



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.488

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com



Analysis and Detection of Women Safety Using NLP Algorithm in Machine Learning on Tweets

Sermakani A M¹, Divya Priya.R², Madhumathi.R², Yuvarani Ms²

Associate professor, Department of Information Technology, S.A Engineering College, Chennai, Tamil Nadu, India¹

B. Tech Student, Department of Information Technology, S.A Engineering College, Chennai, Tamil Nadu, India²

ABSTRACT: Women and girls have been experiencing a lot of violence and harassment in public places in various cities starting from stalking and leading to sexual harassment or sexual assault. This research paper basically focuses on the role of social media in promoting the safety of women in Indian cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram. This paper also focuses on how a sense of responsibility on part of Indian society can be developed the common Indian people so that we should focus on the safety of women surrounding them. Tweets on Twitter which usually contains images and text and also written messages and quotes which focus on the safety of women in Indian cities can be used to read a message amongst the Indian Youth Culture and educate people to take strict action and punish those who harass the women.

KEY WORDS: Machine Learning , NLP Algorithm , Women Safety.

I. INTRODUCTION

There are certain types of harassment and Violence that are very aggressive including staring and passing comments and these unacceptable practices are usually seen as a normal part of the urban life. There have been several studies that have been conducted in cities across India and women report similar type of sexual harassment and passing off comments by other unknown people. The study that was conducted across most popular Metropolitan cities of India including Delhi, Mumbai and Pune, it was shown that 60 % of the women feel unsafe while going out to work or while travelling in public transport. Women have the right to the city which means that they can go freely whenever they want whether it be too an Educational Institute, or any other place women want to go. But women feel that they are unsafe in places like malls, shopping malls on their way to their job location because of the several unknown Eyes body shaming and harassing these women point Safety or lack of concrete consequences in the life of women is the main reason of harassment of girls. There are instances when the harassment of girls was done by their neighbors while they were on the way to school or there was a lack of safety that created a sense of fear in the minds of small girls who throughout their lifetime suffer due to that one instance that happened in their lives where they were forced to do something unacceptable or was sexually harassed by one of their own neighbors or any other unknown person. Safest cities approach women safety from a perspective of women rights to the affect the city without fear of violence or sexual harassment.

II. RELATED WORK

Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or

information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto tagging, recommender system, and many more.

A Machine Learning system learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it. The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.

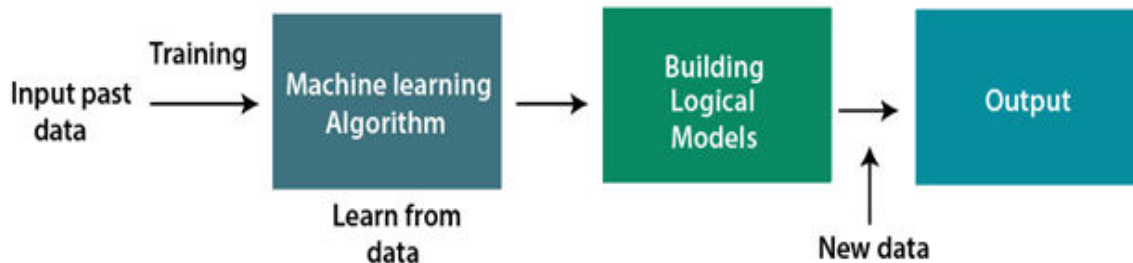


fig (1.1)ML diagram

1) Supervised Learning

Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output. The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not. The goal of supervised learning is to map input data with the output data. The supervised learning is based on supervision, and it is the same as when a student learns things in the supervision of the teacher. The example of supervised learning is spam filtering.

2) Unsupervised Learning

Unsupervised learning is a learning method in which a machine learns without any supervision. The training is provided to the machine with the set of data that has not been labeled, classified, or categorized, and the algorithm needs to act on that data without any supervision. The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns. In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data.

III. PROPOSED SYSTEM

This research paper basically focuses on the role of social media in promoting the safety of women in Indian cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram and also crime details over all India. This paper also focuses on how a sense of responsibility on part of Indian society can be developed the common Indian people so that we should focus on the safety of women surrounding them.

