IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS OF WOMEN SAFETY IN INDIAN CITIES USING MACHINE LERNING ON TWEETS

[1]M.SUSHREETH,[2]K.AKSHAYA,[3]N.PRASANNA,[4]R.MAHENDAR,[5]V.NEELIMA,
[6]R JEGADEESAN
[1,2,3,4]FinalYear – Department of Computer Science and Engineering [5]Asso.Professor,[6]Professor Jyothishmathi Institute of Technology and Science Karimnagar, Telangana.

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 abuse harassment or abuse 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?

Key Words: Women Safety, Sexual Assault, Hash Tags,Sentimental Analysis, Tweets on Tweeter.

I.INTRODUCTION

Twitter in this modern era has emerged as a ultimate micro blogging social network consisting over hundred million users and generate over five hundred million messages known as 'Tweets' every day. Twitter with such a massive audience has magnetized users to emit their perspective and judgemental about every existing issue and topic of internet, therefore twitter is an informative source for all the zones like institutions, companies and organizations. On the twitter, users will share their opinions and perspective in the tweets section. This tweet can only contain 140 characters, thus making the users to compact their messages with the help of abbreviations, slang, shot forms, emoticons, etc. In addition to this, many people express their opinions by using polysemy and sarcasm also. On social media, people can freely express what they feel about the Indian politics, society and many other thoughts. Similarly, women can also share their experiences if they have faced any violence or sexual harassment and this brings innocent people together in order to stand up against such incidents. From the analysis of tweets text collection obtained by the twitter, it includes names of people who has harassed the women and also names of women or innocent people who have stood against such violent acts or unethical behaviour of men and thus making them uncomfortable to walk freely in public.

II.LITERATURE SURVEY

[1]Fadi Biadsy "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams."
We present a classifier to predict contextual polarity of subjective phrases in a sentence. Our approach features lexical scoring derived from the Dictionary of Affect in Language (DAL) and extended through WordNet, allowing us to automatically score the vast majority of words in our input avoiding the need for manual labeling.Our results show significant improvement over a majority class baseline as well as a more difficult baseline consisting of lexical n-grams.

[2]Barbosa "Robust sentiment detection on twitter from biased and noisy data." Proceedings of the 23rd international

conference on computational linguistics: posters.
We proposed an approach to automatically detect sentiments on Twitter messages (tweets) that explores some characteristics of how tweets are written and meta-information of the words that compose these messages.

1

[3]Bermingham "Classifying sentiment in microblogs: is brevity an advantage?."Proceedings of the 19th ACM international conference on Information and knowledge management.
Microblogs as a new textual domain offer a unique proposition for sentiment analysis. Their short document length suggests any sentiment they contain is compact and explicit. However, this short length coupled with their noisy nature can pose difficulties for standard machine learning document representations.

[4]Gamon, Michael. "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis."
He demonstrated that it is possible to perform automatic sentiment classification in the very noisy domain of customer feedback data. We show that by using large feature vectors in combination with feature reduction, we can train linear support vector machines that achieve high classification accuracy on data that present classification challenges even for a human annotator.

[5]Kim, Soo-Min "Determining the sentiment of opinions." Proceedings of the 20th international conference on Computational Linguistics.
Identifying sentiments (the affective parts of opinions) is a challenging problem. We present a system that, given a topic, automatically finds the people who hold opinions about that topic and the sentiment of each opinion.

III.EXISTING 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 are method of sentiment that can be categorized like lexicon-based learning.

DISADVANTAGES OF EXISTING SYSTEM

- 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 are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning.

IV.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 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 abusely harassed by one of their own neighbour 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.

ADVANTAGES OF PROPOSED SYSTEM

- Analysis of twitter texts collection also includes the name of people and name of women who stand up against abuse harassment and unethical behaviour of men in Indian cities which make them uncomfortable to walk freely.
- The data setthat was obtained through Twitter about the status of women safety in Indian society

2

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

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.

