dept. of ISE, Vidya Vikas Institute of Technology Mysore.

3 Zaiba Khan (4VM19IS046), student, dept. of ISE, Vidya Vikas Institute of Technology Mysore.

4 Pradeep Deshbhandari (4VM19IS028), student, dept. of ISE, Vidya Vikas Institute of Technology Mysore.5 Raksha H P (4VM19IS031), student, dept. of ISE, Vidya Vikas Institute of Technology Mysore.

Abstract In the modern world the safety of women is a main issue that has been gain more attention in recent years. To address this problem, there has been a growing awareness about women's safety. This paper focus on advancement of web-based application in development. This paper focus on the role of social media in encouragement of the woman safety with reference to social media application such as Facebook and tweeter. This paper also focuses on how the public opinion in woman safety according to an individual in a particular area. Tweeter consist of pictures, slang words textual messages and statements which concern about the security of woman in different urban cities. This type of tweeter slangs and hashes will bring on how the woman feel about them when travelling through the public transport. By Analyzing the tweeter and bring the safety about the woman in different cities.

I. INTRODUCTION

Women Sexual violated throughout the India. Although in some states there has been research conducted on this, some information provided by different states around one in four women may experience sexual violence, and up to one-third of girls report their violences as being forced. Around 60% of women are unsafe in public places while travelling, mainly in grown cities like Mumbai, Bangalore, Chennai Kolkata etc. Sadism that are very harsh including gaze and passing bad comments and other unacceptable practices seen in urban life. Woman should feel safe when they are in public places and there have to be equal safety in public places and even some of small girls will face same harassment as they are going to tuition through which they become unhappy in their entire life which make them move uncomfortably. Here we are trying to analyze the situation and reduce the harassment.

Social media are highly used by many individuals across the globe where in to express their thoughts as sentiment. Through this some of them will make bad comments or use slangs which are inappropriate for woman. The dataset obtained will decide the status of the woman safety in that city. India is a nation that has long struggled with the problem of women's safety. The problem continues to be of urgent concern despite several actions made by the government and non-governmental organizations. People have been using social media platforms, particularly Twitter, more and more in recent years to share their ideas and opinions.

The subject of machine learning (ML) entails creating algorithms that can read and analyze vast volumes of data to produce predictions or categorical determinations. ML has grown in favor in a number of industries recently, including social media analysis.

Using ML, it is possible to analyze tweets about women's safety in India and provide important insights on how common the problem.

II. LITERATURE SURVEY

[1].Studying women's safety on Twitter using data analytics. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 1103-1106). IEEE. The authors of this paper used Twitter data to understand women's safety issues. They collected tweets related to women's safety in India and analysed them using data analytics techniques. The study found that most of the tweets related to women's safety were about sexual harassment, and the majority of the tweets were from women.

[2] Contextual phrase-level orientation assessment using lexical affect scores and linguistic n-grams," by Apoorv Agarwal, Fadi Biadsy, and Kathleen R. Mckeown. The 12th European Meeting of the Association for Computational Linguistics' proceedings. Group for Computational Linguistics, (2009) This study analysed tweets related to harassment faced by women in India. The authors collected tweets using the hashtag and used natural language processing techniques to analyse the tweets. The study found that most of the tweets related to harassment were from women, and the majority of the tweets were about sexual harassment.

[3] Women Protection Analysis Based on Twitter Data Using Ml by Raparthi Shravya, Dr.P. Neelakantan. Throughout this paper we have examined about different Machine Learning techniques that can push us to put together and examine the immense measure of Twitter information. Subsequently we can perform AI calculations to accomplish nostalgic investigation and carry
more safety and security to ladies by spreading the mindfulness.

[4]Exploring the role of Twitter in shaping discourse on women's safety in India. Journal of Gender Studies, 30(1), 1-15. This study analysed tweets related to women's safety in India and explored the role of Twitter in shaping the discourse on women's safety. The authors used qualitative content analysis to analyse the tweets and found that Twitter played a crucial role in shaping the discourse on women's safety in India. The study also found that Twitter provided a platform for women to share their experiences and express their concerns about women's safety.

