

Analyzing tweets using Machine Learning for Woman Safety

Kalyani Dilip kakade, Shivani Sunil kakade, Aditi Singh, Shikha sharma, Prof. Priyanka Kale

¹ Kalyani Dilip kakade, Ballarpur Institute of Technology, Department of Computer Science and Engineering

² Shivani Sunil kakade, Ballarpur Institute of Technology, Department of Computer Science and Engineering

³ Aditi Singh, Ballarpur Institute of Technology, Department of Computer Science and Engineering

⁴ Shikha Sharma, Ballarpur Institute of Technology, Department of Computer Science and Engineering

⁵ Prof. Priyanka Kale, Ballarpur Institute of Technology, Department of Computer Science and Engineering

Abstract –Safety of women and girls is a very concerning and pressing matter in today's society. In today's society, women are continuously harassed in every corner of the neighborhood. This project focuses on the safety of women using the tweets targeting various cities and localities. Many tweeter users and tweeter handles tweets about women safety using messages, tweets, images, or quotes to educate Indian society about women's safety. Using all these tweets, messages, and quotes our team is trying to implement a system targeting small cities to get an average rating of how safe the women are in a particular neighborhood.

Key Words: Women's safety, Sexual Harassment, Sentiments, Safety, Safety Analysis.

1. INTRODUCTION

In our society, women are continuously harassed in one way or the other. 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 shaming 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 implementing a system which will analyse the tweets by the people who stands

up and raise a voice against woman safety, create a dataset of it which will be obtained from Tweeter that will be processed through Machine Learning to remove the zero works and redundant data using Laplace and porter's theory for developing a method of analysis of dataset to obtain a clear view of woman safety in the society.

2. 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.

3. 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.

4. 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.

Algorithm 1 Extract Twitter sentiment

```

1: procedure TWITTER-CONNECTION()
2:   consumer - key = 'xxxxxxxx'
3:   consumer - secret = 'xxxxxxxx'
4:   access - token = 'xxxxxxxx'
5:   access - token - secret = 'xxxxxxxx'
6:   self.auth = OAuthHandler(consumer - key, consumer -
7:   self.auth.set - access - token(access - token, access - tc
8:   self.api = tweepy.API(self.auth)
9: end procedure
10:
11: procedure TWEET-CLEANING(t)
12:   tweet = t.remove - Stop - words
13:   Return tweet
14: end procedure
15:
16: procedure TWEET-CLASSIFICATION(t)
17:   t = Tweet - Cleaning(t)
18:   tweet - polarity = t.sentiment.polarity
19:   tweet - polarity
20: end procedure
21:
22: procedure GET-TWEETS(q, count)
23:   fetched - tweets = self.api.search(q = query, count = co
24:   Return fetched - tweets
25: end procedure

```

Fig. 1. Sample code on a high level programming

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 install tweepy
- 2] pip install textblob

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

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 below.

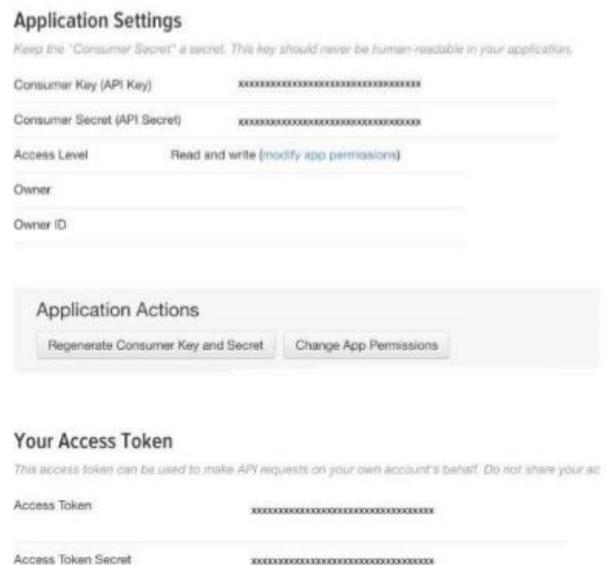


Fig. 2. Twitter API application website

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.

5. CONCLUSIONS

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 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] 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.