International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 45-49 www.ijaem.net ISSN: 2395-5252

Analysing tweets using Machine Learning for Woman Safety

Kalyani Dilip kakade, Shivani Sunil kakade, Aditi Singh, Shikha Sharma
[1,2,3,4]*Student, Ballarpur Institute of Technology, Chandrapur, India*
Corresponding Author: Kalyani Dilip kakade, Shivani Sunil kakade, Aditi Singh, Shikha Sharma

Date of Submission: 01-06-2020 Date of Acceptance: 16-06-2020

ABSTRACT: Twitter has been used for many analysis based on the sphere of research. In this project we implemented a system which can rate the polarity of woman safety based on the tweets and sentiments shared on the twitter in different regions of the country. We have seen woman safety is a priority in many places of World and in India, so as a result we have come up with a tweet analysis project which can analyze the tweets and based on the sentiments shared by the tweeter users, our system can divide the results in positive aspect. Tweeter users share various kind of information, tweets, images, videos, and other media related to woman safety along with the hashtags that goes with it. Our implementation has successfully distinguished between positive and negative users with respect to regions of the users.

KEYWORDS: Women's safety, Sexual Harassment, Sentiments, Safety, Safety Analysis**.**

I. INTRODUCTION

Women are continuously harassed in our society every day. Each and every city has some parts or localities where women harassment is a major issue. The survey of metropolitan cities has uncovered that 60% of women are harassed and do not feel safe while going out of their houses. These harassments range from passing comments to body shamming which is a matter of concern for all of our society. Women while travelling via public transport feel unsafe according to the recent analysis. There are many cases in a society where women are continuously harassed in their neighbourhoods, shopping malls, and on their way to their work.

These issues of harassment lead to the discouragement of women class to work in a safe environment. Building a safe and harassment-free work environment for women can encourage them to work and prosper. One incident of harassment for a woman or girl can carve a lifetime bad memory and leave a scar for that woman or girl. Our society needs to approach woman safety with a perspective which will empower them to live a carefree life without having to concern with their safety and harassment.

In this project, our team is implemented a system which analysed the tweets by the people who stands up and rose their voice against woman safety, created a dataset of it which we obtained from Tweeter that was processed through Machine Learning to remove the zero words and redundant data using Laplace and porter's theory and developed a method of analysation of dataset to obtain a clear view of woman safety in the society.

II. LITERATURE REVIEW

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.

III. TWITTER ANALYSIS

Thousands of users use social media like tweeter to express their emotions, sentiments, feelings, and opinions for the world to read. These tweets can be easily extracted and can be subjected to a polarity test of the phrases using deep learning to determine the rating of woman's safety in particular locality.

Thousands of users use social media like tweeter to express their emotions, sentiments, feelings, and opinions for the world to read. These tweets can be easily extracted and can be subjected to a polarity

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 45-49 www.ijaem.net ISSN: 2395-5252

test of the phrases using deep learning to determine the rating of woman's safety in particular locality.

We implemented API from tweeter in order to collect all the tweets on tweeter which were tagged under woman harassment or woman safety hashtags or contained words like woman harassments and woman safety. After collecting all the tweets, we collectively divided them in datasets which were then processed for polarities and separated by our algorithm.

IV. IMPLEMENTATION OF SENTIMENTAL ANALYSIS OF TWEETS

In this paper, we will analyze the tweets picked up by the tweeter API and create a set of libraries which will be used.

Let us check step by step algorithm used in the analysis of tweets on tweeter.

1. Starting with downloading the sentimental dictionary.
2. Then download the twitter testing data sets and add them as an input to the program.
3. Clean tweets by removing the stop words and noise like repetitive letters.
4. Tokenize each word and allot strength to the words in the dataset and feed it to the program.
5. For each word, compare it with positive sentiments and negative sentiments word dictionary and then increment positive count or negative count of the overall phrase.
6. Finally, based on the positive count & negative count, we can get result percentage about sentiment to decide the polarity which is categorized in Positive, Negative and Neutral.

