Nat. Volatiles & Essent. Oils, 2021; 8(6): 1289-1294

NVEO
Natural Volatiles &
Essential Oils

Applying Machine Learning Techniques To Analyze The Women Safety

N Kavitha [a], N Srinivasa Rao [b], Srinivasa Rao Madala[c]

[a] M.Tech., Scholar, Department of Computer Science and Engineering,

[b] Associate Professor, Department of Computer Science and Engineering,

[c] Professor & HOD, Department of Computer Science and Engineering

PACE Institute of Technology and Sciences

Abstract

There has been a lot of crime and intimidation against women and girls in public areas throughout the country, ranging from stalking to sexual harassment and assault. Social media websites and apps such as Twitter platform, Facebook, and Instagram have an important role in improving women's safety in Indian cities, as discussed in this study article. As a result of this article, the general Indian population will be encouraged to acquire a feeling of responsibility for the protection of women in their immediate surroundings. Tweets on women's safety in Indian cities on Twitter, which often include pictures and text, may be used to spread a message within Indian youth culture and educate people about the need to take severe action and punish harassers of women. The primary goal of this article is to develop a women's safety index that can be used to compare the safety of women in a few major cities, including Delhi, Mumbai, Pune, and Chennai. In many cities, women and girls are subjected to different forms of violence and harassment in public areas, ranging from stalking to sexual harassment and assault. All cities are ranked using various statistical methods that evaluate distinct factors. This mainly focuses on the impact of social media in enhancing the safety of women in Indian cities.

1.Introduction:

In the contemporary day, Twitter has emerged as the ultimate microblogging social network with over 100 million members and generates over 500 million 'Tweets' each day. Because Twitter has such a large audience, people have been drawn to share their opinions and judgments on every problem and topic on the internet. As a result, twitter serves as a valuable information resource for institutions, businesses, and organisations of all kinds. Individuals on twitter utilise the tweets area to express themselves. Due to the character limit of a tweet (140 characters), users are forced to be creative in order to include as much information as possible into their tweets. Polysemy and sarcasm are also common ways for individuals to communicate their thoughts. The unstructured nature of twitter discourse justifies the

name. The tweet's sentiment may be deduced from the message itself. Sentimental analysis is used to extract this data. These findings may be used in a variety of ways, such as gauging consumer sentiment toward a specific brand or product release, examining public opinion on government legislation, or gauging attitudes about women. Many studies have been conducted on the twitter data in order to do categorization and analysis of the results. In this article, we also examine several studies on machine learning and research on how to use sentimental analysis on Twitter data. The focus of the article is on machine learning algorithms and models. Staring and making remarks to women may be forms of aggression and harassment, and these behaviours, which are unethical, are common in metropolitan areas. Many studies in India indicate that women have experienced sexual harassment and other behaviours like those described above. According to such research, most women in well-known metropolises like Delhi, Pune, Chennai, and Mumbai feel uncomfortable when surrounded by strangers. People may openly share their views on Indian politics, society, and a wide range of other topics via social media. The same goes for women who have been victims of assault or sexual harassment, who may share their stories and unite others to fight back against such crimes. Twitter's analysis of tweets text collection yielded identities of harassers of women as well as names of women or innocent people who stood up to such violent actions or unethical behaviour by males, making it difficult for them to move about freely. For the purposes of training the machine learning algorithms and models, the tweet data set will be utilised. With this technique, the tweet data will be smoothed down by removing any zeroes. Using Laplace and Porter's theory, a technique is created to evaluate twitter data and eliminate unnecessary information from the collection. Twitter, Facebook, and Instagram have all garnered massive numbers of users. Text messages, emoticons, and hashtags on social media are common ways for people to voice their opinions on society, politics, and women, among other topics. Sentiment analysis techniques like machine learning and lexicon learning may both be classed as sentiment analysis.

2. Literature Review:

When it comes to Indian culture and politics, people aren't afraid to voice their opinions publicly on social media [1.]. People may openly express their views on social media platforms, and women can discuss their experiences of sexual harassment and how we would have retaliated if it had been forced on them[2]. Women's safety tweets and tales of standing up to sexual harassment on the same social media website or application like Twitter inspire other women.

Other women pass on these messages and tweets, inspiring five more men or ten more women to get up and speak out against those who have turned Indian cities into dangerous places for women. Social media platforms such as Facebook, Twitter, and Instagram have drawn a significant number of individuals in recent years, and most people use them to express their feelings and views on Indian cities and Indian culture. Machine learning hybrid and lexicon-based learning are two examples of sentiment methods. [5] There's still more division to consider. As an example, Janta offered three alternative approaches: statistical, knowledge-based, and based on age. Data extraction, data analysis, and data interpretation are typical practises for gleaning information from publicly accessible social networking data. Use of behavioural analysis based on social networks may improve Twitter analysis and forecast accuracy.

