International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Predicting Safeness of Women in Indian Cities using Machine Learning

Nanditha P[1], Sindhu B M[2], R Geetha[3]

[1]*Department of Computer Science and Engineering, Cambridge Institution Of technology, Bangalore, India.*
[2]*Department of Computer Science and Engineering, Cambridge Institution Of technology, Bangalore, India.*
[3]*Assistent Professor, Department of Computer Science and Engineering, Cambridge Institution Of technology, Bangalore, India.*

Abstract - As we all know ladies and women area unit annoyed everyplace in each a part of the cities. ladies or lady within the country thinks about with characteristic. though she is worshiped and she or he plays several roles and however, she isn't safe. This base paper chiefly concentrate on the way to promote the characteristic for ladies' victimization social media platforms to market women safety. Here, we tend to chiefly concentrate on twitter platform. As we all know many ladies raise their problems within the social media platform like twitter that has hash tags wherever it reaches the message to everyone seems to be one in all the most effective platform were women tell their unsafe facing publicly transport, college, work space etc. we tend to predict the characteristic however ladies feel victimization twitter knowledge whether ladies feel unsafe.

***Key Words*:** Women, safety, Sentiment analysis.

1.INTRODUCTION

In India domestic abuse, sexual abuse and murder square measure common kinds of violence against ladies. Dowry death is associate final style of murder. Indians square measure still with the psychological science that dowry is tradition and girls' fathers lose everything to pay it. The list of crimes against ladies is kind of long, to mention the smallest amount. Acid attack is turning into terribly traditional is varied components of the country. There are a unit sure styles of harassment and Violence that area unit very aggressive together with staring and spending comments and these unacceptable practices area unit typically seen as a standard half of the urban life.

There are many studies that have been conducted in cities across Asian country and girls report similar type of molestation and spending off comments by different unknown individuals. The study that was conducted across most popular Metropolitan cities of Asian country together with metropolis, Mumbai and Pune, it had been shown that hour of the ladies feel unsafe while going intent on work or whereas move publicly transport. Women have the proper to the town which implies that they will go freely whenever they require whether or not it's too an academic Institute, or the other place ladies need to travel. But women feel that they're unsafe in places like malls, searching malls on their thanks to their job location thanks to the many unknown Eyes body shaming and harassing these ladies purpose Safety or lack of concrete consequences within the lifetime of girls is the main reason of harassment of women.

There square measure instances when the harassment of women was done by their neighbors while they were on the thanks to faculty or there was an absence of safety that created a way of worry within the minds of little ladies who throughout their time period suffer because of that one instance that happened in their lives wherever they were forced to try and do something unacceptable or was sexually harried by one in all their own neighbor or the other unknown person. Safest cities approach girl's safety from a perspective of women rights to the have an effect on the town without worrying of violence or sexual harassment. instead of imposing restrictions on women that society typically imposes it's the duty of society to imprecise the necessity of protection of ladies and additionally acknowledges that women and ladies even have a right same as men have to be compelled to be safe within the town.

Analysis of twitter texts assortment additionally includes the name of people and name of ladies World Health Organization get up against sexual harassment and unethical behavior of men in Indian cities which create them uncomfortable to steer freely. the info set that was obtained through Twitter regarding the standing of ladies safety in Indian society was for the processed through machine learning algorithms for the aim of smothering the info by removing zero values and exploitation Laplace and porter's theory is to developer methodology of analyzation information of knowledge of information and take away re-tweet and redundant data from the info set that's obtained thus that a transparent and original read of safety standing of ladies in Indian society is obtained.

2. LITERATURE REVIEW

People usually categorical their views freely on social media regarding what they feel regarding the Indian society and therefore the politicians that claim that Indian cities area unit safe for ladies. On social media websites individuals will freely categorical their read purpose and women will share their experiences wherever they need Janus-faced sexual harassment or wherever we might have fight back against the molestation that was obligatory on them.

