

Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets

S. Sunil Kumar, B. Sruthi, B. Pallavi, T. Rohini

1. Assistant Professor, Department Of Information Technology,
Malla Reddy Engineering College For Women(UGC Autonomous),Hyderabad, India
Email :ssnlkmr@gmail.com

2. Department Of Information Technology,
Malla Reddy Engineering College For Women(UGC Autonomous),Hyderabad, India
Email :bagamsruthi02@gmail.com

3. Department Of Information Technology,
Malla Reddy Engineering College For Women(UGC Autonomous),Hyderabad, India
Email :pallavibagam03@gmail.com

4. Department Of Information Technology,
Malla Reddy Engineering College For Women(UGC Autonomous),Hyderabad, India
Email :thantravedhirohini@gmail.com

Abstract:

Women and girls have been experiencing a lot of violence and harassment in public places in various cities starting from stalking and leading to sexual harassment or sexual assault. This research paper basically focuses on the role of social media in promoting the safety of women in the safety of women is a concern of increasing urgency in India and other countries. The primary issue in the handling of these cases by the police lies in constraints preventing them from responding quickly to calls of distress

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

I. INTRODUCTION

Women's safety involves strategies, practices and policies which aim to reduce gender-based violence (or violence against women), including women's fear of crime. Women's safety involves safe spaces. Space is not neutral. Space which causes fear restricts movement and thus the community's use of the space.

Women have the right to be free from violence, harassment and discrimination and removing the barriers of an unsafe environment can help women fulfil their potential as individuals and as contributors to work, communities and economies. Women's safety must be the prime concern of all because she is not someone who could be trampled over. A woman

takes many roles in our lives, and yet she suffers the most.

Today, the safety of women in India is widely discussed everywhere. Now it has become a serious problem. The crime rate is skyrocketing. Women are not safe either at home or outside. Female travelers from other countries also find themselves in a precarious position when traveling to India. But

There are laws, but there must be adequate security measures that must be strictly followed to protect against violence against women. Increase, we must implement effective measures to make the country safe

for them to live and dream. People and name of women who stand up against sexual harassment and unethical behaviour of men in Indian cities which make them uncomfortable to walk freely. The data set that was obtained through Twitter about the status of women safety in Indian society was for the processed through machine learning algorithms for the purpose of smoothening the data by removing zero values and using Laplace and porter's theory isto developer method of analyzation of data and remove re- tweet and redundant data from the data set that is obtained so that a clear and original view of safety status of women in Indian society is obtained.

II. LITERATURE REVIEW

The device/application FEMME can be made better and the usage can be increased by making the product small so that it can be used as a watch or even a pendent and also there can be a voice keyword recognition which can trigger the device to send an emergency message to the preset contacts.. the device can be made more helpful by adding any defense element in the device in order to protect oneself if the help gets delayed, as a backup option. The device has many advantages but at the same time it has many

disadvantages too. The battery consumption will be extremely high so if there is any battery backup option it will be more efficient. The proposed device is portable which has SMS options, screaming sensors and also defense element, thereby covering almost all needs. It can be more helpful by adding few more sensors like pressure sensors and detecting hidden cameras. the device can be made more useful by adding any alarm to alert the victim's surrounding areas as rapid protection can be given to the victim as Internet cannot be relied on all the time.

The device can be made more effective by adding recording system in order to record the incident taking place which in turn can help get justice to the victim or even add sensors to activate the device automatically when in danger. in addition to all the features present there can also be a defense element which helps women to deal with the threat not completely rely for other to come and rescue her. the smart band should also be able to produce an alarm or buzzer sound so that it can be

used to get public attention and the people can contribute in providing justice. Instead of sending the instant location of the victim a real time location can be shared to respective people as victim will be in panic and there is more chance of running which will lead to missing of the victim.

III.TWITTER ANALYSIS

The process of obtaining the sentiments of tweet includes three steps:

1. Data extraction: The first stage in sentiment analysis is to get data from social media sites such as Twitter. This aids in the extraction of the tweet message, but this message also contains additional information such as tweet likes, dislikes, and comments.

2. Text Cleaning: Once the data has been extracted as datasets from the social media resource, it has to sent to the classifier. Before the analysis, the classifier cleans the dataset by removing unnecessary data such as stop words and emoticons to ensure that non-textual stuff is recognized and removed.

