International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

Women Safety Analysis based on tweets using Machine Learning

Mrs. Chandini U[1], Harika M[2], Chandana K P[3], G Menaka[4], Harshitha s[5]

[1]*Assistant Professor, Dept. of Computer Science and Engineering, Bangalore*
[2,3,4,5]*Student, Dept. of Computer Science and Engineering, Bangalore, Karnataka, India*

Abstract - *Even from the earlier days, women safety has been a major issue to deal with. In the olden days there was discrimination between a girl and a boy however nowadays it has reduced. But in terms of women safety, it is the same, there is not much change. The Women of our society are mistreated in a number of ways, mentally and physically. This makes them feel unsafe. Making them feel that they are not strong enough to fight and protect themselves against harassment. Society is one of the reasons why women feel unsafe and scared. The society should be responsible for women safety and make them feel safe and protected. We aim to create a society where women can feel safe and protected. In this digital era, people tend to share their views and opinions about issues and incidents taking place all over the world with the help of networking sites such as Twitter, Facebook, and Instagram etc. However, some people tend to misuse this opportunity and make unhealthy comments about the women of our society. This will result in women being mentally abused and will instill fear in them. Collecting such data from Twitter help us to analyze the safety levels in the various Indian cities. Applying various machine learning algorithm on tweets, and performing analysis on them to classify them into neutral, negative and positive can help us improve the situation.*

Key Words: **Women safety, Twitter tweets, machine learning algorithms, sentiment analysis.**

1. INTRODUCTION

Social media is on the rise in present times. Every individual gets to put forth their personnel opinion about the various issues from all over the world. Any person from one country can raise his/her voice regarding an issue related to some other country. Social media is made of several networking sites such as Twitter, Facebook, Instagram, among the other widely used sites. Twitter is one such media that is on the rise and accessed by many people all over the world to express their opinions.

Using tweets from Twitter to perform sentiment analysis is a great way to analyse the sentiments of people regarding the various issues in the world. Sentimental Analysis can be performed over a variety of tweets to perform analysis, a variety of topics such as product evaluations, movie reviews, and so on can be understood better using machine learning. The goal here is to use various machine learning algorithms to analyse the tweets and classify them as positive, negative, and neutral. With the rise in the usage of social media Twitter sentiment analysis is an excellent way to find the opinion of people.

Classifying the tweets will help us understand the opinion of the public regarding various issues. Women's safety is a rising issue all over the world. Twitter is a great platform that can provide suitable data for us to analyse the safety levels and the steps being taken to increase the safety levels of women. Performing a Twitter sentimental analysis on tweets related to women can help us understand the society in which we all are living. This understanding can further help us improve the environment for the women of our nation. The tweets that are extracted from Twitter will contain several words and text which is not related to the content and context that we seek. Twitter is a platform where people express their opinions using single-line sentences and emojis. The tweets directly obtained from Twitter cannot be used directly since they are considered noisy data. While performing analysis not only the emojis even hashtags (which are used quite often in tweets) and white spaces should be removed to get better accuracy in the results. The various punctuation marks used in the sentences should also be removed to make the data noise-free.

Especially in India atrocities against women are greatly increasing, under such circumstances analyzing to understand the women's safety levels in the country is essential. Previously talk shows and surveys were conducted on Women's Day to analyze and discuss the precautions taken for women's safety. However, considering the present times using social media especially Twitter, and applying machine learning algorithms can help us analyze the safety levels in the various Indian cities. Various algorithms such as Naive Bayes, Linear regression, and others are available in machine learning to perform analysis and classify the data into positive, negative, and neutral.

Twitter is the most compatible social media platform to perform sentiment analysis. The type of content shared on each social media platform varies with the features provided by it. Facebook is a platform in which the content shared is at times large in size and hence is not suitable for analysis. Instagram is a site where the content shared mainly focuses on pictures and videos rather than text. Thus, when compared with the other social media platforms Twitter is the most suitable networking platform since the content shared on twitter is mainly text and emojis. Though the content from twitter is also not structured and not in the required form for analysis, it can be cleaned and processed easily before use making it more suitable than others.