Data Cleaning:
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.

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.

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.

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.

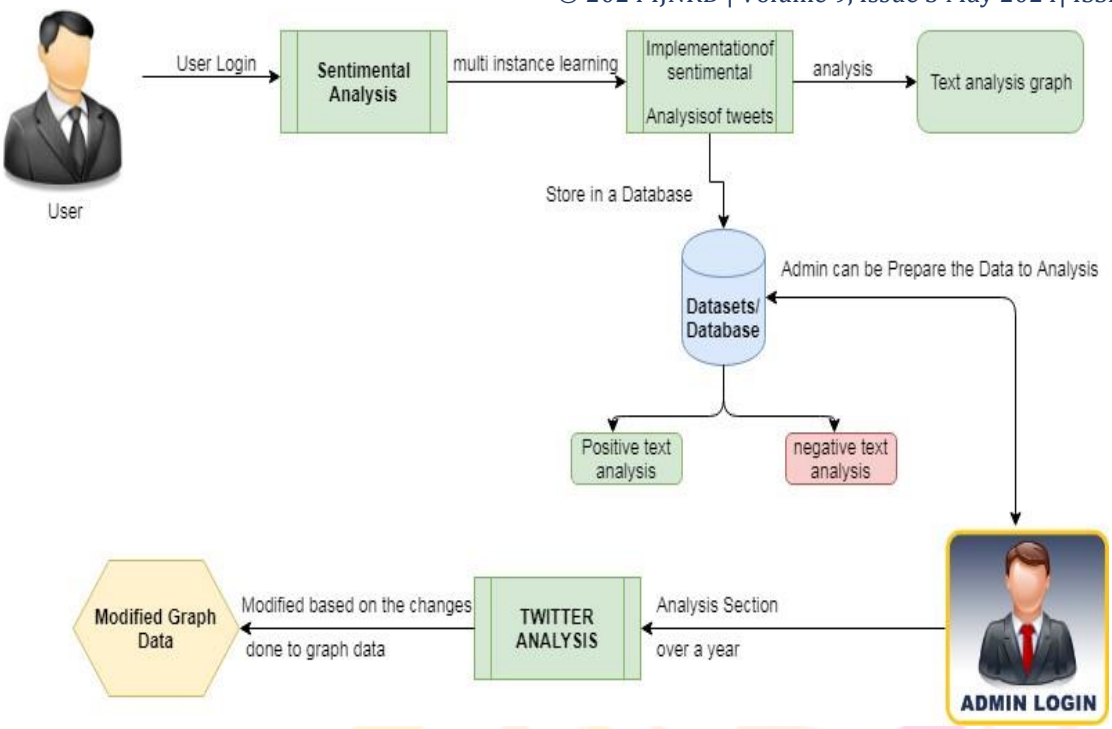

Fig:Architecture

Text Preprocessing

As text is the least structured of all the data kinds, it contains a variety of noise and cannot be easily analyzed without preprocessing.

Text pre-processing refers to the full procedureof standardizing and cleaning text to remove noise and prepare it for analysis.

It typically consists of three steps:
1. Lexicon normalization.
2. Noise reduction.
3. Object standardization.