3. Sentimental Analysis: The data is ready for sentiment analysis after the classifier cleans the dataset. Sentiment analysis employs a variety of techniques, including machine learning, Lexicon-based learning, and hybrid learning.

In this project a set of available libraries has been used.

The approach to extract sentiment from tweets is as follows:

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.

Developers have done different sentimental analysis on Twitter for different purposes and a real-time twitter sentimental analysis of the trending events happening in the world, like elections, crimes, movies etc. Figure 1 shows the sentimental analysis algorithm at a higher level.

Corpora is a structured set of texts/words which we need for analyzing the tweets.

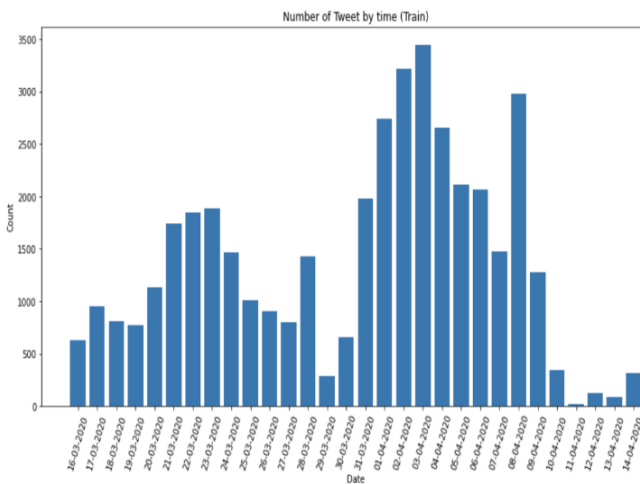


Fig. 1. Sample analysis on a high level programming

A. Initial Setup

In this paper, we have used python to perform sentimental analysis. Some packages have been utilized including tweepy and textblob. We installed the required libraries by following commands:

```
1]pip
```

```
instaltweepy
```

```
2]pip
```

```
installtextb
```

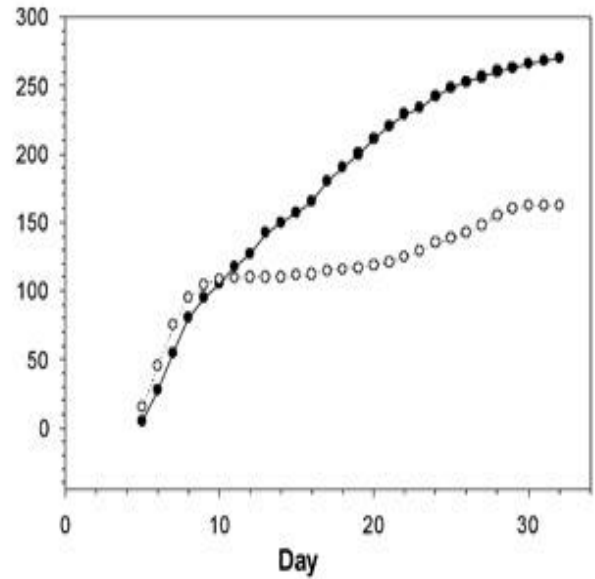
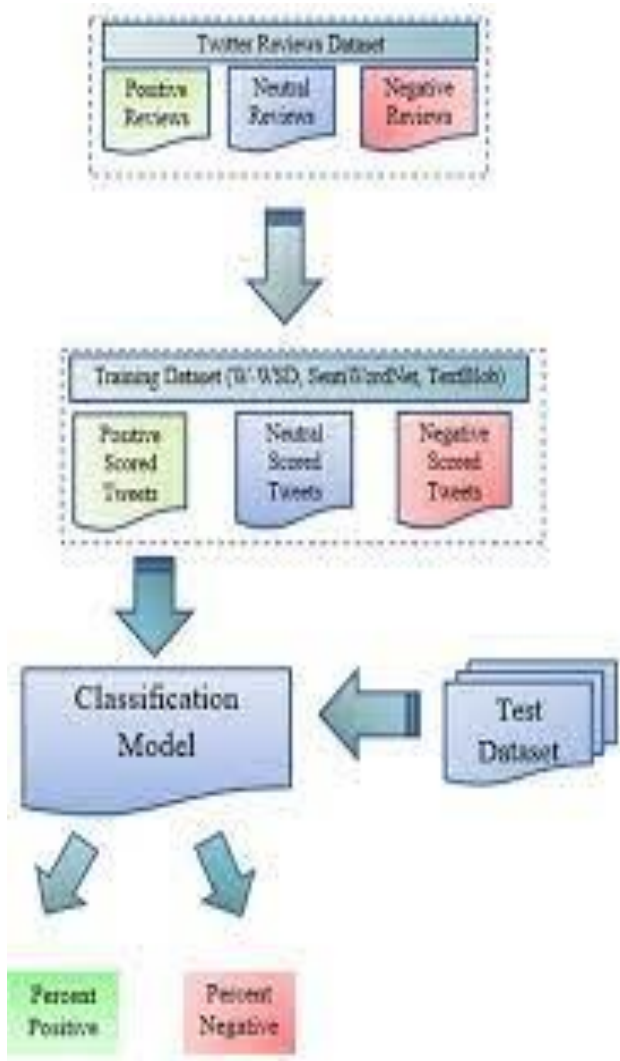
The second step is downloading the dictionary by running the following command:

```
python -m textblob.download_corpora.
```

