Women’s Welfare on the Basis of Sentimental Analysis

Mohammad Abdul Ghani Faiz1,P Goutam Varma2, Christy A3

1,2,3 Department of Computer Science and Technology , Sathyabama Institute of Science and Technology, Chennai

*faizabdu0109@gmail.com, gautamsaivarma800@gmail.com, christy.cse@sathyabama.ac.in*

*Abstract:* Ladies and girls have been subjected to extreme physical violence in the form of sexual harassment. This particular study showcases how social media impacts a crucial part in monitoring the safety of women through different social websites like Facebook, WhatsApp, Twitter, Instagram, etc. This study also focuses on every individual in Indian society and their required responsibility in their societies must be essential for the safety of ladies around them. In the form of Twitter, different parameters like tweets, images, written messages, and quotes can guide today’s youth and punish those people who misbehave with women in every regard. The top feature of Twitter is the hashtags that can be used by women to express their views to the outer world on how they feel when they travel around in a public using public transport, how they feel when are in the presence of unknown men and do they really feel secured or not?

*Keywords : Harassment, tweets, views, hashtags, safety*

1. INTRODUCTION

There are various types of physical violence in the form of stalking, spreading abusive comments, and some unethical practices. The Research was done in many states and places in our country and many women have complained and registered cases of the same kind of harassment activities done by unknown people. This research that has been conducted in the cities like Pune, Kolkata, and Chennai has shown that more than 50% of women don’t feel safe and secure while traveling in public using public transport.

Ladies and girls have the right to move which means that they can travel to the place they want at any time, But ultimately they feel annoyed at the places like shopping malls when are going to their jobs as many unknown eyes come upon them and start making abusive comments passing by, so probably the safety and the absence of concrete situations are the major reasons for this physical and mental harassment. The reports have claimed that most of these harassment cases have been registered in the names of the neighbors as small girls feel that some instances and harassment by their neighbor on their way to school can make them feel that their whole life is destroyed, this situation basically arrives when their absence of security around those small girls on many occasions.

The cities which are deemed to be safe ensure that women’s rights are the only way possible for the general public to be aware that these unethical activities shouldn’t be carried out. Instead of pointing out women on the way they dress and the way they behave, serious actions must be taken upon the men and their approach to society as they shall ultimately know that it’s not the women who shall change their lifestyle but it’s the society who shall change their perspective and realize that men and women are equal under the eyes of the law and shall be treated in an equal manner.

This research comprises the collection of people who claim and perform these unethical activities and also the names of women who took action against this for getting harassed in public areas. The information that is taken out from Twitter through tweets and messages which showcase public harassment is carried out by machine learning algorithms for making the data smoothened Using laborers and porter’s theory the analysis of data is ensured and the repetition of tweets and their redundancy is eliminated for the clear and original result on the measures of women safety.

1. LITERATUREREVIEW

We can observe many people these days trying to give their opinion regarding the society of India through the social media platform and also regarding the politicians who claim to bring changes to society [1]. This platform also gives a great chance to people to showcase their viewpoints and especially women will have this high advantage of expressing their views whenever they faced any harassment and what they do to fight back in that particular situation[2]. Ultimately, these couple of stories displayed by women in social media and also women empowerment tweets help to bring awareness among the society to behave accordingly with women. This forms a chain of action where different people can relate their opinions and forward them to other people so that the people who are responsible for making the place of women unsafe in society. In recent times,

Large sections of people started being attracted to social media sites like Facebook, WhatsApp, Instagram, etc., and started keeping their viewpoints regarding Indian society. There are multiple methods of sentiment that can be divided into machine hybrid learning and learning that is based upon lexicon. [5]Alongside, there are various categories that the people have presented which are broadly differentiated as age-wise, approach-wise, etc. It became normal practice to intake information from a given dataset with the help of data extraction, interpretation of data, and analysis of data. The accurate values and the predicted result can be derived through the help of behavioral analysis taken from the social network dataset.

