International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 06 | June -2020 ISSN: 2582-3930

Survey on Women Safety UsingMachine Learning Techniques

Nireeksha R J[1], Dr. Ananda Babu J[2], Varshini C[3], Vandana H S[4]
BE student, Department of Information Science and Engineering, Malnad College of Engineering, Hassan, India[1]
Associate Professor, Department of Information Science and Engineering, Malnad College of Engineering, Hassan, India[2]
BE student, Department of Information Science and Engineering, Malnad College of Engineering, Hassan, India[3]
BE student, Department of Information Science and Engineering, Malnad College of Engineering, Hassan, India[4]

Abstract: Including babies, all ages of women are facing a lot of harassments in the society. Stalking, eve teasing, ragging, unethical physical harassments are the common problems experienced by most of the women. In such situations, they feel helpless and tend to seek help. Many technologies have been introduced intending to help and protect women. Past few years, Social media has gained a lot of popularity. Along with the positivity and advantages, it also has drawbacks and one among them is the violence against women. In this paper, we have surveyed the existing systems and mechanisms that intends for women safety on social media in India. Based on the views and opinions expressed on the social networking sites by people, comments and tweets are extracted from Twitter and other social networking sites which are then analyzed using Recursive Neural Tensor Network (RNTN) in R-studio and Python program. Using hash tags (#) that are used in the comments by user tweets are downloaded. Based on the results of the analysis, safe cities for women and sentiments of people about particular concept is obtained.

Keywords: *Sexual Harassment, Self Defence, Hash tag, Tweets, Python, Sentimental Analysis, RNTN, R-studio, Social networking sites .*

INTRODUCTION

Women safety is one of the major problems in the entire world. As each country is growing, it is lacking in providing safety for women. There are violent actions and harassments being faced by women in the society not just during night but also in the broad day light. In today's world, women believe themselves as equal to men and get exposure to various fields. Women work in various fields at different time, like men. But face various problems like stalking, staring, receiving comments from unknown people.

After globalization, they have got several opportunities equally to men. Nowadays we can see women working in every field. They drive cabs, run metro trains, works as CEO for the leading companies and also works in army for the nation.

Even though many technologies have been introduced for safety measures of women and rules and laws have been implemented by the government concerning women safety, many feel unsafe in their home, working places and cities. Though rights and opportunities are provided to women like men, women feel unsafe in places like shopping malls, bus-railway stations because of many eyes, body shaming and sexual harassment. [1] According to survey that have been conducted in various cities across India, women have reported facing similar type of harassments and passing off comments in public places. A study was shown that 60% of women feel unsafe in the most popular metropolitan cities like Delhi, Mumbai and Pune.

Since social media is growing widely everyone is free to share their experiences and express their opinions. Women who face threats and harassments can easily share the incidents and bring awareness among others about the person or the place. On the basis of those opinions, many systems are built to analyze the threats against women through various social networking sites.

Messages, comments, text, likes, shares and other form of opinions are retrieved from the multiple social networking sites. The collected data then is used as input for different algorithms and the result is classified into positive, negative or neutral according to human beliefs about particular incident. The analysis result can be represented using graph or chart which provides the most unsafe and safe cities for women. The mechanisms used for analysis may be of

1. Opinions, likes, shares and comments are obtained from Facebook and analysis is done using RNTN algorithm.

2. Tweets are collected from Twitter and using sentimental dictionary and high level programming language, analysis is performed.

Existing System

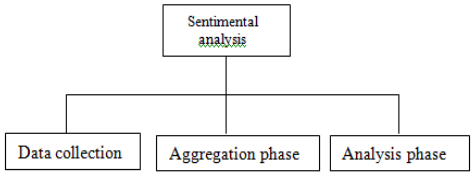
In this paper, we have surveyed the existing mechanisms for analysing women safety on social media and predicting safety measures and safe cities for women in India.

1. Analysis using RNTN

Recursive Neural Tensor Network (RNTN). It takes any length of input phrases. Each phrase is represented through word vectors and parse tree and then vectors are computed for

higher nodes in the tree using tensor-based composition. It also uses recursive neural networks (RNN) for the new datasets to obtain improved performance for predicting the fine-grained sentiment.

1.1 Phases of sentimental analysis:



1.1.1 Data collection:

The data that is suitable for the system to perform sentimental analysis is retrieved from various social media applications like Facebook and Twitter. The data that matches with the concept of harassment is collected and stored using open authentication R-studio.

- R-studio – An integrated development environment (IDE) for R. It includes console, syntax-highlighting editor that supports direct code execution as well as tools for plotting, history, debugging and workspace management.

1.1.2 *Aggregation phase:*

The data set is stored in tabular form contains different views of people that are expressed on the social media sites in the form of feelings and that data set is aggregated based on the sentimental rate. This results in many other tables of sentiments which are used for further analysis.