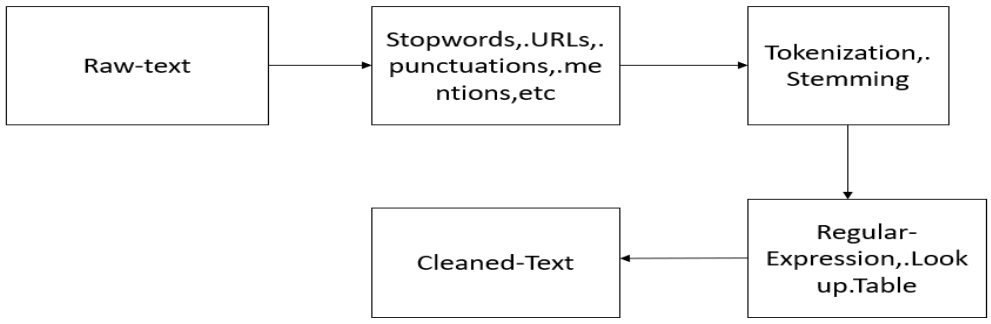

FIG:Text Preprocessing

4

V.RESULT

Staring at women and passing comments can be certain types of violence and harassments and these practices, which are unacceptable, are usually normal especially on the part of urbanlife. Many researches that have been conducted in India shows that women have reported sexual harassment and other practices as stated above. Such studies have also shown that in popular metropolitan cities like

Delhi, Pune, Chennai and Mumbai, most women feel they are unsafe when surrounded by unknown people. On social media, people can freely express what they feel about the Indian politics, society and many other thoughts. Similarly, women can also share their experiences if they have faced any violence or sexual harassment and this brings innocent people together in order to stand up against such incidents. From the analysis of tweets text collection obtained by the twitter, it includes names of people who has harassed the women and also names of women or innocent people who have stood against such violent acts or unethical behaviour of men and thus making them uncomfortable to walk freely in public.

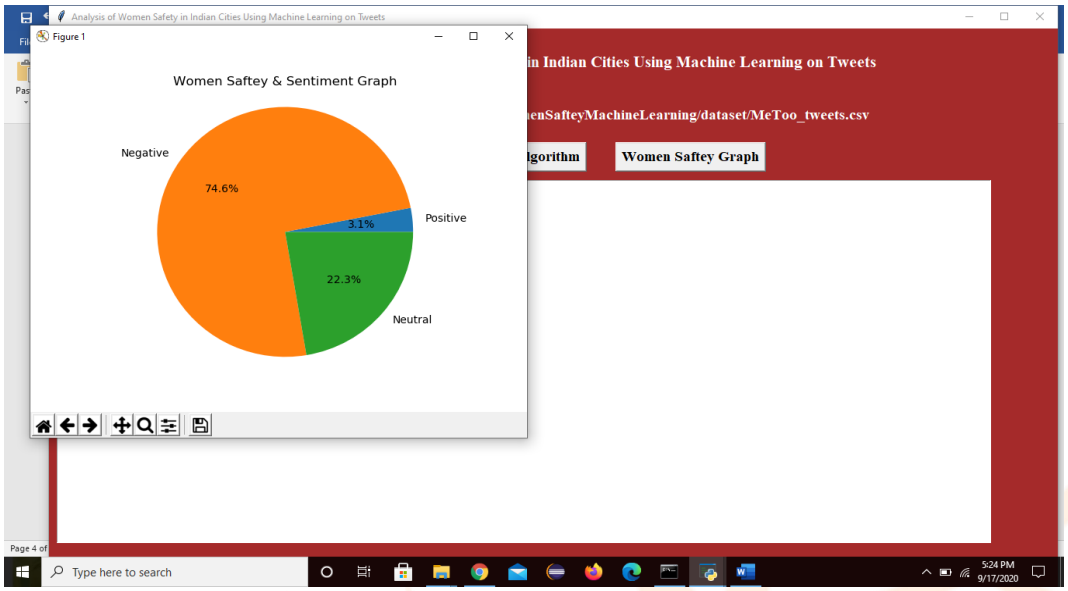

Fig:Women safety and sentimental graph

VI.CONCLUSION

The different machine learning techniques that can help us organize and analyze the enormous amount of Twitter data acquired, including the millions of tweets and text messages posted every day, have been discussed. The SPC method and linear algebraic Factor Model techniques, which help to further categorize the data into meaningful groupings, are two machine learning algorithms that are particularly successful and useful when it comes to evaluating enormous amounts of data. Another machine learning algorithm known as support vector machines is highly popular for extracting useful data from Twitter and gaining insight into the status of women's safety in Indian cities. We coclude that 74% people are talking negative and area is not safe and only 22 and 3% peoples are talking positive and neutral.

