IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 9, Issue 6, Dec - Jan 2022
ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

Analysis of Women Safety in Indian Cities Using Machine Learning On Tweets

Mrs.S.Anitha, T.Pavani, CH.Alekhya, G.Dikshitha

Computer Science Engineering , JNTU-H , Hyderabad

ABSTRACT: Women and girls have been experiencing a lot of violence and harassment in public places in various cities starting from the 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 like twitter ,Facebook and Instagram. This paper also focuses on how 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 contain 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 among the Indian Youth Culture and educate people to take strict action and public those who harass the women. Twitter and 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.

Key words: hash tag, safety, sentimental analysis, women, sexual harassment.

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 neighbor 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.

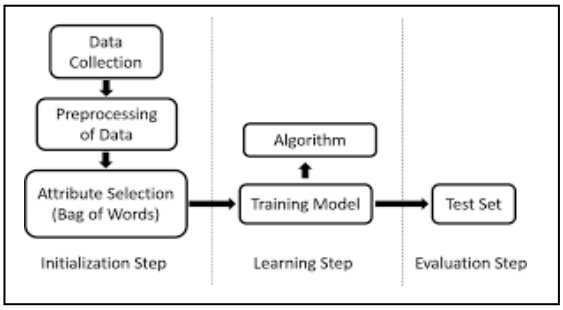
II. PROPOSED SYSTEM

Women have the right to the city which means that they can gofreely 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 isthe 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 abusely harassed by one of their own neighbor 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 abuse 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.

III. TWITTER ANALYSIS

As People communicate and share their opinion actively on social medias including Facebook and Twitter, Social network can be considered as a perfect platform to learn about people's opinion and sentiments regarding different events. There exists several opinion-oriented information gathering and analytics systems that aim to extract people's opinion regarding different topics. Since Twitter contains short texts, people tend to use different words and abbreviations. These phrases are difficult to extract their sentiment by current NLP systems easily. Therefore, many researchers have used deep learning and machine learning techniques to extract and mine the polarity of the phrases. As a large number of people have been attracted towards social media platforms like Facebook, Twitter and Instragram point and most of the people are using it to express their emotions and also their opinion about what they think about the Indian cities and Indian society. Using twitter analysis for business is kind of like getting a monthly twitter analytics report card. Twitter analytics complier all the behavior and action audiences take when they come across your posts or profile-clicks, follows, likes, expands and break down that data to help you track performance.

IV. ANALYSIS OF SENTIMENTS



The process of obtaining the sentiments of tweet includes five steps:

1**) Data extraction**: 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 tweets likes, dislikes and comments.

2) Text Cleaning: Once the data is extracted from the twitter source as the datasets, this information has to bepassed 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.

3) Sentiment Analysis: After the classifier cleans the dataset, the data is ready for the sentimental analysis process. Machine learning and Lexicon based learning and Hybrid learning are some of the approaches of sentimental analysis. There are also some other approaches such as Nero Linguistic Programming and Natural Language Processing. Training the dataset and then testing that trained dataset involves in machine learning approach. Training data and Testing data are useful for the classifier to perform the algorithm. Maximum Entropy, Naives Bayes classification, Bayesian Networks and Network Support Vector Machine are some of the algorithm which can be used to train the classifier. Testing data is used to identify the efficiency of the sentiment classifier.

In case of Lexicon based leaning, training dataset is not used. This approach uses a built-in dictionary in which words associated with sentiments of human are present. The third approach, which is the Hybrid learning, combines both machine leaning approach and lexicon learning approach in order to improve the performance of classifier.

4) Sentiment Classification: At this step, the dataset is ready for the classification. Each and every sentence of the tweet will be examined and opinion will be formed accordingly for subjectivity. Subjective expression sentences are retained and those of objective expression sentences are rejected. Techniques like Unigrams, Negation, Lemmas and so on are used at different levels of sentimental analysis. Sentiments can be distinguished broadly into two groups – Positive and Negative. At this point of sentimental analysis, each of the subjective sentences which will be retained are classified into good, bad or like, dislike or positive and negative.