1.1 SENTIMENT ANALYSIS

Sentiment analysis is a technique that involves understanding the subjectivity of a sentence. It helps us to recognize the underlying intentions and sentiments of the user while making a statement. The sentiment analysis techniques can be broadly sorted into the following categories.

- Lexicon-based approach: It is also referred to as a Rule-based system. This approach follows semantic analysis i.e., makes use of a lexicon (i.e., a record of positive and negative words) to perform the analysis and classify the text.
- Machine-learning based approach: This technique mainly involves the usage of various algorithms such as Naive Bayes, Support vector machine, etc.
- Hybrid system: This technique involves combining the previous two techniques to produce an improved accuracy and results by overcoming the limitations of the individual techniques.

1.2 LITERATURE REVIEW

Now a days social media has been playing an important role in connecting people and letting them share their opinion openly on any topic [1]. The amount of data being shared on social media platforms is so huge that it can be utilized to perform an analysis providing an ocean of information to gain knowledge about people's understanding of a particular topic. From predicting customer review on a product to predicting the box-office collection of a movie, all this can be possible by analyzing people's comment.

Women have been facing various forms of violence since the ancient times and it continues even now in the modern times. People nowadays share their reaction on such events with the public using social media platforms [2]. Analyzing such data will lets us know what kind of violence and at what rate it is being performed in a particular city. There are people who misuse social media platforms and comment negatively against a woman which affects the women mentally and creates a fear inside them which in turn makes them feel unsafe [3]. Twitter is one such social media platform that allows people to connect and share content with each other. Sentiment analysis is a process performed to understand language, where the data collected from twitter will undergo several levels of processing and will then be analyzed to come forth with a result regarding a topic [4]. The various levels include pre-processing of data and then applying different algorithms on them to classify them as per our need to perform analysis. Women safety level in the different Indian cities can be predicted by performing sentiment analysis on twitter data obtained directly from Twitter using Twitter API. The analysis can help one understand the mentality of people regarding the women of their society and this data can further be used by the NGO's or various other departments of women development to spread awareness about the importance and role of a women in our society. This can help improve the women safety level of in a particular city. One's mentality towards anything is the foundation to the action performed by them, thus if the mentality of a person can be improved eventually the actions of a person will improve.

2. IMPLEMENTATION

Sentiment analysis is a task of natural language processing that can be implemented to classify the data (i.e., tweets) into different categories based on the sentiment associated with them. In this technical paper, machine learning concepts are being used to classify tweets into positive, negative and neutral and display them in the form of graph, pie-chart and in the terminal. Analyzing tweets involves various stages which are discussed further in the subsections:

A. Initial Setup

In this project we have used python as a programming language and PyCharm as IDE to carry out the analysis. Before performing the sentiment analysis, we are required to install some required libraries. The required libraries can be installed using the following commands:

a. pip install tweepy
b. pip install textblob
c. pip install Django
d. pip install matplotlib

The next step is to install the dictionary. In this paper we are not creating our own dictionary, but using the inbuilt dictionary Corpora, which is a structured set of words that are needed to analyze the data. The following command should be used to download the dictionary.

- Python -m textblob.download_corpora

B. Data extraction

Data here refers to tweets collected from Twitter that are related to women and their safety. These tweets can be easily acquired from twitter using TwitterAPI. Twitter API is a programming interface that provides us with various tools that enable easy access of tweets. TwitterAPI makes use of HTTP get requests to retrieve tweets. Following these simple steps can enable us to easily connect and retrieve data from twitter.

1. First, create/log in to your Twitter developer account.

2. Click on create new app (By filling in all the required details on the app creation page).
3. Now, the project will be created, once it is created, to get the consumer secret and consumer key click on "Keys and Access Tokens" tab. These consumer secrets and consumer keys will be used for authentication purposes.