1. **ANALYSIS OF TWITTER DATA**

We know that Social Media is being an active platform where people generally express their views and opinions on various occasions, we can easily identify the emotions and sentiments of people through various social apps like Twitter, Facebook, etc. There would be different forms of opinions coming from a set of public and analytics systems that help in extracting these multiple forms of responses. Since people got used to making short abbreviations and shortcuts for expressing their views through short texts, it became a hurdle for the present Natural Language Processing systems to analyze the sentiment behind the particular text. Ultimately, Researchers have got rid of this complication within a short period of time by implementing deep learning and most importantly machine learning.

1. **EXECUTION OF SENTIMENT ANALYSIS USING TWEETS ON TWITTER DATA**

The input Tweets are extracted through API for Twitter directly given by Twitter. Because of the Abundant availability of API for Twitter, sentimental analysis has become quite efficient in calculating. The libraries which are already available within the data have been used. The execution of sentimental analysis can be performed like the:

* 1. Initially, Download the Sentimental Dictionary.
  2. Secondly, Download the test data of tweets from Twitter, referred to as the input for the program.
  3. Remove the repetitive letters and stop words to clean the provided dataset.
  4. Each and every word of the dataset have to be tokenized and added to the program.
  5. Each word in the dataset must be compared to the positive and negative sentiments of the dictionary which is been used and then we shall increase the count of total phase positively or negatively accordingly.
  6. Depending upon the count, the Result percentage can be calculated regarding the sentiment and classify the polarity as either positive or negative or else moderate.

The developers have performed sentimental analysis on various types of situations and occasions, Major analysis can be performed on the events that are trending worldwide such as sports, politics, crime, etc. The following fig 1. depicts the sentimental analysis and its algorithm on a larger scale.

We are provided with different sorts of processes to get subjected to Twitter API and take out the tweets which are visible from the algorithm used. Then cleaning and removal of repetitive words and classifying based on the polarity and at last returning the result must be considered as the further stage.

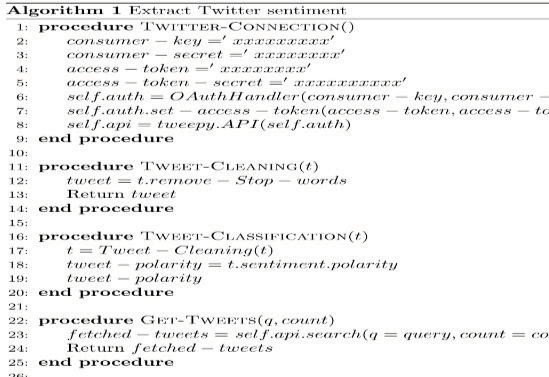


Fig.1.Model code for programming on a large scale

1. *Starting Setup*

Python is the high-level language that is used for performing the sentimental analysis we make use of the following packages like tweeps and text blob and we obtain them by executing the below commands:

1] pip install tweetpy

2] pip install text blob

Secondly, we have to download the required dictionary by executing the mentioned command:

* python-mtext blob.install\_corpora.

From the available Python libraries, we make use of text blob for the sake of Processing natural language which uses a natural language toolkit. To analyze tweets, we make use of corpora which is the finite set of texts.

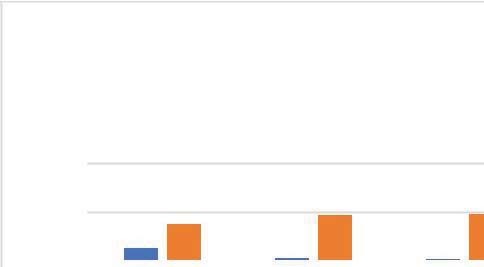
1. *Joining Twitter API*

To get through the Twitter API, and record the latest tweets made on Twitter and take them into the database, An account must be created, and then we need to create an application followingly. We need to go through apps.twitter.com/app/new to collect the required API keys and are necessary to connect these to the given program. In the following fig 2, The Application settings are displayed and API keys are prevented from being displayed due to the security permissions.

parliament, No one who is in #outfit is charged from #rape punished in 40 years”. <tweet>=”@UnderageOrphans, rape accused in refugee camp asking for justice and security. #refugees #multan #pakistan”.