The tweets regarding safety of ladies and stories of standing up against molestation more motivates different

girls' knowledge on an equivalent social media web site or application like Twitter. Other girls share these messages and tweets that more motivates different five men or ten girls to face up and lift a voice against those that have created Indian cities and unsafe men or ten girls to face up and lift a voice against people who have created Indian cities and unsafe place for the ladies. within the recent years an out sized variety of people are attracted towards social media platforms like Face book. it's a typical observe to extract the knowledge from the information that's obtainable on social networking through procedures of knowledge extraction, knowledge analysis and knowledge interpretation methods.

The accuracy of the Twitter analysis and prediction will be obtained by the utilization of behavioral analysis on the idea of social networks.

3. TWITTER ANALYSIS

As individuals communicate and share their opinion actively on social medias together with Facebook and Twitter, Social network can be thought-about as an ideal platform to be told concerning people's opinion and sentiments relating to totally different events. There exists several opinion-oriented operation and analytics systems that aim to extract people's opinion relating to different topics. Since Twitter contains short texts, individuals tend to use totally different words and abbreviations. These phrases are difficult to extract their sentiment by current information processing systems easily. Therefore, several researchers have used deep learning and machine learning techniques to extract and mine the polarity of the phrases.

4. EXISTING SYSTEM

Women have the proper to town which suggests that they can go freely whenever they need whether or not it's too an academic Institute, or the other place women wish to travel. however, girls feel that they're unsafe in places like malls, looking malls on their way to their job location owing to the many unknown Eyes body shaming and harassing these women purpose.

People typically categorical their views freely on social media regarding what they feel regarding the Indian society and the politicians that claim that Indian cities area unit safe for women. On social media websites folks will freely categorical their read purpose and girls will share their experiences wherever they need Janus-faced abuse harassment or wherever we might have fight back against the abuse harassment that was obligatory on them. The tweets regarding safety of girls and stories of standing up against abuse harassment more motivates alternative women information on constant social media web site or application like Twitter. alternative girls share these messages and tweets that more motivates alternative five men or ten girls to face up and lift a voice against people who have created Indian cities and unsafe place for the ladies. within the recent years an outsized variety of people are attracted towards social media platforms like Face book. it's a typical follow to extract the data from the information that's out there on social networking through procedures of information extraction, information analysis and information interpretation methods. The accuracy of the Twitter analysis and prediction is obtained by the utilization of activity analysis on the idea of social networks.

5. PROPOSED SYSTEM

Analysis of twitter texts collection also includes the name of people and name of women who stand up against sexual harassment and unethical behavior 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.

Women have the correct to town which implies that they will go freely whenever they require whether or not it be too an academic Institute, or the other place women need to travel. however, girls feel that they're unsafe in places like malls, searching malls on their way to their job location as a result of the many unknown Eyes body shaming and harassing these women purpose Safety or lack of concrete consequences in the lifetime of girls is that the main reason of harassment of girls. There are a unit instances once the harassment of girls was done by their neighbors whereas they were on the thanks to college or there was an absence of safety that created a way of worry within the minds of tiny women World Health Organization throughout their life suffer thanks to that one instance that happened in their lives wherever they were forced to try to one thing unacceptable or was abusely harassed by one among their own neighbor or the other unknown person. Safest cities approach girl's safety from a perspective of girl's rights to the have an effect on the city without concern of violence or abuse harassment. Rather than imposing restrictions on girls that society typically imposes it's the duty of society to imprecise the requirement of protection of girls and additionally recognizes that women and girls even have a right same as men got to be safe within the town.