Crime details in India which usually contains images and text and also written messages and quotes which focus on the safety of women in Indian cities can be used to read a message amongst the Indian Youth Culture and educate

people to take strict action and punish those who harass the women. It is a common practice to extract the information from the data that is available on Crimes through procedures of data extraction, data analysis and data interpretation methods using Exploratory data analysis (EDA) concept, the accuracy of the Crime analysis and prediction can be obtained by the use of Time series analysis on the basis of Crimes happened in India.

IV. SYSTEM ARCHITECTURE

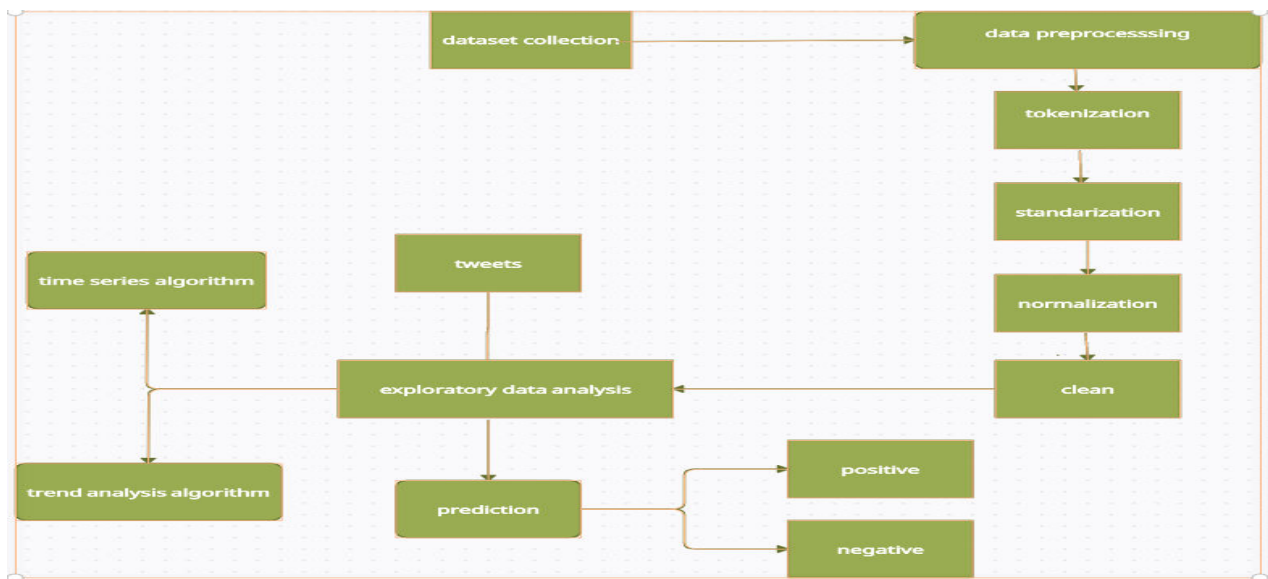


Fig (1.2) System Architecture

MODULES

MODULE 1:Dataset Collection

MODULE 2:Data Pre-processing

MODULE 3:Exploratory Data Analysis

MODULES 4:Execution Of Time Series And Trend Analysis Algorithm

MODULES 5:Prediction

MODULES DESCRIPTION

MODULE 1:

DATASET COLLECTION

Collecting the data, information, tweets and hash-tags from Twitter. First step involved in analysis of sentiment is the collection of information from the social network website like twitter. This process helps in extracting the tweet message and data like crime rates in India



MODULE 2:

DATA PRE-PROCESSING:

Once the data is extracted from the twitter as the datasets, this information has to be passed to the classifier. The classifier cleans the dataset by removing redundant data like **STOP WORDS**. Stop words are generally thought to be a “single set of words”. We would not want these words taking up space in our database. Using tokenization we can

divide a chunk of text into words, or we can divide it into sentences. Normalization- This improves the performance and training of the model.