Algorithm 1 Extract Twitter sentiment

```
procedure TWITTER-CONNECTION()
    consumer − key =' xxxxxxxx'
    consumer − secret =' xxxxxxx'
    access − token =' xxxxxxx'
    access − token − secret =' xxxxxxxxx'
    self.auth = OAuthHandler(consumer − key, consumer −
    self.auth.set − access − token(access − token, access − to
    self.api = tweepy.API(self.auth)
end procedure

procedure TWEET-CLEANING(t)
    tweet = t.remove − Stop − words
    Return tweet
end procedure

procedure TWEET-CLASSIFICATION(t)
    t = Tweet − Cleaning(t)
    tweet − polarity = t.sentiment.polarity
    tweet − polarity
end procedure

procedure GET-TWEETS(q, count)
    fetched − tweets = self.api.search(q = query, count = co
    Return fetched − tweets
end procedure
```

Fig. 1. Sample code on a high level programming

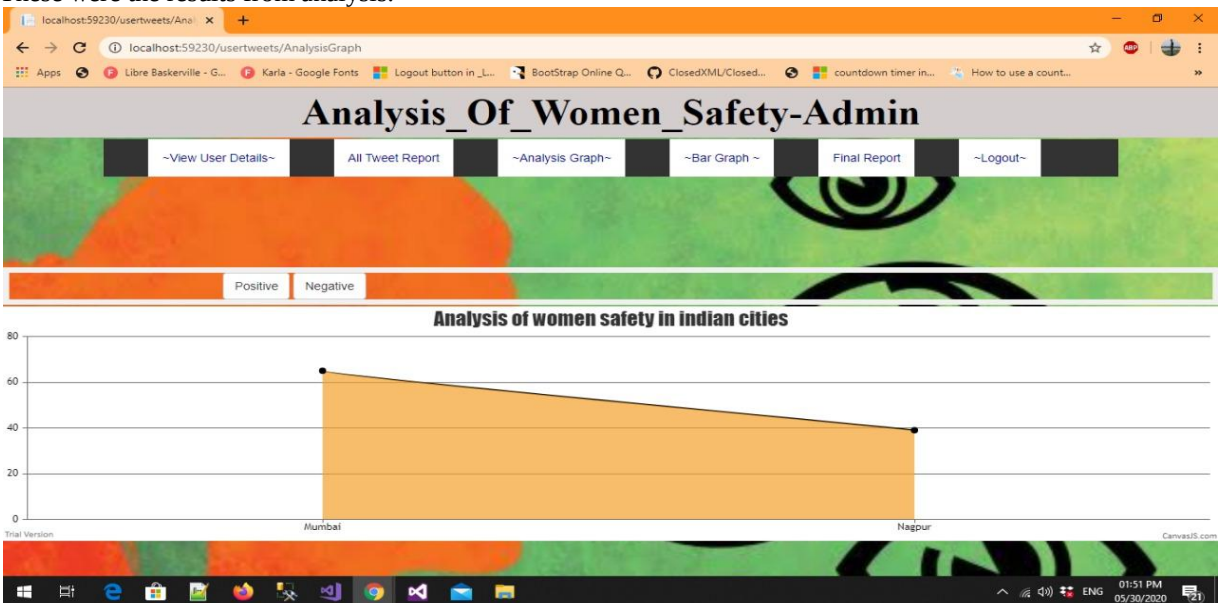
This algorithm will fetch the tweets which will be related to woman safety, and based on the words and phrases used in the tweets the algorithm will create a library of dataset after deleting the zero words which will ultimately check the polarity based on the negative or positive sentiments of the words and phrases.