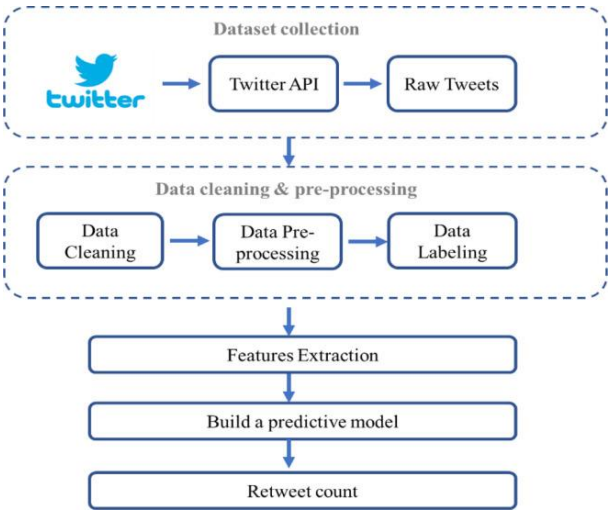

Fig -1: Process of analysis

6. SENTIMENTAL ANALYSIS

Sentiment analysis is that the method of extracting the sentiment behind any sentence or statement. It will be referred to as a classification technique that is employed to get the opinion from tweet. This opinion is beneficial in formulating a sentiment which may further be wont to accomplish sentiment classification. Sentiments area unit personal to the subject and so we'd like to decide what reasonably specifications is developed out of it. Person performing arts the sentimental analysis needs to search out the class of entities of the tweet's victimization the programming model. The dimension of the sentimental category is a very important issue in order to decide the potency of the rule. For instance, there will be 2 category sentimental classification of tweets – Positive and Negative or there will be 3 category classification – Positive, Negative and Neutral. Approaches of sentimental analysis will be generally differentiated into 2 types – machine learning primarily based} and lexicon learning based.

Machine learning approach includes the method of extraction of options, programming model coaching victimization dataset of options. Whereas lexicon learning based mostly approach uses the vocabulary and grading methodology to discover opinions. during this paper, we tend to use machine learning approach. Collection of information, pre-processing the information, extraction of features, selecting base options, detection of sentiments and classification of sentiments victimization machine learning approaches or straightforward computations area unit the fundamental steps to perform sentimental analysis.

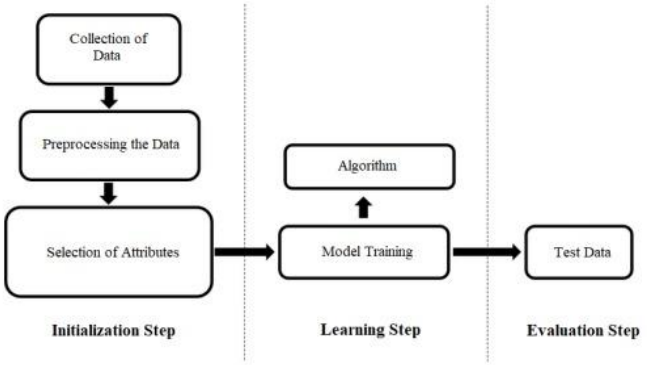

Fig -2: Process of Sentiment Analysis

7. SYSTEM SPECIFICATION

HARDWARE REQUIREMENTS:

System : Pentium IV 2.4 GHz.

Hard Disk : 40 GB.

Floppy Drive : 1.44 Mb.

Monitor : 14' Color Monitor.

Mouse : Optical Mouse.

Ram : 512 Mb.

SOFTWARE REQUIREMENTS:

Operating system : Windows 7 Ultimate.

Coding Language : Python. anaconda

Front-End : html

8. SYSTEM ARCHITECTURE

Every user knowledge like credentials, new tweets, re-tweets and tweet score are keep within the information for the admin to monitor and perform the analysis. The sentiment analysis is applied on the user knowledge so as to watch and ensure whether any tweets area unit abusive to ladies or not. Admin performs this analysis on every and each user tweets to provide safety for the ladies. Sentimental analysis is implemented on the tweets of user that area unit keep within the database. Admin will currently prepare the information to perform the analysis.

