

COPY RIGHT



ELSEVIER
SSRN

2023 IJEMR. Personal use of this material is permitted. Permission from IJEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJEMR Transactions, online available on 10th Apr 2023. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04](http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04)

10.48047/IJEMR/V12/ISSUE 04/85

Title **WOMEN SAFETY BASED ON TWEETS BY USING MACHINE LEARNING**

Volume 12, ISSUE 04, Pages: 699-704

Paper Authors

K. THRILOCHANA DEVI, KOTA TARUN, KOLLURU HEMANTH RAHUL, MUNTHA VAMSI RAJ,

P SUMANTH CHOWDARY



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

WOMEN SAFETY BASED ON TWEETS BY USING MACHINE LEARNING

K. THRILOCHANA DEVI¹, Assistant Professor, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

**KOTA TARUN², KOLLURU HEMANTH RAHUL³, MUNTHA VAMSI RAJ⁴,
P SUMANTH CHOWDARY⁵**

^{2,3,4,5} UG Students, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

¹ thrilochanaprakash@gmail.com , ² tarunkota2002@gmail.com

³ hemanthrahulkolluru@gmail.com, ⁴ vamsimuntha31@gmail.com

⁵ Pamidisumanth9@gmail.com

Abstract

Women and girls encounter violence and abuse in public places across many different nations. Today, there is a serious issue with girls protection. Social media platforms like Twitter are increasingly becoming a space for women to share their experiences and raise awareness about these issues. With the help of machine learning techniques, it is possible to analyse the tweets related to women's safety and gain insights into patterns and trends of such incidents. In this research paper we focused on the role of social media in promoting women's safety in Indian towns. Tweets are essentially just people's views because people frequently share their tweets on Twitter in the form of voice, video, images, etc. One of the platforms where people can post their opinions openly is Twitter. Tweets about the protection of women in Indian towns can be interpreted as a call to action by polite people, and they may result in women who are in difficulty being punished.

Keywords: Women safety, Sentimental Analysis, Machine Learning, pre-processing.

Introduction

Women's safety has always been a critical concern in our society, and with the increasing usage of social media platforms, such as Twitter, it has become even more important to understand the trends and patterns related to women's safety. Twitter has emerged as a popular platform for individuals to share their

experiences and thoughts in recent years, there have been several incidents reported on Twitter that have raised serious concerns regarding women's safety, and it has become a platform for women to voice their concerns and experiences related to harassment and abuse. By using of several machine learning models and sentimental analysis techniques, it has become likely to analyse large amounts of data on Twitter and extract valuable understandings associated with women's safety.

Sentimental analysis is being used. Sentimental Analysis indicates it is a natural language processing (NLP) technique used to determine whether information is indicates either negative, or positive. In this analysis, we will use machine learning algorithms to examine Twitter data related to women's safety and explore the patterns and trends in pain, abuse, and violence against women.

We can develop models that can predict the likelihood of certain events or identify areas of concern related to women safety. This type of analysis can help organizations and policy makers to better understand the issue and develop effective strategies to address it.

In this we will explore the use of machine learning for analyzing tweets related to women's safety. We will discuss the methods and techniques used to preprocess and analyze the data, as well as the challenges and limitations of this approach. Over all Our goal is to demonstrate the potential of this approach for understanding and addressing the issue of women's safety.

Literature Survey

S. Rammorthy and R. Poorvadevi[1] proposed a method for identifying various forms of violence and threats against women using data from popular social media platforms. They collected data such as likes, tweets, comments, blogs, and posts related to incidents involving women, and analysed the sentiment associated with these data points, classifying them as positive, negative, or

neutral. The sentiment analysis was conducted using a range of emotional states such as anger, disgust, positivity, negativity, and sadness, and was graphically represented.

To perform sentiment analysis, the authors used a supervised machine learning algorithm called a Neural Network. This involved applying recursive neural and composition functions to the collected data, analyzing the sentiment, and predicting sentiment labels. The resulting sentiment labels were then used to represent measures for ensuring women's safety.

Madhura Shridokar, Sayali, Ashela Ingole, and Pragati Mishra [2] proposed that Supervised machine learning algorithms like Naive Bayes and SVM are utilized to categorize data based on political parties. Previous studies have mainly utilized Twitter data, but it is possible to gather data from other social media platforms such as Facebook and Instagram. Typically, the data is analysed in the form of words and sentiment analysis is performed to determine the word's polarity. Special characters like '@' and URLs can be stripped off to overcome the noise. Enhancing the integrity of the data by eliminating noise is one of the most crucial pre-processing objectives.

