An Analysis on Woman Safety in Indian Cities

M V Nithya

Dept of Computer Science, SV University, Tirupati

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. Twitter and other Twitter handles which include hash tag messages that are widely spread across the whole globe sir as a platform for women to express their views about how they feel while we go out for work or travel in a public transport and what is the state of their mind when they are surrounded by unknown men and whether these women feel safe or not?*

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 neighbours 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 neighbours 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. Rather than imposing restrictions on women that society usually imposes it is the duty of society to imprecise the need of protection of women and also recognizes that women and girls also have a right same as men have to be safe in the City. Analysis of twitter texts collection also includes the name of people and name of women who stand up against sexual harassment and unethical behaviour of men in Indian cities which make them uncomfortable to walk freely. The data set that was obtained through Twitter about the status of women safety in Indian society was for the processed through machine learning algorithms for the purpose of smoothening the data by removing zero values and using Laplace and porter's theory is to developer method of analyzation of data and remove retweet and redundant data from the data set that is obtained so that a clear and original view of safety status of women in Indian society is obtained.

II. LITERATURE SURVEY

Hashtag Feminism and Twitter Activism in India Elizabeth Losh, University of California

Authors: Elizabeth Losh - 2018

The use of Twitter by activists protesting violence against women, particularly sexual violence, is complicated by the fact that microblogging services use hashtags to identify relevant content to their audiences. Activist communities congregate around these particular keywords, and archives that map the history and morphology of controversies in public discourse online depend

on shared terminology in the metadata. It is noteworthy that trending topics are particularly likely to reference proper names. After a prominent fatal sexual attack in India, the mobilization of activists through online organizing progressed through several stages, and some users privileged #delhirapecase or Delhi gangrape – which protected the private identity of the victim – while others made her into a public martyr by using her proper #jyotisinghpandey. Many also used a series of pseudonyms, such as #damini or #amanat. This study focuses on the informational labour of two specific activist groups in India Breakthrough and Blank Noise -- and how careful hashtag use reflected their policy decisions and deliberative activities about metadata management, which is becoming an increasingly important aspect of transformative social movements that bring citizens out into the streets.

Women Protection Analysis Based on Twitter Data Using ML

Authors: Raparthi shravya, Neelakantan - 2020

Girls and Women have been encountering a ton of savagery and badgering in broad daylight places in different urban communities beginning from following and prompting inappropriate behaviour or rape. This paper examines essential centres around the function of web-based media in advancing the security of ladies in different areas with exceptional reference to the part of online media sites and applications including Twitter stage Facebook and Instagram. This paper additionally focuses around how a feeling of obligation on part of culture can be built up the basic Indian individuals. Tweets on Twitter which typically contains pictures and text and furthermore composed messages and statements which centres around the security of ladies in different urban areas can be utilized to peruse a message among the Youth Culture and instruct individuals to make exacting move and rebuff the individuals who disturb the ladies. Twitter and other Twitter handles which incorporate hash label messages that are generally spread over the entire globe as a stage for ladies to communicate their perspectives about how they feel while we go out for work or travel in a public vehicle and what is the condition of their brain when they are encircled by obscure men and if these ladies have a sense of security? By analyzing the tweets polarity from the Twitter API. In Further improvements, we can use it in any Social Media Platform.

Women Wellbeing Assessment in Indian Metropolises Using Machine Learning models

Authors: Bonthala Prabhanjan Yadav, M. Sheshikala - 2020

In recent times, women working in many metropolises have experienced harassment and nuisance in civic places. In this regard, the research centres taking place the responsibility of social media in endorsing care of these sensitive cases in Indian metropolises with exceptional reference to Twitter, Facebook and Instagram platforms. This paper makes special efforts to understand how the sense of responsibility in Indian society can advance the safety of these women around us.

Identifying Victim Blaming Language in Discussions about Sexual Assaults on Twitter:

Authors: Ashima Suvarna, Grusha bhalla – 2020

Increasing instances of sexual assault have presented an opportunity for these heinous crimes to be discussed on social platforms. Oftentimes, victims are slut shamed and held culpable for the assault by the community which further discourages such personal disclosures and assault reporting. Victim Blaming has multiple psychological effects on the victim and further discourages formal reporting of such crimes. Therefore, it is important to devise computationally relevant methods to identify and prevent victim blaming to protect the victims. Additionally, specific datasets to devise models should also be developed. In our work, we present an exhaustive statistical analysis of victim blaming and gender attributes along with a single step transfer learning-based classification method to identify victim blaming language on Twitter. Finally, we compare the performance of the proposed model against various deep learning and machine learning models on a manually annotated domain-specific dataset.