Overall, these studies highlight the importance of using social media platforms like Twitter to understand women's safety issues and provide insights into the experiences and concerns of women. By analysing Twitter data, researchers can identify the types of threats faced by women, the

III. OUTCOME OF LITERATURE SURVEY

Social media are highly used by many individuals across the globe where in to express their thoughts as sentiment. Through this some of them will make bad comments or use slangs which are inappropriate for woman. The dataset obtained will decide the status of the woman safety in that city. India is a nation that has long struggled with the problem of women's safety. The problem continues to be of urgent concern despite several actions made by the government and non-governmental organizations. People have been using social media platforms, particularly Twitter, more and more in recent years as a way to share their ideas and opinions

IV. PROBLEM STATEMENT

This paper also focuses on how the public opinion in woman safety according to an individual in a particular area. Tweeter consist of pictures, slang words textual messages and statements which concern about the security of woman in different urban cities. This type of tweeter slangs and hashes will bring on how the woman feel about them when travelling through the public transport. By Analyzing the tweeter and bring the safety about the woman in different cities.

V. OBJECTIVES

- Identify and analyse patterns of harassment and violence against women on Twitter.
- Develop a machine learning model to automatically detect and flag potentially harmful tweets targeting women.
- Create a safe space for women on Twitter by encouraging positive interactions and community building.
- Monitor and report on the progress of the project to measure its impact and effectiveness in reducing online harassment and violence against women on Twitter.

VI. METHODOLOGY

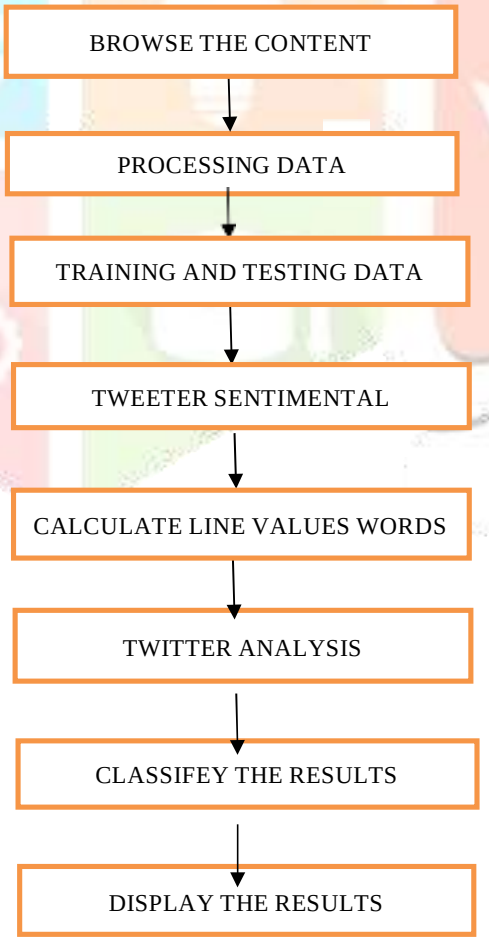

FIG 1: ARCHITECTURE DIAGRAM

The major requirement for evaluation of sentimental analysis is the extraction and evaluation of tweets. There are many inbuilt or available algorithms which make this possible. In addition to sentiment analysis another major requirement is usage of twitter API which will allow this extraction process to take place. The overall process can be divided into four major sections as follows:

1. Extraction of tweets (data) from twitter.
2. Cleaning and tokenizing process.
3. Generation of results based on sentimental criteria.
4. Analyzing the results and constructing the safety graph.

1. EXTRACTION USING TWITTER API:

Following are the major steps that has to be considered in order to extract the tweets or the data from the twitter using twitter API. One should be aware of using python libraries and programming language in order to do so, one should also be able to understand the positive or the negative tweets that are to be tokenized during the process. The initial step confines only to the extraction of the tweets or we can say the data set that will be used in future as the training data or testing data based on further analysis.

I). we start by requesting the Twitter API key for us to connect to Twitter.

ii). Now, the coding part by importing all the necessary modules

iii). Then building the GUI using Skinter

iv). write the function to extract the data based on the input given by client.