MODULE 3:

EXPLORATORY DATA ANALYSIS:

Exploratory data analysis (EDA) is a term for certain kinds of initial analysis and findings done with data sets, usually early on in an analytical process. Some experts describe it as “taking a peek” at the data to understand more about what it represents and how to apply it. Exploratory data analysis is often a precursor to other kinds of work with statistics and data. In our system to visualize the data using EDA concept, which include bar chart, pie chart etc., To show dataset in different visualization like crime wise, district wise, State wise, Zone wise, Year wise.

MODULE 4:

EXECUTION OF TIME SERIES AND TREND ANALYSIS ALGORITHM

In this project we are using time series analysis algorithm that deals with time series data which means that data is in a series of particular time or interval. Then cross sectional data method and pooled data method is also used. Cross sectional data means collection of data at same time and pooled data is combination of both cross sectional and time series data. Here we also used trend series data analysis which is used to predict future movements based on recently observed trend data.

MODULE 5:

PREDICTION

To generate useful and meaningful information out of the raw data, sentimental analysis plays vital role. Once the algorithm is completed, the outcome of the analysis can be visualized by creating different types of graphs. Bar graphs, Time series and Pie charts are some of the examples which can be used to display the output. To measure the tweets in terms of Positive and Negative, Bar graphs can be used. Similarly, to measure in terms of likes, dislikes, average length of tweet for a certain period, Time series can be used. To obtain the initial source of the tweet, pie charts can be used.

IV. CONCLUSION

Throughout the research paper we have discussed about various machine learning algorithms that can help us to organize and analyze the huge amount of Twitter data obtained including millions of tweets and text messages shared every day. These machine learning algorithms are very effective and useful when it comes to analyzing of large amount of data including the Time series algorithm and Trend analysis Model approaches which help to further categorize the data into meaningful groups. EDA is yet another form of machine learning that is very popular for visualize the dataset information from the Crime report and get an idea about the status of women safety in Indian cities.



REFERENCES

1. Suvarna, Ashima, Grusha Bhalla, Shailender Kumar, and Ashi Bhardwaj. "Identifying Victim Blaming Language in Discussions about Sexual Assaults on Twitter" In International Conference on Social Media and Society, pp. 156-163. 2020.
2. Suvarna, Ashima, Grusha Bhalla, Shailender Kumar, and Ashi Bhardwaj. "Identifying Victim Blaming Language in Discussions about Sexual Assaults on Twitter." In International Conference on Social Media and Society, pp. 156-163. 2020.
3. Milusheva, Sveta, Robert Marty, Guadalupe Bedoya, Sarah Williams, Elizabeth Resor, and Arianna Legovini. "Applying machine learning and geolocation techniques to social media data (Twitter) to develop a resource for urban planning." Plos one 16, no. 2 (2021): e0244317.
4. Vanitha, G. "Security women through analysis of twitter messages."
5. Yadav, Bonthala Prabhanjan, M. Sheshikala, N. Swathi, Kanegonda Ravi Chythanya, and E. Sudarshan. "Women Wellbeing Assessment in Indian Metropolises Using Machine Learning models." In IOP Conference Series: Materials Science and Engineering, vol. 981, no. 2, p. 022042. IOP Publishing, 2020.
6. Oh, Onook, Manish Agrawal, and H. Raghav Rao. "Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises." MIS quarterly (2013): 407-426.
7. Shravya, Raparathi, and Dr P. Neelakantan. "Women Protection Analysis Based On Twitter Data Using ML." European Journal of Molecular & Clinical Medicine 7, no. 7(2021): 5820-5825.
8. Das, Priyanka, and Asit Kumar Das. "Crime analysis against women from online newspaper reports and an approach to apply it in dynamic environment." In 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), pp. 312-317. IEEE, 2017.
9. Swain, Sweta, and K. R. Seeja. "Analysis of Epidemic Outbreak in Delhi Using Social Media Data." In International Conference on Information, Communication and Computing Technology, pp. 25-34. Springer, Singapore, 2017.
10. Amusan, Lere, Adebowale Idowu Adeyeye, and Samuel Oyewole. "Women as Agents of Terror: Women Resources and Gender Discourse in Terrorist and Insurgency." Politikon 46, no. 3 (2019): 345-359.



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details