The textblob is a python library for natural language processing and it uses NLTK.

B. Connecting to the Twitter API

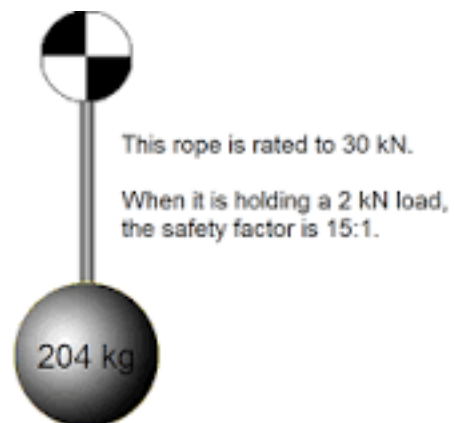
To connect to the Twitter API and query latest tweets to store it in the database, we need to create an account on twitter and create an application. We were supposed to visit apps.twitter.com/app/new and generate the api keys required to feed the program. The application settings are shown in the figure2. Due to the security reasons the API keys are not shown.



The "safety factor" is the ratio between the force that will be applied to a component in a system and the minimum breaking strength of the component. To calculate the safety factor, divide the gear's minimum breaking strength by the maximum force it will support.

How do you calculate safety factor?

The Factor of Safety of the structure is defined as $F = C/D$ and failure is assumed to occur when F is less than unity.



C. Result

Following shows the sample output of the program for the ‘rape’ as a query based on the last 300 tweets from Twitter.

- Positive tweets percentage: 14.73%
- Negative tweets percentage: 70.03%
- Neutral tweets percentage: 10.89%

Few Sample Tweets picked up from the database are:
 <tweet>="@BinaNepram Sexual assaults on women #Manipur #NortheastIndia happening since armed conflict started 1960s~In a region where hundreds thousands armed forces operate under #AFSPA with impunity sanctioned by FMME.

D.Final Report

Here we have run the report three times to know about the percentages occurred for positive, negative and neutral tweets.If the neutral tweets percentage is high that means people are not willing to have any interest in that topic and are not having any positive or negative side for the topic.

Here we will get the safety of the women by percentages.

MORE PERCENTAGE = MORE SAFETY

Here by the above experiment done by us, we can conclude that Chennai is more safer than any other cities mentioned.

II.CONCLUSION

Women have the right to be free from violence, harassment and discrimination and removing the barriers of an unsafe environment can help women fulfil their potential as individuals and as contributors to work, communities and economies. Female travelers from other countries also find themselves in a precarious position when traveling to India. So to fulfil

their potential as individuals they have to know the place they feel the safest which can be done by this software that is developed.

III.REFERENCES

[1] Agarwal, Apoorv, Fadi Biadisy, 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] Birmingham, 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.

[6] Klein, Dan, and Christopher D. Manning. "Accurate unlexicalized parsing.

[7] Klein, Dan, and Christopher D. Manning.

Metro City	Cases tweeted about	Safety factor (tweet vs population)
Delhi	1213205	72%
Mumbai	48202	94%
Kolkata	679466	86%
Chennai	8796	98%

"Accurate unlexicalized parsing." *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2003..