Social Media Analysis

People regularly interact and express their opinions on social media, such as Facebook and Twitter, thus the social network may be seen as a great platform for learning about people's views and emotions on various events. There are a number of opinion-based data collection and analysis tools out there that are designed to glean people's thoughts on various issues. People on Twitter use a variety of terms and acronyms because of the little space available. Current NLP algorithms have a hard difficulty parsing these words and figuring out what they mean. As a result, several academics have used deep learning and machine learning methods to extract and mine the polarity of the words.

3. System Analysis:

People often share their opinions on Indian culture and politics, claiming that Indian cities are safe for women on social media. Online communities such as Facebook and Twitter allow individuals to openly express their ideas and opinions, and women may tell their stories of times when they have been the victims of abuse or harassment and how they responded. Other women on the same social media website or application, such as Twitter, are further motivated by tweets about women's safety and tales of rising up against abuse harassment. Other women pass on these messages and tweets, inspiring five more men or ten more women to get up and speak out against the males who have turned Indian cities into dangerous places for the women who now live there. People have been more interested in social media sites like Facebook in recent years. Data extraction, data analysis, and data interpretation are typical practises for gleaning information from the social networking data. The accuracy of the Twitter analysis and prediction may be achieved by using behavioural analysis based on social networks. Twitter and Instagram have become popular platforms for expressing one's feelings and thoughts on Indian cities and society in general. Machine learning hybrid and lexicon-based learning are two examples of sentiment analysis methods. Additionally, Janta offered yet another classification, this time based on statistical, knowledge-based, and age-based differentiation methods. As a result, women have the freedom to travel anywhere they choose in the city, whether it to an Educational Institute or any other location they want. In areas like malls, shopping malls, and on their way to work, women worry about being harassed and shamed for their bodies by a variety of unseen individuals. For the most part, harassment of girls is motivated by concerns about women's safety or the absence of clear repercussions in their daily lives. Sometimes, girls are harassed by their neighbours on the way to school or because of a lack of safety, which instils fear in the minds of young girls. As a result, these girls suffer throughout their lives as a result of that one incident in which they were forced to do something unacceptable or were harassed by a stranger. Women in safest cities have the freedom to make decisions about the city without fear of violence or abuse. The responsibility of society is to imprecisely acknowledge the need for women's protection rather than place limits on their freedom, and to accept that women and girls have the same right as men to feel secure in the City. The collection of twitter messages contains the names of men and women in Indian cities who stand up to abuse, harassment, and unethical behaviour, making it difficult for them to move about freely. The statistics about women's safety in Indian society that was collected through Twitter.

4. Sentiment Evaluation

Sentiment analysis is the process of figuring out what a phrase or a statement really means. It's a method for classifying tweets that's used to determine the public's viewpoint. This point of view is helpful in forming a feeling, which may then be classified based on that emotion. Sentiments are unique to the subject, therefore we must determine what sort of requirements may be derived from them. The person tasked with emotional analysis wishes to discover the class of entities represented by the tweets in the programming model. The sentimental class's size has a significant bearing on the algorithm's performance. For example, tweets may be classified in two ways based on their sentiment: positively or negatively. They can also be classified in three ways: positively, negatively, or neutrally. Machine learning-based sentimental analysis and lexicon learning-based sentimental analysis are the two main approaches. When utilising machine learning, you'll use the extraction of features, as well as model training using a feature dataset. Vocabulary and score methods are used in the lexicon learning-based technique to identify views. We use a machine learning strategy in this study. Data collection, pre-processing, feature extraction, feature selection, sentiment detection, and sentiment classification using machine learning methods or simple calculations are the fundamental stages in sentimental analysis.
There are five stages involved in getting a tweet's sentiments:

4.1 Data collection: The first stage in sentiment analysis is to get data from social media sites like twitter. When you use this technique, you can get more information out of a tweet than you would otherwise be able to get from it alone.