Security Women Through Analysis of Twitter Messages

Authors: Vanitha G – 2020

Ladies and young ladies have been encountering a ton of brutality and badgering in broad daylight places in different urban areas beginning from following and prompting inappropriate behaviour or rape. This examination paper essentially centers around the job of online networking in advancing the wellbeing of ladies in Indian urban communities with unique reference

to the job of web-based life sites and applications including the Twitter stage Facebook and Instagram. This paper likewise centers around how an awareness of other's expectations on some portion of Indian culture can be built up the normal Indian individuals with the goal that we should concentrate on the security of ladies encompassing them. Tweets on Twitter which as a rule contain pictures and text and furthermore composed messages and statements which center on the wellbeing of ladies in Indian urban communities can be utilized to peruse a message among the Indian Youth Culture and instruct individuals to make exacting move and rebuff the individuals who irritate the ladies. Twitter and other Twitter handles which incorporate hashtag messages that are generally spread over the entire globe sir as a stage for ladies to communicate their perspectives about how they feel while we go out for work or travel out in the open vehicle and what is the condition of their psyche when they are encircled by obscure men and whether these ladies have a sense of security.

Problem Statement: 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.

Earlier System: In the recent years a large number of people have been attracted towards social media platforms like Facebook, Twitter and Instagram point and most of the people are using it to express their emotions and also their opinions about what they think about the Indian cities and Indian society.

- There is method of sentiment that can be categorized like lexicon-based learning**.**

Disadvantages

- Less Accurate
- Data analysis is not proper
- Prediction is inaccurate.

III. PROPOSED WORK

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.

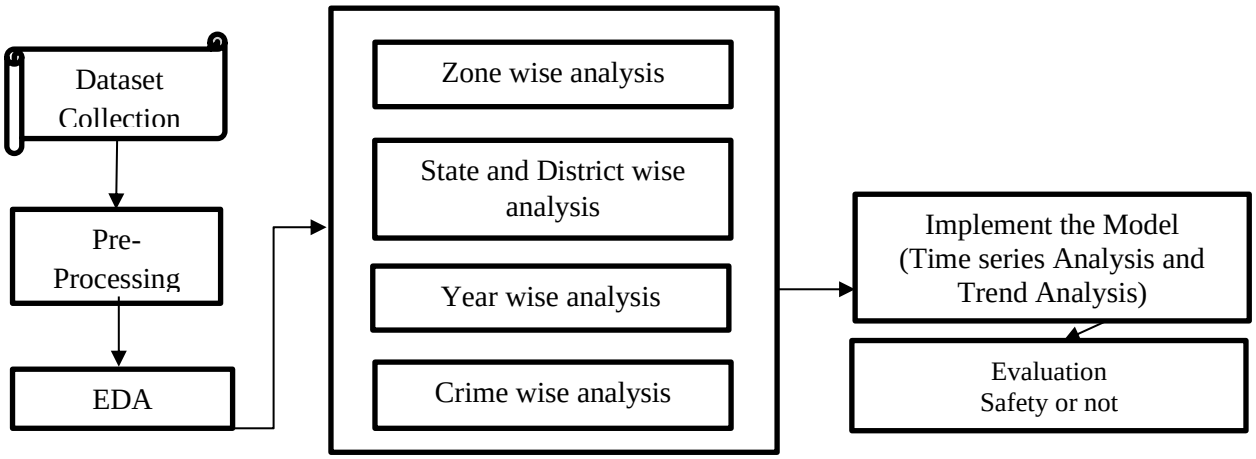
This paper also focuses on how a sense of responsibility on part of Indian society can be developed the common Indian people so that we should focus on the safety of women surrounding them. Crime details in India which usually contains images and text and also written messages and quotes which focus 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.

Advantages

- Accuracy is high when implementing the model.
- Crime report dataset are well handled in this system.
- Prediction of woman safety in India is well analyzed.

IV. SYSTEM ARCHITECTURE



Dataset Co1lection: First step involved in analysis of sentiment is the collection of information from the social network website like twitter. This helps in extracting the tweet message but this message also includes extra data like crime rates in India.

Unnamed: 0	STATE/UT	DISTRICT	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls	
0	0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	34	175	0
1	1	ANDHRA PRADESH	ANANTAPUR	2001	23	30	7	118	24	154	0
2	2	ANDHRA PRADESH	CHITTOOR	2001	27	34	14	112	83	186	0
3	3	ANDHRA PRADESH	CUDDAPAH	2001	20	20	17	126	38	57	0
4	4	ANDHRA PRADESH	EAST GODAVARI	2001	23	26	12	109	58	247	0
...	...	...	...	...	...	...	...	...	...	...	...
10672	832	Lakshadweep	Lakshadweep	2014	1	0	0	1	2	0	0
10673	833	Lakshadweep	Total District(s)	2014	1	0	0	1	2	0	0
10674	834	Puducherry	Karaikal	2014	3	1	0	12	1	1	0
10675	835	Puducherry	Puducherry	2014	7	6	1	20	7	3	0
10676	836	Puducherry	Total District(s)	2014	10	7	1	32	8	4	0

Data Pre-Processing: Once the data is extracted from the twitter source as the datasets, this information has to be passed to the classifier. The classifier cleans the dataset by removing redundant data like stop words, emoticons in order to make sure that non textual content is identified and removed before the analysis.