The tweets created by each user of the applying will be referred to as because the initial input for the sentiment analysis and thus they'll be the dataset. alongside this, text analysis graph also can be shown. Admin can store the filters in the information. Filters area unit the keywords that the tweet context is probe for so as to declare as abusive or not. There will be 2 forms of filters – positive keyword and negative keyword. Positive keywords area unit those words that are abusive or disrespect the ladies by any means that. Negative keywords area unit the words that area unit traditional and cannot abuse the ladies. There will be 'n' variety of positive and negative keywords stored within the information. once the admin implements the sentimental analysis, each keyword within the information are compared with every and each word within the tweet of the user.

If anybody of the positive keyword is found within the tweet, that tweet is classified as positive sentimental analysis and these area unit abusive to ladies. If negative keyword is found in the tweet, it'll be classified because the negative sentimental analysis that isn't abusive to ladies. Hence, by this stage there will be 2 forms of sentimental analysis created primarily based on the filter within the information. underneath positive sentimental analysis, there'll be an inventory of all the tweets within the application that area unit abusive to ladies. Similarly, under negative sentimental analysis there'll be an inventory that's clean and aren't abusive tweets. alongside the tweet context, user details also will be provided at every of the analysis list.

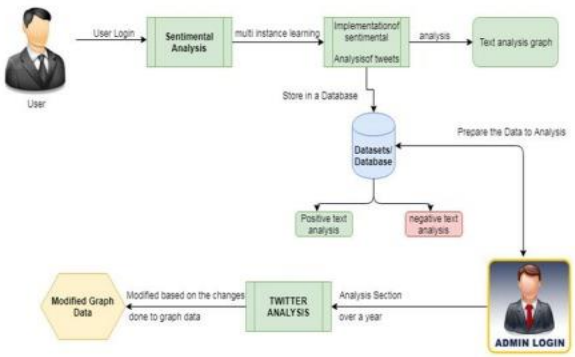

Fig -3: System Architecture

9. RESULT

In this work, we took Twitter API database as input database, after that Pre-process the data set (remove incomplete and noisy data) then apply feature extraction method as bagging of words and finally used Naïve Bayes classification. In this work, we used python language to develop the system. The following figure is the result of the given Twitter API dataset Tweets Accuracy Percentage.

Following shows the sample output of the program for tweets from Twitter.

Positive tweets percentage : 0.03095%
Negative tweets percentage : 0.74571%
Neutral tweets percentage : 0.2233293%

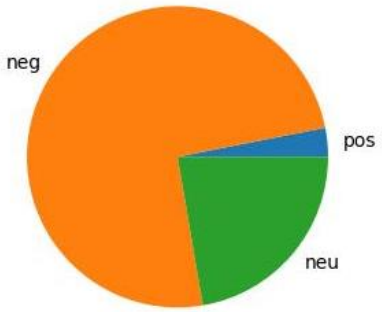

Fig -4: Result

10. CONCLUSIONS

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 tweets to provide more safety.

REFERENCES

[1] Agarwal, Apoorv, 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] Barbosa, Luciano, 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] Bermingham, Adam, 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] Gamon, Michael. "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] Kim, Soo-Min, and Eduard Hovy. "Determining the sentiment of opinions." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

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

[7] Charniak, Eugene, and Mark Johnson. "Coarse-to-fine n-best parsing and MaxEnt discriminative reranking." Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2005.

[8] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., & Badhani, P. (2017). Study of Twitter sentiment analysis using machine learning algorithms on Python. International Journal of Computer Applications, 165(9), 0975-8887.

[9] Sahayak, V., Shete, V., & Pathan, A. (2015). Sentiment analysis on twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.

[10] Mamgain, N., Mehta, E., Mittal, A., & Bhatt, G. (2016, March). Sentiment analysis of top colleges in India using Twitter data. In Computational Techniques in Information and Communication Technologies (ICCTICT), 2016 International Conference on (pp. 525-530). IEEE.