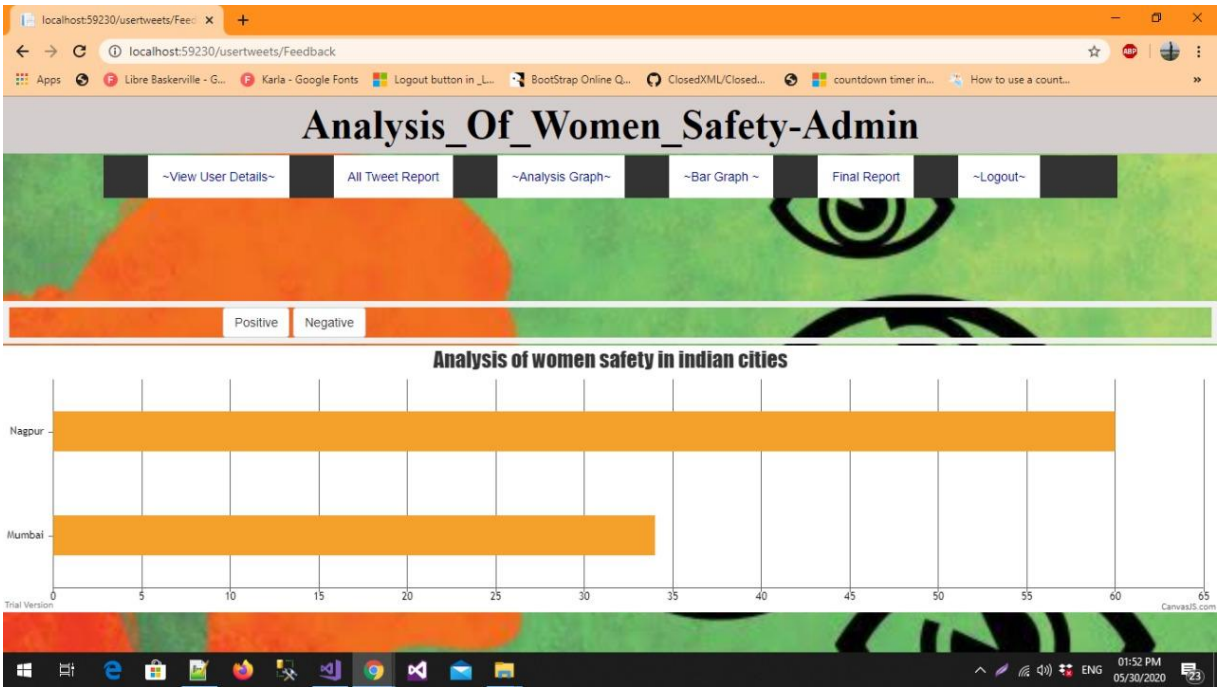
International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 45-49 www.ijaem.net ISSN: 2395-5252

V. RESULTS AND EVALUATION

We have implemented the system in various localities, and these were the results from the implemented system. We have analyzed tweets from various regions and created various datasets to work on, and then divided them according to their polarities.

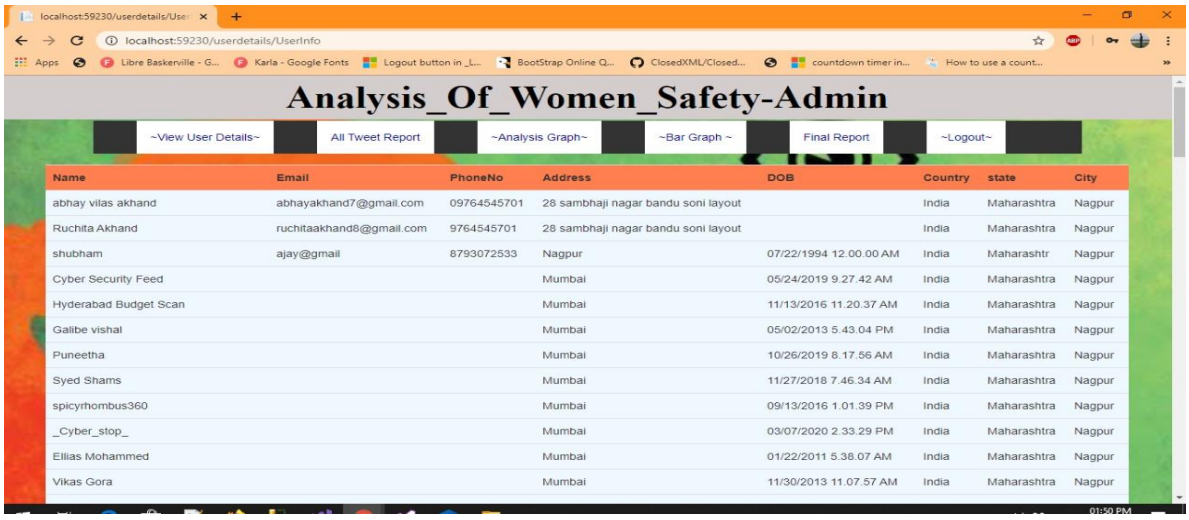
These were the results from analysis.