C. Sentiment analysis

The tweets will be retrieved from twitter using keywords like as "women safety mumbai". The keyword is matched with tweets to retrieve the related tweets for analysis. Once the tweets are collected, preprocessing of these tweets plays a highly important role because the maximum data on social media or networking sites are unstructured. In order to convert it into structured or formatted data or the required format, preprocessing is done. Preprocessing of the collected data involves a number of processes such as Lemmatization, Tokenization, etc. Some of the processes performed in the paper are:

The classification of the data i.e., tweets related to women safety is done using the sentiment method of textblob package which makes use of polarity, subjectivity, and intensity to classify the data. The polarity ranges from [-1,1] where a -1 indicates negative, 0 indicates neutral, and 1 indicates positive.

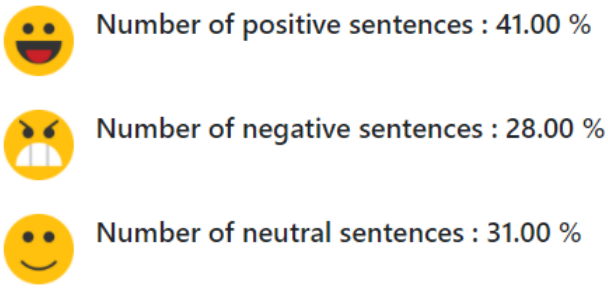
ALGORITHM:

```
function Connect_with_Twitter()
    consumer_key = 'xxxxxxxxx'
    consumer_secret = 'xxxxxxxx'
    access_token = 'xxxxxxxx'
    access_token_secret = 'xxxxxxxx'
    self.auth = OAuthHandler(consumer_key, consumer_secret)
    self.auth.set_access_token(access_token, access_token_secret)
    self.api = tweepy.API(self.auth)
end function

function collect_tweets(tweet_count)
    collected_tweets = self.api.search(q=query,count=tweet_count)
    return collected_tweets
end function

function clean_tweets(tweets)
    t= tweets.remove_stop_words
   return t
end function

function classify_tweets(tweets)
    pre_processed_tweet = clean_tweets(tweets)
    tweet_polarity = pre_processed_tweet.sentiment.polarity
    Classify using tweet_polarity and Return value.
end function
'
```

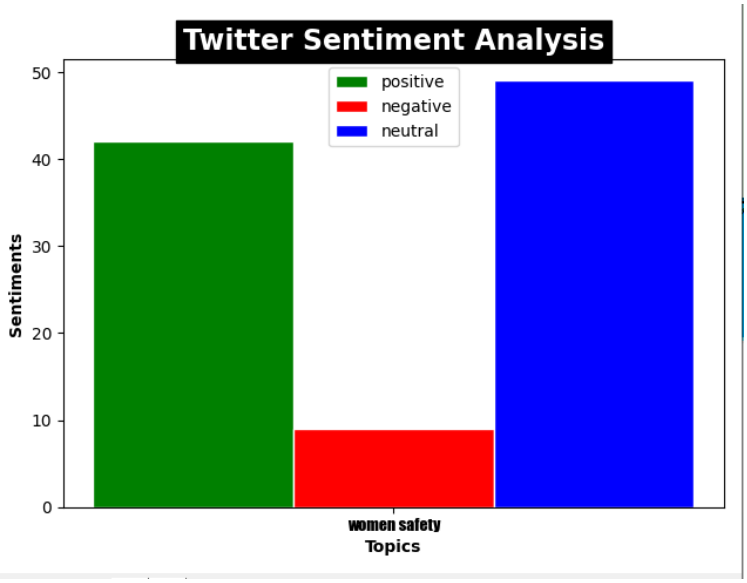
D. Results

Finally, the analysis and classification is displayed in an easily understandable manner. The results are displayed in the form of graphs. The bar graph shows the positive, negative, and neutral percentages, whereas the pie chart shows a more diverse classification representing strongly positive, positive, weakly positive and so on for the negative sentiment as well.