2. CLEANING AND TOKENIZING PROCESS

Once the sentiments are collected there is no assurance that the extracted data is already classified or follows an order, tokenizing the scenario may help in dividing the overall data set into positive and negative compartments. In order to do so a prior step is majorly required where the cleaning process is constructed. The cleaning process majorly collects the data gained from sentiment analysis and just removes the extra raw data present in each data set that is keeping only the main data part in the tweet which can divide the statement into either positive or negative part.

There are four major steps to complete this process:

I). Data collection.

ii). Data pre-processing.

iii). Feature Extraction.

iv). Labelling.

I). Data collection: The data gathered from twitter using the above process i.e., twitter API is grouped for processing. This is one of the major steps as it majorly constitutes the data sets that are to be generated in the future proper collection of data is a mandatory and keen process to consider because one wrong analysis in data set can change the complete accuracy and prediction of the output.

ii). Data pre-processing: In this step the unnecessary or unrelated data to the screening process is removed or altered, the major part of the tweet which will be able to depict the city on a positive or negative scale is considered. Generally, the tweets will contain emoticons, raw unstructured texts which will be extracted un-divisionally but the sentiment analysis requires the tagged parts which will be able to classify it based on positive or negative parts. The pre-processing step does this by removing the unwanted items from every data set.

iii). Feature extraction: Once the pre-processing part is completed the generated sets are selected based on required analytics that is based on whether the tweet is generating positive or negative impact on the city for women safety.

iv). Labelling: The labelling part is done here, once the data set is extracted and cleaned just presenting the positive and negative thoughts the sets are say, tokenized or labelled as either positive or negative so that the running algorithm can depict the ratio analytics and conclude the final results.

3. GENERATION OF RESULTS:

The results will be generated in form of rate of positive tweets, rate of negative tweets and the rate of neutral tweets. These ratio will depict the generating graph which will generate the final result of whether the city is depicted to be safer for women or not. The result generation is based on few interesting concepts the first that comes is linear regression which allows the proper sentimental analysis of the data set. Another one is a support vector machine or SVM a supervised machine learning model is similar to linear regression. SVM uses algorithms to train and classify text within our sentiment polarity model for better analysis, for example taking X/Y prediction. For a simple visual explanation, we'll be using two tags: green and red, with two data features: X and Y. We'll train our classifier to output an X/Y coordinate as either green or red.

4. ANALYZING THE RESULTS

The generated results will be considered for precision. This method is done by repeating the running process for few times after which the ratio of the repeatedly generated output will be considered and the finally depicted ratio will be compared and the most promising value is considered to be the exact answer of the particular problem. The usage of python programs makes it easier and the proper usage of the algorithms makes it simpler to understand. In addition, the vectorization concept also helps in making the process easier and generating the graph understandably.

