

MACHINE LEARNING OF TWITTER FEEDS AND WOMEN SAFETY IN INDIAN CITIES

S. Gokulakrishnan

Department of CSE, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram, India
Email: s.gokulakrishnan@kanchiuniv.ac.in

Chitte Ashok Reddy, Palutla Venkata Sai Anudeep

Department of CSE, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram, India

Received: 10th September 2022, Accepted: 3rd October 2022 and Published: 3rd October 2022

ABSTRACT

Aim: The paper focuses on the role of Twitter feeds in finding safety aspects of women and girls in Indian cities using machine learning algorithms.

Results: The data set obtained through Twitter about women and girls' safety status in Indian cities is analyzed using machine learning tools.

Conclusion: Machine learning algorithms help organize and analyze Twitter data, including millions of daily tweets and messages. The same can be extended to other social media platforms.

Keywords: Accuracy, Harassment, Sarcasm, Emoticons.

HIGHLIGHTS:

A simple machine learning algorithm will help analyze tweet feeds concerning girls' safety.

PROPOSED SYSTEM

Women and girls sometimes face safety situations or harassment, and most of the time, nowadays, they post on Twitter. Analysis of Twitter texts collection will help understand harassment and amoral behaviour of men in Indian cities [1]. The tweets are downloaded from Twitter API, and the following methodology is adopted.

1. Starting with downloading the sentimental dictionary
2. Then download the Twitter testing data sets and insert them as input to the program.
3. The tweets are cleaned by removing stop words and repeated letters.
4. Tokenize each word, allot strength to the words within the dataset, and feed it to the program.
5. For each word, compare it with positive sentiments and negative sentiments word dictionary, then increment the general phrase's positive or negative count.
6. Finally, based on the positive count & negative count, we can get the result percentage about sentiment to finalize the polarity, categorized as Positive, Negative and Neutral.

Women are monitored and confirmed tweets by using sentiment analysis on user data. The admin analyzes every user's tweet to ensure women's safety on social networking sites like Twitter and Instagram. It will be possible to analyze users' tweets recorded in the database. Now, the administrator can prepare the data for analysis. The initial input for sentiment analysis will be the tweets produced by each application user, and they will comprise the dataset. Text analysis graphs can be displayed in addition to this. The filters will be stored in the database by the admin. *Filters* are keywords searched for in the tweet context to determine if it is abusive or not. Positive keyword and negative keyword filters are two types of filters that can be used. A positive keyword is not abusive to women in any way. Keywords that women will not find offensive are considered negative keywords. Positive and negative keywords can be stored in the database in infinite combinations. An emotional analysis will be conducted by an administrator who will compare every keyword in the database against the user's Tweet. Tweets categorized as "positive sentimental analysis" contain at least one of the optimistic keywords in their content. It will be categorized as a negative sentimental analysis if a negative keyword is identified in the tweet. Based on the filter in the database, two types of sentimental analysis will be performed at this point.

RESULTS

Social networking communications and machine learning algorithms can assess women's safety. In order to analyze women's safety, the author has implemented the following modules: MEETOO tweets on women's safety and security into a dataset folder instead of utilizing the TWEETPY tool from Python (Figure 1). NLTK (natural language tool kit) cleans up the tweets. The TEXT BLOB corpora package and dictionary count positive, negative, and neutral polarity. After clicking the 'Run Machine Learning Algorithm' button, the program estimates the Tweet data using machine learning. Each tweet is displayed on the above panel, along with its text and polarity score. All tweets can be viewed by scrolling down above the input field. Now, clicking on the "Women's Safety Graph" button helps to examine the results, which will help you determine whether or not the location is safe for women. Because of the negative sentiment expressed in the results, 74% of the location can be confirmed unsafe for women.

CONCLUSION

Machine learning algorithms help organize and analyze Twitter data, and the women's safety graph in the program helps understand the safety aspects in Indian cities. The machine learning algorithms can be extended to other social media platforms such as Facebook and Instagram.