The percentage of positive, negative, and neutral tweets for the keyword 'women safety mumbai' is as follows:

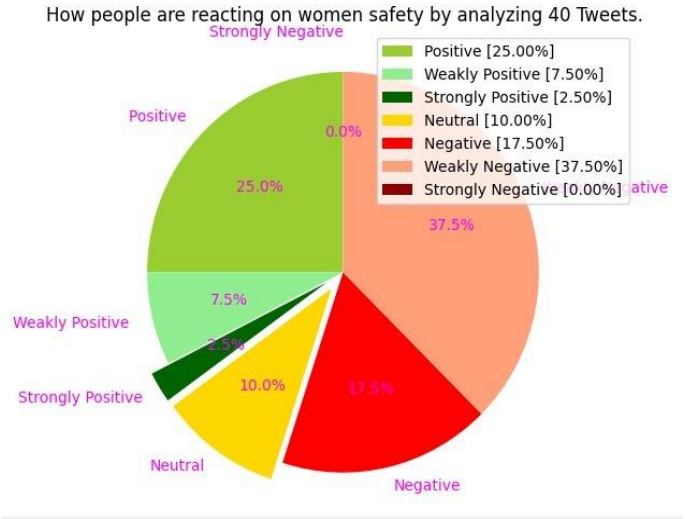


The bar graph indicates the percentage of positive tweets in green, percentage of negative tweets in red, and percentage of neutral tweets in blue. The bar graph showing the positive, negative, and neutral percentages for Mumbai city is as follows:



Graph1. Bar graph representation of results

The pie chart gives a more diverse classification, representing the classification as strongly positive or negative, weakly positive or negative, and so on.



Graph 2. Pie Chart representation of results

3. CONCLUSIONS

In this paper we have discussed about the various approaches of sentimental analysis available to analyze data and perform meaningful classifications. In the paper we have used the Lexicon-based approach to classify the data (i.e., Sentiment method). We can further make use of Machine-Learning based approach or the Hybrid approach to perform further classification to obtain subgroups of the classification. Naïve Bayes and Support Vector machine are some of the widely used classifiers which can be used to perform in-depth classification making use of attributes and considering the factors that affect women safety and in turn the classification.

REFERENCES

[1] S. Asur and B. A. Huberman, "Predicting the Future with Social Media," 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010, pp. 492-499, DOI: 10.1109/WI-IAT.2010.63.

[2] S. Ramamoorthy, R. Poorvadevi, "Safety Measures against Women Violence in India using Sentimental Analysis," 2019 IJITEE, ISSN: 2278-3075, Volume-8, Issue-6S3, April 2019.

[3] Vikram Chandra, Rampur Srinath, "Analysis of Women Safety using Machine Learning on Tweets," e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume: 07 Issue: 06 | June 2020.

[4] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, "Sentiment Analysis of Twitter Data," Department of Computer Science, Columbia University, New York, NY 10027 USA.

[5] D. Kumar and S. Aggarwal, "Analysis of Women Safety in Indian Cities Using Machine Learning on Tweets," 2019 Amity International Conference on Artificial Intelligence (AICAI), 2019, pp. 159-162, DOI: 10.1109/AICAI.2019.8701247.

[6] Reena G. Bhati, "Sentiment analysis a deep survey on methods and approaches", International Journal of disaster recovery and business Continuity, Pune, India.

[7] Abdullah Alsaeedi, Madinah, "A study on sentiment analysis techniques of Twitter data", International Journal of advanced computer science and applications, KSA.

[8] Nie, Dong & Guan, Zengda & Hao, Bibo & Bai, Shuotian & Zhu, Tingshao. (2014). Predicting Personality on Social Media with Semi-supervised Learning. 158-165. 10.1109/WI-IAT.2014.93.