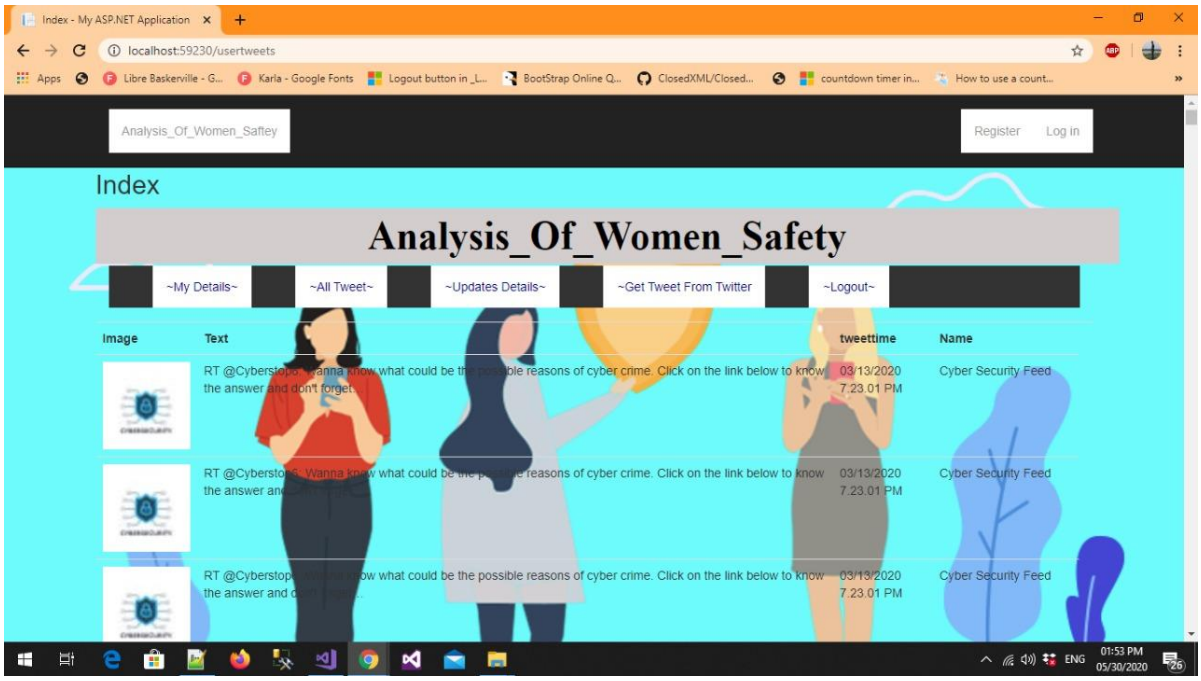




These are some of the tweets and datasets that were created with the help of tweeter API and algorithm.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 1, pp: 45-49 www.ijaem.net ISSN: 2395-5252





We have tested the program multiple times depending on the activity time throughout the day. Our team has collected various datasets and the results were different each time with bit of variance. If the neutral tweets are significantly high, means that people have a lower interest in the topic and are not willing to haves a positive/negative side on it. This is also important to mention that depends on the data of the experiment we may get different results as people's opinion may change depending on the circumstances for example rape news it becomes the most trending news of the year in 2017.

VI. CONCLUSION

Throughout the paper 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.

Kalyani Dilip kakade, et. al. "Analysing tweets using Machine Learning for Woman Safety." *International Journal of Advances in Engineering and Management (IJAEM)*, 2(1), 2020, pp. 45-49.

International Journal of Advances in Engineering and Management
ISSN: 2395-5252

IJAEM

Volume: 02 Issue: 01 DOI: 10.35629/5252

www.ijaem.net
Email id: ijaem.paper@gmail.com