Text pre-processing is an essential a part of any NLP method and the significance of the NLP pre-processing are

To minimize indexing (or knowledge) records dimension of the textual content records

1. Stop words bills 20-30% of total phrase counts in a special textual content record
2. Stemming may just diminish indexing size as much as forty- 50%

To make stronger the efficiency and effectiveness of the IR method

1. Stop words aren't valuable for shopping or textual content mining
2. Stemming used for matching the similar words in a text record

The sklearn.preprocessing package provides several common utility functions and transformer classes to change raw feature vectors into a representation that is more suitable for the downstream estimators.

In general, learning algorithms benefit from standardization of the data set. If some outliers are present in the set, robust scalers or transformers are more appropriate. The behaviours of the different scalers, transformers, and normalizers on a dataset containing marginal outliers is highlighted in Compare the effect of different scalers on data with outliers.

Standardization, or Mean removal and Variance Scaling: Standardization of datasets is a **common requirement for many machine learning estimators** implemented in scikit-learn; they might behave badly if the individual features do not more or less look like standard normally distributed data: Gaussian with **zero mean and unit variance**.

Scaling features to a range- In practice we often ignore the shape of the distribution and just transform the data to center it by removing the mean value of each feature, then scale it by dividing non-constant features by their standard deviation.

For instance, many elements used in the objective function of a learning algorithm (such as the RBF kernel of Support Vector Machines or the l1 and l2 regularizes of linear models) assume that all features are centred around zero and have variance in the same order. If a feature has a variance that is orders of magnitude larger than others, it might dominate the objective function and make the estimator unable to learn from other features correctly as expected.

An alternative standardization is scaling features to lie between a given minimum and maximum value, often between zero and one, or so that the maximum absolute value of each feature is scaled to unit size. This can be achieved using MinMaxScaler or MaxAbsScaler, respectively.

The motivation to use this scaling include robustness to very small standard deviations of features and preserving zero entries in sparse data.

MaxAbsScaler works in a very similar fashion, but scales in a way that the training data lies within the range [-1,1] by dividing through the largest maximum value in each feature. It is meant for data that is already centred at zero or sparse data.

Normalization: Normalization is the process of **scaling individual samples to have unit norm**. This process can be useful if you plan to use a quadratic form such as the dot-product or any other kernel to quantify the similarity of any pair of samples. This assumption is the base of the Vector Space Model often used in text classification and clustering contexts.

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. To starts with, I imported necessary libraries (for this example pandas, NumPy, matplotlib and seaborn) and loaded the data set. 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

Implement the Model: To train and test the system using algorithm such that Time Series algorithm and Trend analysis algorithm

Time series Analysis Algorithm: Time series analysis is a statistical technique that deals with time series data, or trend analysis. Time series data means that data is in a series of particular time periods or intervals. The data is considered in three types:

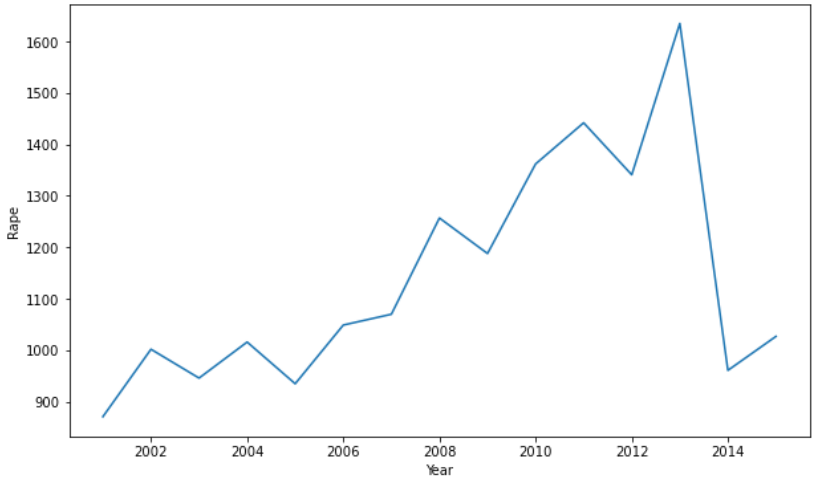
Time series data : A set of observations on the values that a variable takes at different times.

Cross-sectional data **:** Data of one or more variables, collected at the same point in time.

Pooled data **:** A combination of time series data and cross-sectional data.

Trend Analysis Algorithm: The trend is the component of a time series that represents variations of low frequency in a time series, the high and medium frequency fluctuations having been filtered out. The objective of this analysis is to understand if there is a trend in the data and whether this pattern is linear or not. The best tool for this job **is visualization.** Let's write a function that can help us to understand the trend and the movements of the time series. We want to see within the plot some rolling statistics such as:

- Moving Average: the unweighted mean of the previous *n* data (also called "rolling mean")
- Bollinger Bands: an upper band at k times an *n*-period standard deviation above the moving average, and a lower band at k times an *N*-period standard deviation below the moving average.



Evaluation: 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.

V. 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, Raparthi, 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 Terrorism and Insurgency." *Politikon* 46, no. 3 (2019): 345-359.