ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10, Issu 10, 2021

ANALYZING TWEETS USING A MACHINE LEARNING APPROACH

[1]*Bathula Srikanth,*[2]*Sanga Ravikiran,*[3]*Thambala Ramesh,*[4]*Badisa Vineesha*
[123]*Assistant Professor,*[4]*Student*
Department of CSE Engineering
Abdul Kalam Institute of Technological Sciences, Kothagudem,Telangana

ABSTRACT

In public spaces in several locations, women and girls have been subjected to a great deal of violence and harassment. It often begins with stalking and escalates into abuse harassment or abuse assault. This study primarily examines how social media, namely the websites and apps associated with Twitter, Facebook, Instagram, and other platforms, contributes to women's safety in Indian cities. This essay also discusses how Indian society might help the ordinary Indian people create a feeling of responsibility, leading us to prioritize the protection of the women who are around them. One way to convey a message among Indian youth culture and teach people to take rigorous action and punish those who harass women is via Twitter, where tweets about the safety of women in Indian cities are often composed of photos, text, written words, and quotations. Twitter and other Twitter accounts, which contain hash tag messages that are extensively disseminated worldwide, sir, serve as a forum for women to voice their opinions about how they feel when they travel in public transportation or go out for work. What goes through these women's minds when they are surrounded by unknown men, and do these women feel safe or not?

1. INTRODUCTION

Twitter in this modern era has emerged as a ultimate microblogging 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. Hence twitter language can be termed as the unstructured. From the tweet, the sentiment behind the message is extracted. This extraction is done by using the sentimental analysis procedure. Results of the sentimental analysis can be used in many areas like sentiments regarding a particular brand or release of a product,

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10, Issu 10, 2021

analyzing public opinions on the government policies, people thoughts on women, etc. In order to perform classification of tweets and analyze the outcome, a lot of study has been done on the data obtained by the twitter. We also review some studies on machine learning in this paper and research on how to perform sentimental analysis using that domain on twitter data. The paper scope is restricted to machine learning algorithm and models.

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 urban life. 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.

The data set of the tweet will be used to process the machine learning algorithms and models. This algorithm will perform smoothening the tweet data by eliminating zero values. Using Laplace and porter's theory, a method is developed in order to analyze the tweet data and remove redundant information from the data set. Huge numbers of people have been attracted to social media platform such as Twitter, Facebook, Instagram. People express their sentiments about society, politics, women, etc via the text messages, emoticons and hash-tags through such platforms. There are some methods of sentiment that can be classified like machine leaning based and lexicon based learning.

2. LITERATURE SURVEY

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.

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. We augment lexical scoring with n-gram analysis to capture the effect of context. We

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10, Issu 10, 2021

combine DAL scores with syntactic constituents and then extract ngrams of constituents from all sentences. We also use the polarity of all syntactic constituents within the sentence as features. Our results show significant improvement over a majority class baseline as well as a more difficult baseline consisting of lexical n-grams.

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.

In this paper, we propose 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. Moreover, we leverage sources of noisy labels as our training data. These noisy labels were provided by a few sentiment detection websites over twitter data. In our experiments, we show that since our features are able to capture a more abstract representation of tweets, our solution is more effective than previous ones and also more robust regarding biased and noisy data, which is the kind of data provided by these sources.

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.

In today's world, opinions and reviews accessible to us are one of the most critical factors in formulating our views and influencing the success of a brand, product or service. With the advent and growth of social media in the world, stakeholders often take to expressing their opinions on popular social media, namely Twitter. While Twitter data is extremely informative, it presents a challenge for analysis because of its humongous and disorganized nature. This paper is a thorough effort to dive into the novel domain of performing sentiment analysis of people's opinions regarding top colleges in India. Besides taking additional preprocessing measures like the expansion of net lingo and removal of duplicate tweets, a probabilistic model based on Bayes' theorem was used for spelling correction, which is overlooked in other research studies. This paper also highlights a comparison between the results obtained by exploiting the following machine learning algorithms: Naïve Bayes and Support Vector Machine and an Artificial Neural Network model: Multilayer Perceptron. Furthermore, a contrast has been presented between four different kernels of SVM: RBF, linear, polynomial and sigmoid.