1.1.3 *Analysis phase:*

Using packages like syuzhet, the data that is stored in R-studio is analyzed and different graphs are plotted based on data set. The syuzhet package helps in obtaining the sentimental feelings like anger, disgust, sadness, happy, positive, negative from users opinions. These feelings are then tabulated as shown in Table 1 and later retrieved and coded for analysis.

The comments by users on social networking sites are given as input and analysis is done based on the given input. The output is displayed as graphs based on the sentiments as shown in Fig 1.1 and Fig 1.2.

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	2	0	0	2	1	0	0	1	0	1
2	0	1	0	0	1	0	1	1	0	1
3	0	2	0	0	2	0	1	3	0	5
4	3	3	2	5	1	3	1	3	5	2
5	0	0	0	1	0	0	0	1	0	0
6	1	0	2	2	0	0	0	0	3	1
7	0	0	0	0	2	0	0	2	0	2
8	1	0	0	1	0	0	0	0	1	0
9	0	0	0	1	1	0	0	1	2	2
10	0	1	0	0	1	0	1	1	0	1
11	0	1	0	1	2	0	0	2	2	4
12	0	0	0	0	0	0	0	0	0	0
13	0	0	0	1	0	0	0	0	1	1
14	1	0	0	0	0	0	0	1	0	1
15	0	0	0	0	0	0	0	0	0	1

Table 1

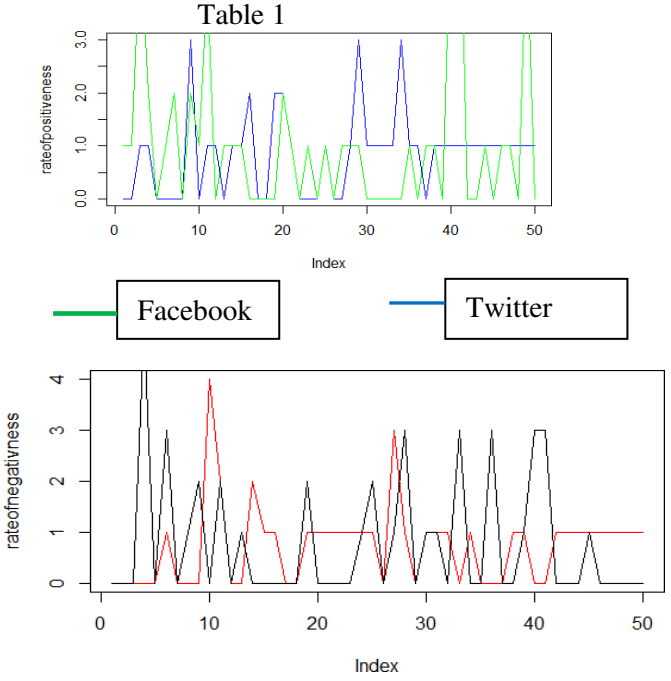


Fig 1.1 Positive Graphical Representation

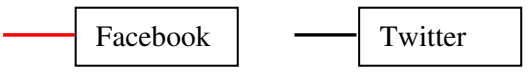


Fig 1.2 Negative Graphical Representation

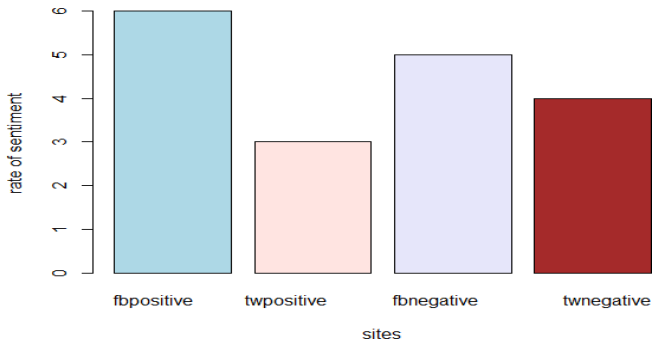


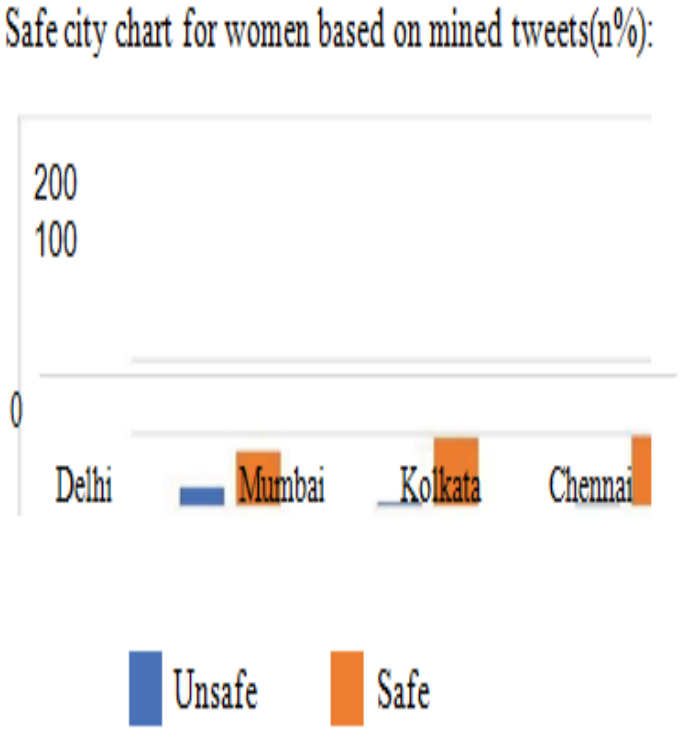
Fig 1.3 Graphical representations of integrated sentimental data