REFERENCES

[1]Agarwal, Apoorv, Fadi Baidsy andKathleenR.Mckeown.“Contextual phrase -level polarity analysis using lexicalaffect scoring and syntactic n-grams.” proceedings of the12th European chapter of the association forcomputational linguistics, associations forcomputational linguistice,2009.Volume: 10 Issue: 02 | Feb 2023 www.irjet.net p-ISSN: 2395-0072International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056© 2023, IRJET | Impact Factor value: 8.226 | ISO 9001:2008 Certified Journal |

[2] Barbosa Lucianoand Junla Feng. “Robust sentiment detection on twitter from biased and noisy data.”Proceedings of the 23rd international conference oncomputational linguistics: posters. associations forcomputational linguistic, 2010.

[3] Bemingham, Adam, and Alan F. Smeaton. “Classifying sentiment in micro blogs: is brevity and advantage?”proceedings of the 19th ACM international conferenceon information and knowledge management ACM, 2010.

[4] Gamon, Michael. “Sentiment classification on customer Facebook data: noisy data, large feature vectors, and the role of linguistic analysis:” proceedings of the 20th international conference on computational linguistics association from computational linguistics,2004.

[5] Kim, Soo-min, and Eduard hovy. “Determining the s of options.” proceeding of the 20th international conference on computational linguistics Associations from computational linguistics, 2004.

[6] Keindan, and Christopher D. Manning, "Accurate Unlexicalized parsing." proceedings of the 41st annual meeting on association f or computational linguisticsvolume1. Association from computational linguistics,2003.

[7] Charniak, Eugene, and mark Johnson."Coarse-to-fine nbest parsing and maxent discriminative re-ranking".proceedings of the 43rd annual meeting onAssociationsfor computational linguistics. Associations forcomputational linguistics,2005.

[8] Gupta B., negi M., Vishwakarma., 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] R Jegadeesan, Durga Bhavani, Dava Srinivas, 2023, "A Robust Approach for Effective Spam Detection using Supervised Learning Techniques", Journal of Xi'an University of Architecture & Technology, Volume XV, Issue 10, ISSN No : 1006-7930, Page No: 13-25, https://doi.org/10.37896/JXAT15.10/32639 (UGC-SCOPUS)

[11] R. Jegadeesan1, N. Sankarram2, C. Bagath Basha3 , K. Vijay4 , R. Jaichandran5 , and P. Nancy6, 2023, Forecasting of origin-to-destination requests for taxis using DNN algorithm with NYU database, AIP Conf. Proc. 2587, 050008-1–050008-13; https://doi.org/10.1063/5.0150597 Published by AIP Publishing. 978-0-7354-4710-3/$30.00, 23.11.23,

[12] Madhavarapu Prathima Rao1 and R Jegadeesan2, 2023, "Electricity Theft Detection In Smart Gridsbased On Deep Neural Networks", Journal of Xi'an University of Architecture & Technology, ISSN No : 1006-7930, Volume XV, Issue 11, 2023, Page No: 175-181

[13] R Jegadeesan, Rajitha Nelli, 2024, A Forecasting and Classification of Attacks used on Machine Learning, Journal of Xi'an University of Architecture & Technology, ISSN No : 1006-7930, Volume XVI, Issue 1, 2024, Page No: 469-477, https://doi.org/10.37896/JXAT16.1/33137

[14] R Jegadeesan, Velmula Shwetha, 2024, Predicting Air Pollution Using Algorithms from Machine Learning, Strad Research, ISSN: 0039-2049, VOLUME 11, ISSUE 3, Page No. 1-10, https://doi.org/10.37896/sr11.3/001

[15] Mamgain, N., Mehta,E., Mittal, A., &Bhatt, G.(2016,march).sentiment analysis of top colleges India using twitter data. in computational techniques in information and communication Technologies (ICCDICT),2026 international conference on (pp.525-530). IEEE