However, algorithms classify the data, analysing sentimental analysis is done on classified data and then on a sentence level may yield more accurate results than analysing individual words. To obtain the outcome, the sentiment percentages

obtained by using the algorithms can be compared across different political parties.

Abdullah Alaseedi, and Mohammed Zubair Khan [3] proposed to Analyse the emotional state of users by studying their tweets. Sentimental Analysis is done here with different classification techniques like supervised, and unsupervised approaches. To classify unlabelled data various machine learning model were used. Maximum entropy gives better compared to naive Bayes. when they used the Multinomial naives Bayes algorithm gave better performance. The accuracy is increased when compared to naive bayes when the algorithm is logistic regression.

Problem Identification

Nowadays women's safety is a major issue in the world. Attacks and insults on women are increasing day by day. Ladies need to be resilient on the inside and out. NCW reportedly highest number of 31,000 complaints crimes against women in the last year. Today almost everyone has social media accounts like Twitter etc...People tweet from any place, at any time. It tells how it impacts women whether it is good or bad on women. On Twitter people tweet about women either in a respectable manner or a disrespectful able manner. Indirectly tweets are nothing but their opinion. On Twitter, everyone expresses their thoughts. Women don't know idea how much they safe are. In earlier days polarity detection is done using naive Bayes, Maximum Entropy,

and Random trees but by using Support Vector Machine we achieved high accuracy and Performance. Pre-processing and applying sentimental analysis by using a classification algorithm helps women to know and visualize graphically how much positive, and negative impacts they have in some Indian cities. So, they take care of themselves while traveling and can move themselves according to it.

Methodology

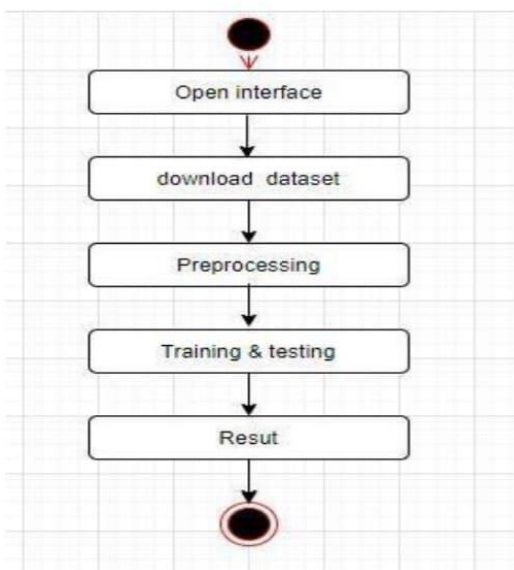
The dataset which contains both positive and negative tweets. preprocessing and preparing raw data and making it into the clean dataset. which is useful for the machine learning model.

A sentimental Analysis was going to be done. Sentimental Analysis means it is a natural language processing technique deciding whether it is positive or negative.

Training data and testing data means training data is subset of the original data and it is used for train the machine learning model.

Testing data means it is used to checking the accuracy of the model

It categorizes data into graphical forms and helps visualize to us.

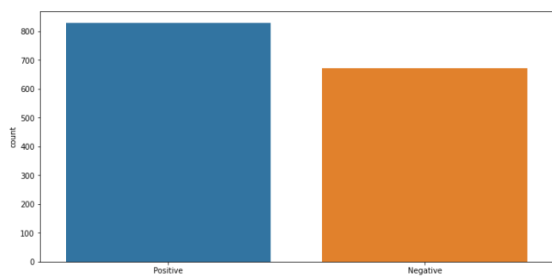


Implementation

We will need following:

1. Having access to the Google Colab.
2. A dataset contains tweets, and it is .csv file.

Under any application, we start with fetching a database that makes it appropriate for our model. Initially, we begin with importing the necessary library files like import pandas, numpy, nltk, and Gensim packages by using packages we can read the data and works with analyzing, cleaning, and examining. NLTK package in python is used to perform some statistical operations and after sentimental analysis, happened.



Calculated how many positive and negative tweets were by the given dataset.