<tweet>="@So much disturbing to come across the big names in the #MeToomovement like ~~#~~AveshRath,now ~~#~~KaranArora?Terrible feeling witnessing the telecast. "

The implementation of sentimental analysis (final graph) is below: Security chart for women on the basis of tweets in different cities(in terms of percentage):



200

100

0

Delhi

Mumbai

Kolkata

Chennai

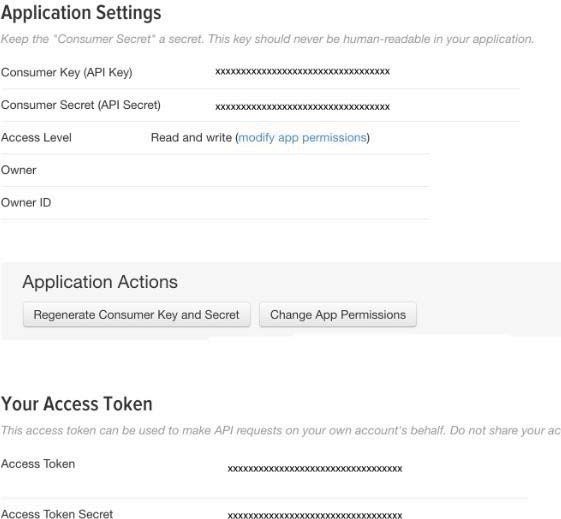
Unsafe Safe

TABLE1:Security parameter for population

|  |  |  |
| --- | --- | --- |
| Metropolitan City | Cases tweeted count | Security factor(tweets  strength) |
| Delhi | 173,943 | 74% |
| Mumbai | 42,945 | 93.2% |
| Hyderabad | 23,992 | 96.9% |
| Bangalore | 13,440 | 94% |

Fig.2.TwitterAPI application website

1. *Result*

The following result shows the sample output for the query word ‘rape’ on the basis of the previous 300 words or tweets from Twitter.

* Positivetweetspercentage:15.39%
* Negativetweetspercentage:71.13%
* Neutraltweetspercentage:13.47%

Some of the tweets taken from the database :

<tweet>=”@ReenaYadav sexual violence on ladies #Mizoram in a place of army which work under#AFSTA from

1. *Final Report*

When the program is been run numerous times, the results would be altered in all the instances with a minimal variance on the basis of tweets taken as inputs. The program runs for 3 times and the result would be the required average for provided outputs.

When the tweets that are neutral appear to be more, then it illustrates that most people are not opting for either the positive or negative side of opinion. It is also necessary to ensure that the results may vary on the basis of the different datasets used and also depending on the complexity of the situation for instance rape scenarios became a worldwide trend in the year 2015. From the depicted data, we could conclude that Hyderabad would be the most secure city and Delhi has the least percentage in the terms of safety and security.

1. CONCLUSION

This research has been so essential to gather up abundant data from Twitter like tweets and text messages which are countless in a day through the means of different machine learning algorithms, as they constitute a large part in collecting big amounts of data, especially like SPC algorithm which plays a vital role in classifying data into the respective groups. SVM is another such algorithm that is used to obtain data and examine the safety measurement of ladies in various cities of India.

REFERENCES

1. Agarwal, Apoorv, Fadi Biadsy, and Kathleen R. Mckeown."Contextual phrase-level polarity analysis using lexical affects coring and syntacticn-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. Albert Mayan J, Anto Praveena M D , Telkar Bharath Rao , Uravakonda Uday Sagar (2018),"Optimized test data generation over suspicious implementation of oracle problem",IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017),pp.2559-2563.
6. 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.
7. Joshila Grace L.K,Asha P, Selvan M.P,Sujihelen L,Christy A(2022),"Detection of Network Intrusion Using Machine Learning Technique",Communications in Computer and Information Science, Vol. 1591, pp.373–382.