REFERENCES:

1. <https://doi.org/10.1109/AICAI.2019.8701247>

FIGURES:

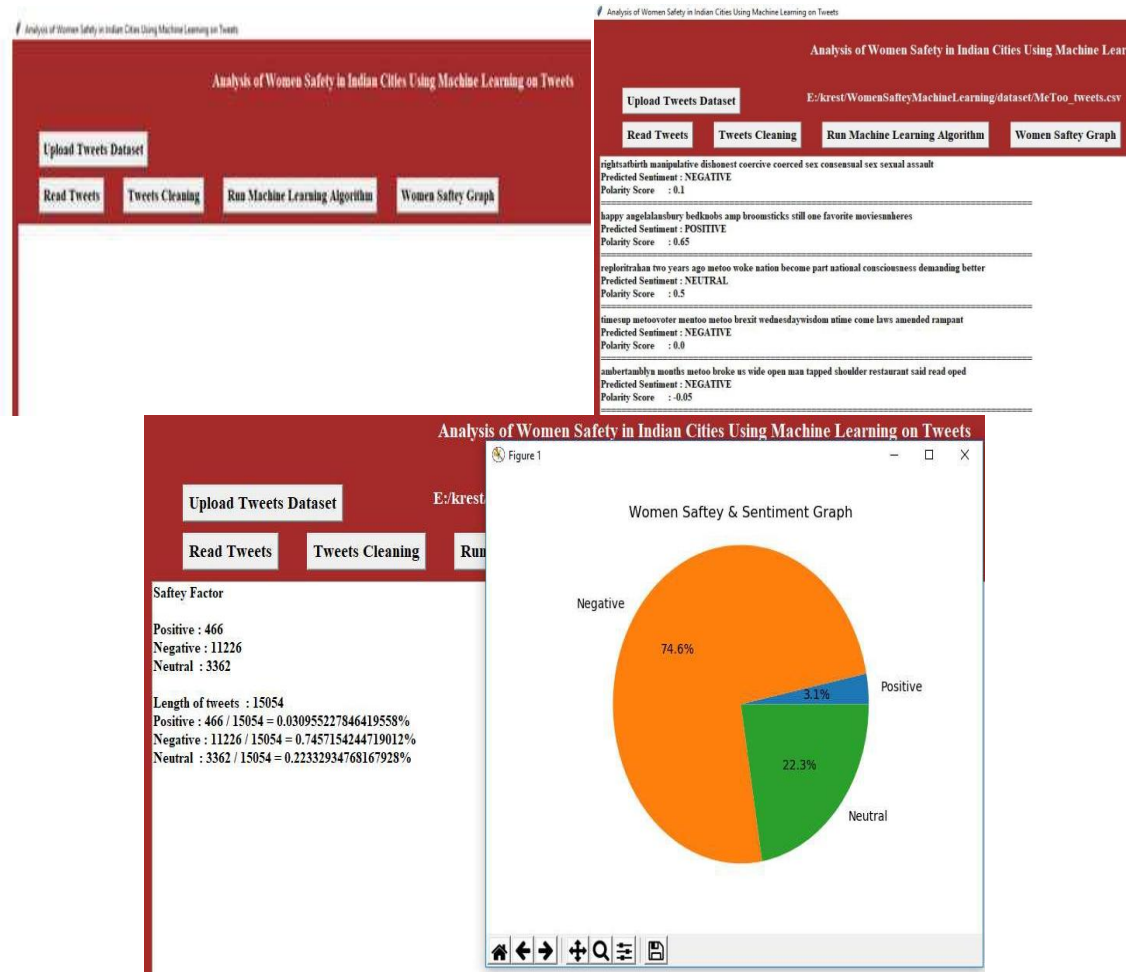


Figure 1: Machine Learning Program, loading tweets and generated results