5) Output Presentation: 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 sentiment of 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. System Architecture

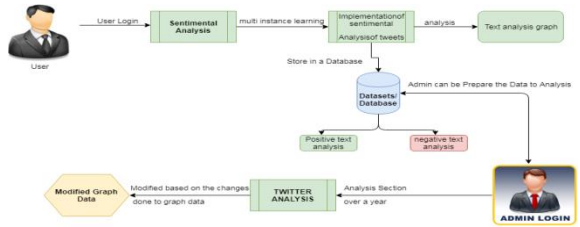


Fig -2: Architecture

Every user data such as credentials, new tweets, re-tweets and tweet score will be stored in the database for the admin to monitor and perform the analysis. The sentiment analysis is applied on the user data in order to monitor and confirm whether any tweets are abusive to women or not. Admin performs this analysis on each and every user tweets to provide safety for the women. Sentimental analysis will be implemented on the tweets of user that are stored in the database. Admin can now prepare the data to perform the analysis. The tweets made by every user of the application will be called as the initial input for the sentiment analysis and hence they will be the dataset. Along with this, text analysis graph can also be shown. Admin will store the filters in the database. Filters are the keywords for which the tweet context will be searched for in order to declare as abusive or not. There can be two types of filters – positive keyword and negative keyword. Positive keywords are those words which are abusive or disrespect the women by any means. Negative keywords are the words which are normal and will not abuse the women.

There can be 'n' number of positive and negative keywords stored in the database. When the admin implements the sentimental analysis, every keyword in the database will be compared with each and every word in the tweet of the user. If any one of the positive keyword is found in the tweet, that tweet will be classified as positive sentimental analysis and these are abusive to women. If negative keyword is found in the tweet, it will be classified as the negative sentimental analysis which is not abusive to women. Hence, by this stage there will be two types of sentimental analysis made based on the filter in the database. Under positive sentimental analysis, there will be a list of all the tweets in the application that are abusive to women. Similarly, under negative sentimental analysis there will be a list that is clean and are not abusive tweets. Along with the tweet context, user details will also be provided at each of the analysis list.

VI. CONCLUSION AND FUTURE WORK

Machine learning algorithm has been discussed throughout the project. For the twitter data that includes millions of tweet and messages every day, machine learning algorithm helps to organize and perform analysis. SPC algorithm, linear algebraic are some of the algorithms which are effective in analyzing the large data that provide categorization and convert into meaningful datasets. Hence we can perform machine learning algorithms to achieve sentimental analysis and bring more safety to women by spreading the awareness.

For the future enhancement, we can extend to apply these machine learning algorithms on different social media platforms like facebook and instagram also since in our project only twitter is considered. Present ideology which is proposed can be integrated with the twitter application interface to reach larger extent and apply sentimental analysis on millions of tweet to provide more safety.

VII. REFERENCES

[1] Apoorv Agarwal, Fadi Biadsy, and Kathleen R. Mckeown. "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams." Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2009.

[2] Luciano Barbosa and Junlan Feng. "Robust sentiment detection on twitter from biased and noisy data." Proceedings of the 23rd international conference on computational linguistics: posters. Association for Computational Linguistics, 2010.

[3] Adam Bermingham and Alan F. Smeaton. "Classifying sentiment in microblogs: is brevity an advantage?." Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010.

[4] Michael Gamon. "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[5] Soo-Min Kim and Eduard Hovy. "Determining the sentiment of opinions." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[6] Dan Klein and Christopher D. Manning. "Accurate unlexicalized parsing." Proceedings of the 41st Annual Meeting on Association for Computational LinguisticsVolume 1. Association for Computational Linguistics, 2003

[7] Eugene Charniak and Mark Johnson. "Coarse-to-fine nbest parsing and MaxEnt discriminative reranking." Proceedings of the