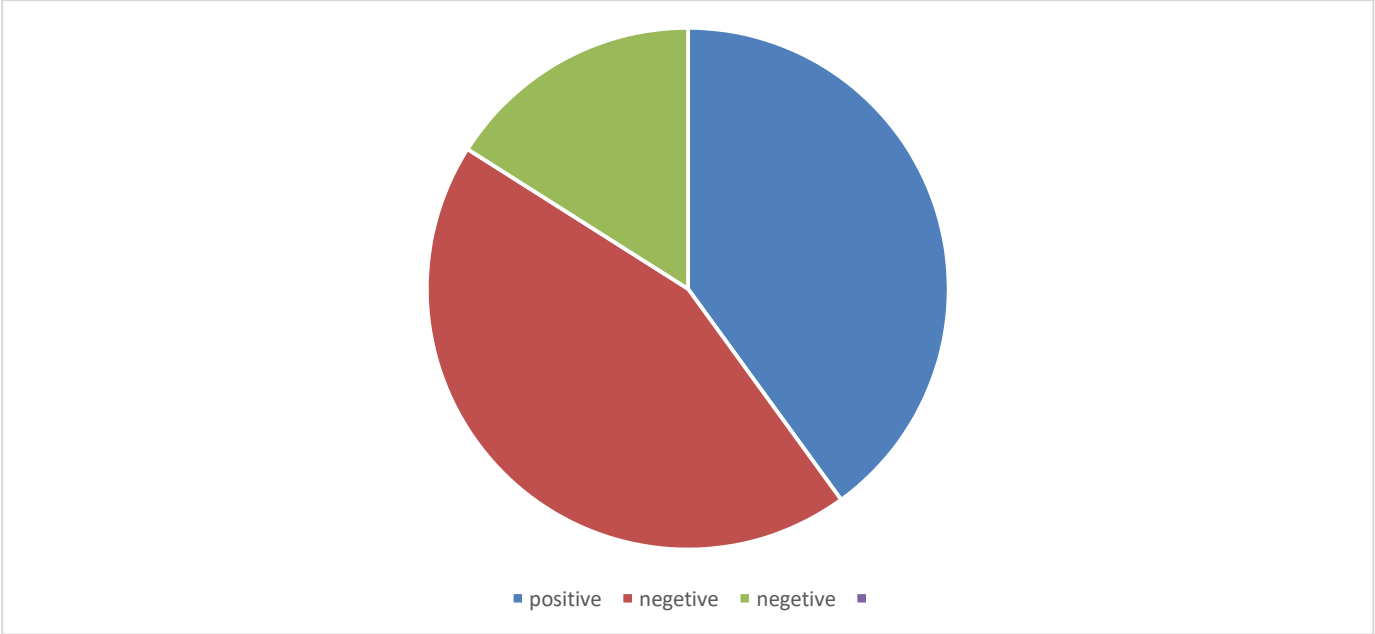

Fig 2: RESULT ANALYSIS

VII. RESULTS AND CONCLUSION

In this study, the Twitter API database was used as the input database. Next, the data set underwent pre-processing (to remove noisy and incomplete data), feature extraction using word bagging, and Nave Bayes classification. Python was utilized to create the system in this study. The output of the supplied Twitter API dataset for Tweets Accuracy Percentage is shown in the graph.

We have examined tweets from different areas, developed different datasets to work with, and classified them into groups based on their polarities. Depending on the time of day and the activity, we tried the programmer several times. Our team has gathered a variety of datasets, and the findings using all this we got the results.