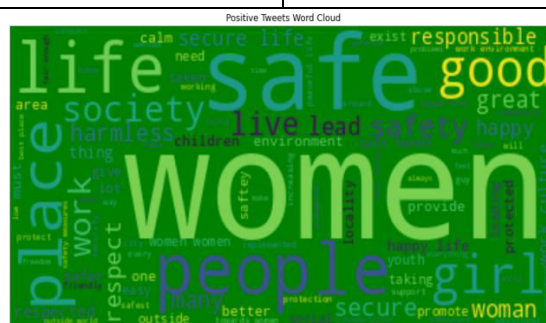
By using Pandas Package dataset with tweets are organised into tweets.

S. No	Tweets	Sentiment	City
1	is having safest transportation in tr	1	mumbai
2	this is haing software industries wh	1	Kolkata
3	they the women are having good re	1	chennai
4	we women are having a joyful atm	1	mumbai
5	this is promoting many rules about	1	kolkata
6	the people there are very geneous	1	chennai
7	it is place where women are having	1	mumbai
8	it is the best place for women stud	1	kolkata
9	it has a good proteins for women to	1	chennai
10	this place has many women as infl	1	mumbai

Intially tweets are downloaded, and it contains punctuations and many stop words so all noisy data will be cleaned and pre-processing is performed on tweets and only tweets are extracted into clean tweets.

To give more visualization word cloud is used. From Word cloud package means the visual depiction of knowledge or data is a word cloud. Making the most used words stand out from the surrounding terms illustrates.

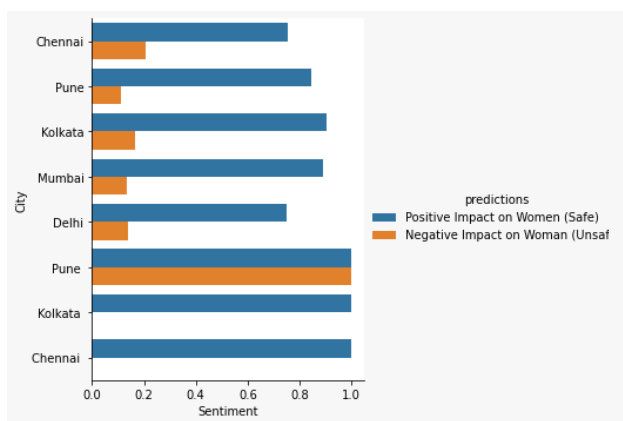
Positive	Negative
Safe	Unsafe
Happy	Difficult
Responsible	Harassment
Good	Violence
Secure	Harmful
Great	Abuse





By using SVM model we can achieve highest accuracy's SVM means Support Vector machine. The SVM algorithm's objective is to establish the best line or judgement boundary that can divide n-dimensional space into groups so that we can quickly categorize new data points in the future.

Result



This bar chart illustrates about on positive and negative on women's safety by comparing different cities.

Conclusion

We've talked about the machine learning algorithms. By applying some algorithms that are efficient at classifying big amounts of data into useful databases and which contains both positive and negative tweets are the SVM algorithm and linear algebraic. Consequently, we can improve the protection of women by employing machine learning techniques to carry out emotional analysis and raise awareness

Limitations

It is only relates to Twitter applications that are built solely on tweets. It is only for a few large towns. It did not provide varying safety standards for various cities.

Future Scope

We took only in our research paper, but by applying the required machine learning algorithms to other social media platforms like Facebook, Instagram, etc., we can expand it. The suggested ideology could be integrated into the Twitter application to reach a wider audience and apply sentiment analysis to millions of tweets to increase security. Depending on the safeties in metropolitan cities we can improve our safety measures and spread awareness about it.

References

[1] S.Rammorthy et al .”Safety Measures Against Women Violence in India using Sentimental Analysis”, International Journal of Innovative Technology and Exploring Engineering, ISSN:22782 3075, p. no:150-154, April 2019

[2] MadhuraShirodkar “Result Prediction using Twitter data”, International Research Journal of Engineering And Technology, ISSN:2395-0056, 2 Feb 2020

[3] AbdullahaAlasedi “Sentimental analysis of Twitter data”, International Journal of Computer Application, ISSN:2395-0072, 2 Nov 2019

[4] EmanAlyami,” Mining Saudi Twitter sphere for expressive support towards women” Fourth World Conference on Smart Trends in system, 26 Oct 2020

[5] Sahayak V, Shete V, Shete V Pathan A(2015),”Sentimental analysis of twitter data”, International Journal of Innovative Research in Advanced Engineering, 2(1), 178-183

[6] 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