3. EXISTING SYSTEM:

People often express their views freely on social media about what they feel about the

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10, Issu 10, 2021

Indian society and the politicians that claim that Indian cities are safe for women. On social media websites people can freely Express their view point and women can share their experiences where they have faced abuse harassment or where we would have fight back against the abuse harassment that was imposed on them . The tweets about safety of women and stories of standing up against abuse harassment further motivates other women data on the same social media website or application like Twitter. Other women share these messages and tweets which further motivates other 5 men or 10 women to stand up and raise a voice against people who have made Indian cities and unsafe place for the women. In the recent years a large number of people have been attracted towards social media platforms like Facebook, . It is a common practice to extract the information from the data that is available on social networking through procedures of data extraction, data analysis and data interpretation methods. The accuracy of the Twitter analysis and prediction can be obtained by the use of behavioral analysis on the basis of social networks.

DISADVANTAGES:

1. 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.
2. There are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning.
3. Also there are another categorization Janta presented with categories of statistical, knowledge-based and age wise differentiation approaches

4. PROPOSED SYSTEM:

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 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

ISSN PRINT 2319 1775 Online 2320 7876

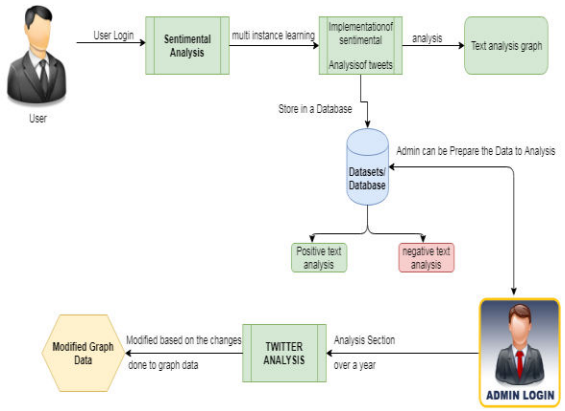
 UGC CARE Listed (Group -I) Journal Volume 10, Issu 10, 2021

right same as men have to be safe in the City.

ADVANTAGES:

1. 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.
2. The data set that was obtained through Twitter about the status of women safety in Indian society

5. ARCHITECTURE DIAGRAM



6. IMPLEMENTATION

MODULES:

TWITTER ANALYSIS

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.

IMPLEMENTATION OF SENTIMENTAL ANALYSIS OF TWEETS

Report the tweets picked up from Twitter API provided by Twitter itself. Due to the presence of Twitter API, there are many techniques available for sentimental analysis of data on Social media. In this project a set of available libraries has been used.

GRAPH

A Depressed interaction graph *G_* is generated via some social graph model, minimizing the distance between the real and Depressed interaction graphs.An *interaction graph G* is extracted from the input (real) social media data. An interaction graph represents how social network actors interact with each other [25], [26]. Entities and their interactions in social media are identified, and an interaction graph is built with a vertex set *V* , including entities, an edge set *E* representing interactions, and an attribute set *A*, which includes both vertex (entity) attributes and edge (interaction) attributes

Final Report

If the neutral tweets are significantly high, means that people have a lower interest in the topic and are not willing to haves a positive/negative side on it. This is also important to mention that depends on the data of the experiment we may get different results as people's opinion may change depending on the circumstances for example rape news it becomes the most

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10, Issu 10, 2021

trending news of the year in 2017. For some queries, the neutral tweets are more than 60% which clearly shows the limitation of the views. By above analysis that we have done, it an be clearly stated that Chennai is the safest city whereas Delhi is the unsafe city.

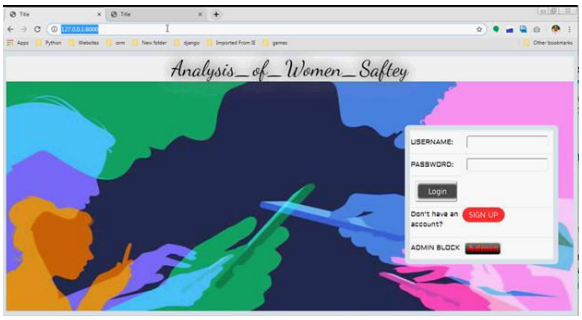
ALGORITHM:

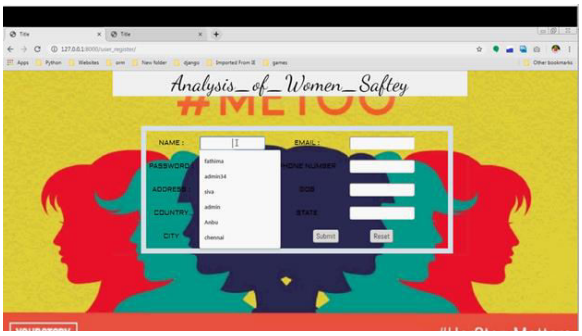
SUPPORT VECTOR MACHINE

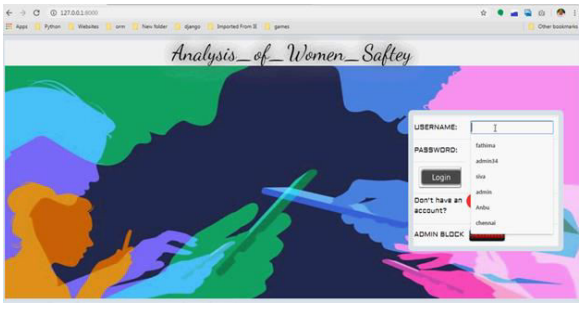
"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot). Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line). More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-

dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space.

7. SCREENSHOTS

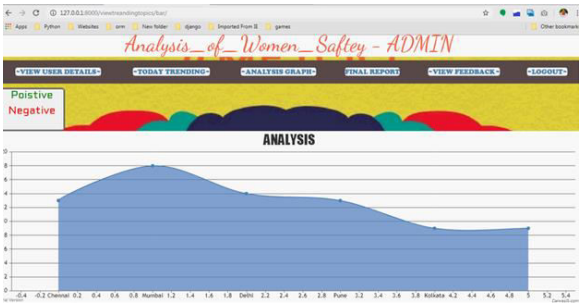


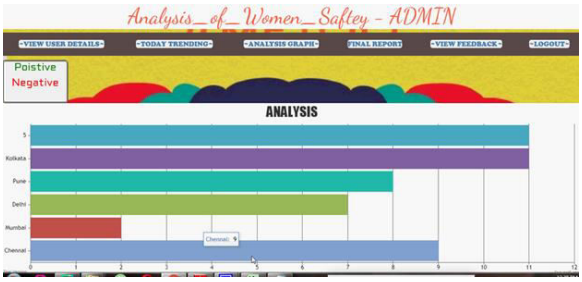


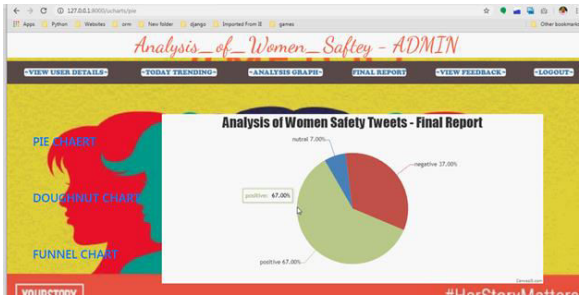


ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10, Issu 10, 2021







8. CONCLUSION

We have covered a number of machine learning algorithms in this research article that may assist us in sorting through and analyzing the massive quantity of Twitter data that we have collected, which includes millions of tweets and text messages sent on a daily basis. When it comes to evaluating vast amounts of data, certain machine learning algorithms such as the SPC algorithm and linear algebraic Factor Model approaches are very successful and helpful in further classifying the data into meaningful categories. Another well-liked kind of machine learning technique for gathering useful data from Twitter and gaining insight into the state of women's safety in Indian cities is support vector machines.

FUTURE ENHANCEMENT

Since just Twitter is taken into consideration in our experiment, we may expand the application of these machine learning algorithms in the future to include other social media sites like Facebook and Instagram. The suggested philosophy can be more fully included into the Twitter application interface, allowing sentiment analysis to be applied to millions of tweets and increasing safety.

REFERENCES

[1] Apoorv Agarwal, Fadi Biadsy, and Kathleen R. Mckeown. "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams." Proceedings

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 10, Issu 10, 2021

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 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.