2. *Analysis of tweets using Python*

Tweets concerned to violence and sexual harassment against women are collected from Twitter API available in Twitter. It is easy to obtain tweets from twitter using hashtags about particular incident or opinion. The sentimental analysis on tweets can be performed as follows:

- Sentimental dictionary is downloaded and packages like tweepy and textblob are used.
  - Tweepy – A python library for accessing the Twitter API. It is great for simple automation and creating twitter bots.
  - Textblob – A python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation and more.
- A high level program is designed using Python to clean, classify and retrieve live tweets.
- The downloaded data sets of tweets are given as input to the program where they are cleaned by removing repetitive letters and stop words.
- Each word of every tweet is compared with the positive and negative sentiments word dictionary and the count of positive and negative sentiments are recorded.
- Based on the count of positive and negative sentiments, the percentage of sentiments is obtained and polarity is decided which is characterized in Positive, Neutral and Negative.

Based on the tweets extracted, the output of the program differs with each execution by small variance.



Metro City	Cases tweeted about	Safety factor (tweet vs population)
Delhi	173,947	75%
Mumbai	42,940	93.6%
Kolkata	23,990	96.5
Chennai	13,442	98%

Table 2 Safety factor chart

Table 3 Comparison between the mechansims surveyed in this paper

Functions	RNTN	Python
Data collection	Data suitable for system is retrieved	Using hash tag data is retrieved.
Packages and libraries	Syuzhet - pckg	Corpora - pckg Tweepy - lib textblob - lib
Software	R-Studio	Python IDE
Social networking sites	Facebook and Twitter	Twitter
Data analysis	Opinions are converted into sentiments using package and that data is used as input for program to perform analysis	Collected data is given as the input to the program using sentimental dictionary tweets are compared and classified as +ve, -ve or neutral sentiments
Result	Output is displayed as graph for number of emotions and based on the graph sentiments of people can be obtained	+ve, -ve and neutral sentiments in percentage is obtained. Based on the percentage cities safe or unsafe for women is determined
Advantages	Data can be collected from more than one networking site and used as input for the same program	Analysis can be performed for the live data.
Drawbacks	Analysis can be done only for the stored data.	Different programs should be designed for different networking sites.

CONCLUSION

From the above survey, we analyzed that sentimental analysis using RNTN can be done only to the stored data set. Live data sets from the social networking sites cannot be analyzed using this approach. Analysis of tweets using Python provides facility to extract live data set from the social media applications and perform analysis to obtain the sentiments of people based on which graph can be plotted to display the unsafe cities for women. But this approach needs different programs depending on the networking site. So a new system needs to be developed which would extract live data sets from all the social networking sites and analyze it to provide view on women safety in our society.

ACKNOWLEDMENT

None

REFERNCES

[1] S. Ramamoorthy, R. Poorvadevi, "Safety Measures Against Women Violence in India using Sentimental Analysis". International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-6S3, April-2019.

[2] Deepak Kumar, Shivani Aggarwal, " Analysis of Women Safety Using Machine Learning on Tweets". 978-1-5386-9346-9/19/$31.00 ©2019 IEEE.

[3] B.Sindhu Bala, M.Swetha ,M.Tamilarasi and D.Vinodha "Survey on women safety using IOT", E-ISSN: 2349-7084, Volume-5, Issue-2 ,2018 Regular Edition

[4] Abhijit Paradkar and Deepak Sharma, "All in one Intelligent Safety System for Women Security". International journal of computer applications(0975-8887) Vol 130-No 11, November 2015.

[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.

[6] Sahayak V, Shete V & Pathan A. "Sentiment analysis on twitter data"(2005). International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.

[7] Gupta B, Negi M, Vishwakarma K, Rawat G & Badhani P. "Study of Twitter sentiment analysis using machine learning algorithms on Python" (2017). International Journal of Computer Applications, *165*(9), 0975-8887.

[8] Sexual harassment of women at work place in INDIA, journal of business management and social sciences research, volume3, No.7, July 2014