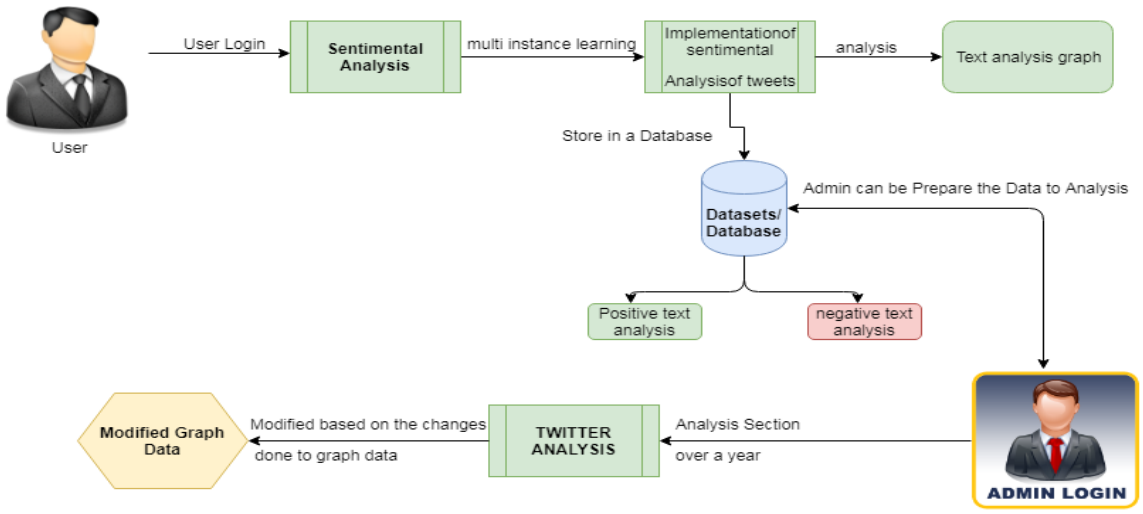
4.2 Data extraction: from the twitter source as datasets must be followed by text cleaning before the datasets can be given to the classifier. In order to identify and eliminate non-textual material before the analysis, the classifier cleans the dataset by eliminating redundant data such as stop words and emoticons.

4.3 Sentiment Analysis: The dataset is ready for sentimental analysis once the classifier has cleaned it. Sophisticated methods to emotional analysis include machine learning and Lexicon-based learning, as well as hybrid learning. Nero Linguistic Programming and Natural Language Processing are two more methods. The machine learning method includes both training and testing the dataset. The classifier may utilise the algorithm's output if it has access to sufficient Training and Testing data. Some of the algorithms that may be used to train the classifier are Maximum Entropy, Naive Bayes classification, Bayesian networks, and Network Support Vector Machine. An evaluation of the sentiment classifier's performance is conducted using test data. Lexicon-based learning does not make use of a training dataset. This method makes use of a built-in lexicon that includes terms linked with human emotions. In order to enhance the performance of a classifier, the Hybrid learning method combines machine learning with lexicon learning.

4.4 Sentiment Classification: The dataset is now ready for sentiment classification. Every tweet sentence will be scrutinised, and a personal opinion will be developed based on it. Sentences that represent ideas subjectively are accepted, whereas sentences that express ideas objectively are rejected. At various levels of emotional analysis, techniques like as Unigrams, Negation, Lemmas, and so on are used. Positive and negative feelings are the two main types of emotions. At this stage of

sentimental analysis, each of the preserved subjective phrases is categorised as excellent, bad, like, dislike, or positively and negatively.

4.5 Output Presentation: Sentimental analysis is essential for extracting valuable and meaningful information from large amounts of unstructured data. Once the algorithm has run its course, various kinds of graphs may be used to show the results of the study. Some examples of how the output may be shown include bar graphs, time series, and pie charts. Bar graphs may be used to gauge the overall mood of the tweets based on how positive or negative they are. Similarly, Time series may be used to track metrics like as likes, dislikes, and the average duration of a tweet over time. Pie charts may be used to find out where a tweet came from in the first place.



5. Conclusion:

Throughout the article, several deep learning and machine learning techniques have been explored that may assist in analysing large amounts of data collected through twitter to help evaluate the safety of women in society. For managing the massive amounts of data from social media sites, the machine learning algorithms employed are very effective and perform well on a variety of platforms. Using these techniques, we can have a significant impact on women's safety while also gathering data and generating new datasets to work with. We're excited to put in additional time and modify it so it's even more effective in the near future.

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.

[11] Madala, S. R., Rajavarman, V. N., & Vivek, T. V. S. (2018). Analysis of Different Pattern Evaluation Procedures for Big Data Visualization in Data Analysis. In Data Engineering and Intelligent Computing (pp. 453-461). Springer, Singapore.

[12] Madala, S. R., & Rajavarman, V. N. (2018). Efficient Outline Computation for Multi View Data Visualization on Big Data. International Journal of Pure and Applied Mathematics, 119(7), 745-755.

[13] Vivek, T. V. S., Rajavarman, V. N., & Madala, S. R. (2020). Advanced graphical-based security approach to handle hard AI problems based on visual security. International Journal of Intelligent Enterprise, 7(1-3), 250-266.