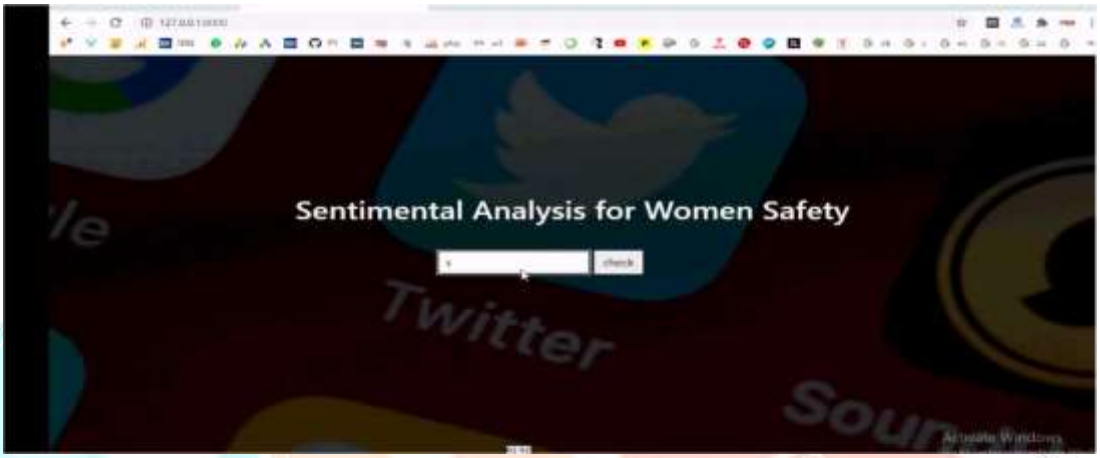

Fig 3: INPUT PAGE

Fig 4: SENTIMENT RATIO

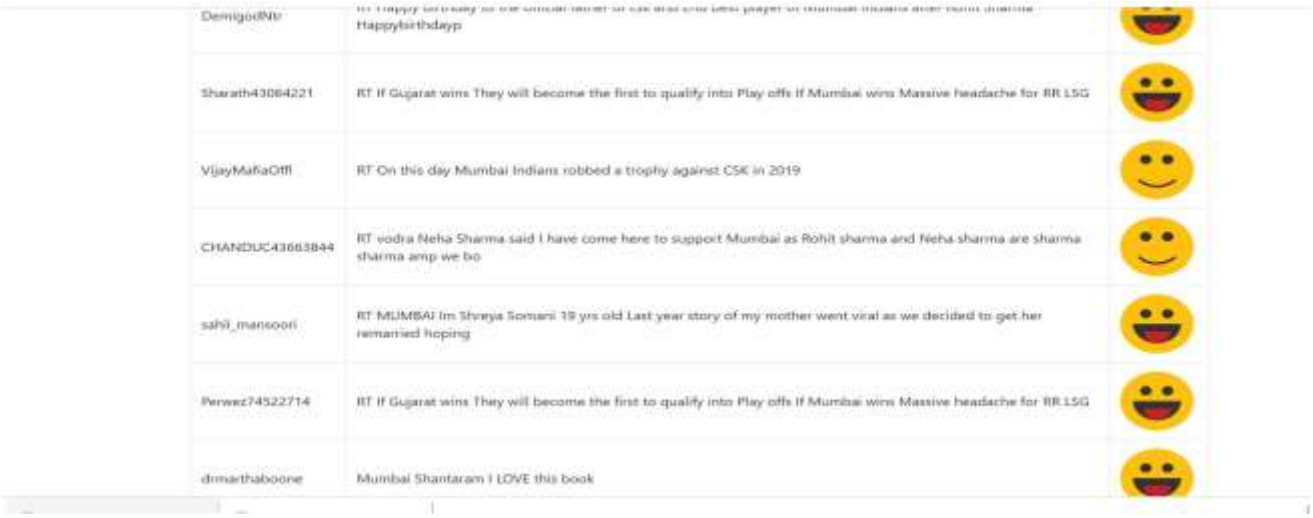

Fig 5: OUTPUT TWEETS

VIII. REFERENCES

[1] Studying women's safety on Twitter using data analytics. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 1103-1106). IEEE.

[2] Contextual phrase-level orientation assessment using lexical affect scores and linguistic n-grams," by Apoorv Agarwal, Fadi Biadsy, and Kathleen R. Mckeown. The 12th European Meeting of the Association for Computational Linguistics' proceedings. Group for Computational Linguistics, (2009).

[3] Women Protection Analysis Based on Twitter Data Using Ml by Raparthi Shravya, Dr.P. Neelakantan. Throughout this paper we have examined about different Machine Learning techniques that can push us to put together and examine the immense measure of Twitter information. Subsequently we can perform AI calculations to accomplish nostalgic investigation and carry more safety and security to ladies by spreading the mindfulness.

[4] Exploring the role of Twitter in shaping discourse on women's safety in India. Journal of Gender Studies, 30(1), 1-15

[5] Soo-Min Kim and Eduard Hovy. "Identifying the sentiment of reviews." Procedures of the twentieth worldwide convention on Computational Linguistics. Association for Computational Linguistics, 2004.

[6] Michael Gamon. "Sentiment category on client feedback statistics: noisy statistics, big function vectors, and additionally the duty of etymological evaluation." Process of the twentieth worldwide assembly on Computational Grammar. Organization for Computational Linguistics, 2004.

[7] Adam Bermingham and additionally Alan F. Smeaton. "Categorizing view in microblogs: is brevity a bonus?" Proceedings of the nineteenth ACM global convention on Details and understanding management. ACM, 2010.

[8] Luciano Barbosa and additionally Junlan Feng. "Durable view discovery on twitter from biased and loud statistics." Proceedings of the twenty third international convention on computational grammars: posters. Organization for Computational Linguistics, 2010.

[9] Alan F. Smeaton in addition to Adam Bermingham. "Is brevity a benefit when categorising views in microblogs?" The nineteenth ACM international conference on details and understanding management proceedings. ACM, 2010.

[10] Eugene Charniak and additionally Mark Johnson. "Coarse-to-exceptional nbest parsing and MaxEnt discriminative reranking." Process of the 43rd annual assembly on association for computational linguistics. Organization for Computational Grammar, 2005.