International Journal of Research in Engineering, Science and Management
Volume 6, Issue 2, February 2023
https://www.ijresm.com | ISSN (Online): 2581-5792

Analysis of Women Safety in Public Places Using Machine Learning on Tweets

Farha Parveen[1*], Rahil Shaik[2], Shaik Raheem[3]
[1,2,3]*Student, Department of Computer Science and Engineering, Nalla Narasimha Reddy Education Society Group of Institutions, Hyderabad, India*

***Abstract*: Now-a-days women are experiencing lots of violence such as harassment in places in several cities. This starts from stalking which then leads to abusive harassment or also called abuse assault. This research paper basically focuses on the role of social media in promoting the safety of women in several cities with special reference to the role of social media websites and applications including Twitter platform Facebook and Instagram. Tweets on Twitter which typically contains pictures and text and furthermore composed messages and statements which centres around the security of ladies in different urban areas can be utilized to peruse a message among the Youth Culture and instruct individuals to make exacting move and rebuff the individuals to make exacting moves and rebuff the individuals who disturb the ladies. Applying various machine learning algorithms on tweets, and performing analysis on them to classify them into neutral, negative, and positive can help us improve the situation.**

***Keywords*: Social media, harassment, abuse assault, women.**

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. Sentiment analysis can be performed over a variety of tweets to perform analysis, a variety of topics such as product evaluations, movie reviews, and so on can be understood better using machine learning. 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. The goal here is to use various machine learning algorithms to analyse the tweets and classify them into positive, neutral and negative. 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.

Classifying the tweets will help us understand the opinion of public regarding various issues. Women's safety is arising issue all over the world. Twitter is a great platform that can provide suitable data for us to analyse the safety levels and the steps being taken to increase the safety levels of women. Performing a twitter sentiment analysis on tweets of tweets related to women can help us understand the society in which we all are living. This understanding can further help us improve the environment for the women of our world. The tweets that are extracted from twitter can contain several words and text which is not related to the content and context that we seek. Twitter is a platform where people express their feeling in single lines and emojis. The tweets directly obtained from twitter cannot be used directly since they are considered as noisy data. While performing analysis not only emojis but also hashtags and white spaces should be removed to get better accuracy in the results. The various punctuations that are used should be removed to make the data nose-free.

2. Sentiment Analysis

Sentiment analysis is the process of extracting the sentiment behind any sentence or statement. It can be called as a classification technique which is used to obtain the opinion from the tweet.

This opinion is useful in formulating a sentiment which can further be used for achieving sentiment classification. The dimension of the sentiment class is an important factor in order to decide the accuracy or efficiency of an algorithm. For instance, there can be three class sentimental classification-positive, negative, neutral.

A. Analysis of Sentiment Data

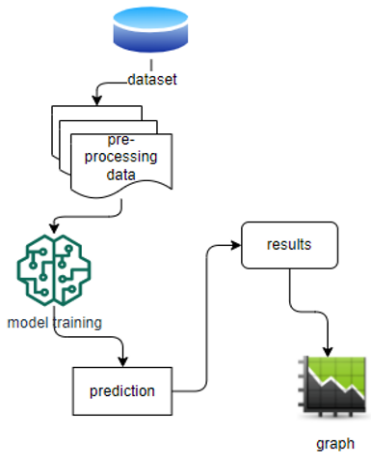

Fig. 1. Architecture

*Corresponding author: farhaparveen606@gmail.com

The process of obtaining the sentiments of tweets include 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 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.

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 Representation

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.

3. Conclusion and Future Work

Machine learning algorithm has been discussed throughout the project. For the twitter data that includes a huge number of tweets and messages consistently, Machine Learning algorithms helps to organize and perform analysis. SVM, Random Forest, NB and Logistic Regression are some of the algorithms which are effective in analysing 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, these machine learning algorithms can be reached out to be utilized on various virtual entertainment platforms, for example, Facebook and Instagram additionally 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] Green William H. (2012). Economic Analysis (Seventh ed.). Boston: Pearson Education. pp. 803-806. ISBN 978-0-273-75356-8.

[2] R. Plutchick. “Emotions: A general psycho evolutionary theory.” In K.R. Scherer & P. Ekman (Eds) Approaches to emotion. Hillsdale, NJ; Lawrence Earlbaum Associates, 1984.

[3] Russell, Stuart; Norvig, Peter (2003) [1995]. Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall. ISBN 978-0137903955.

[4] Bermingham, Adam, and Alan F. Smeaton. "Classifying sentiment in microblogs: is brevity an advantage?", ACM, 2010.

[5] 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.

[6] Charniak, Eugene, and Mark Johnson. "Coarse-tofine n-best parsing and MaxEnd discriminative reranking". Association for Computational Linguistics, 2005.

[7] Sahayak, V., Shete, V., & Pathan, A. Sentiment analysis on twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178183, 2005.

[8] Klein, Dan, and Christopher D. Manning. "Accurate unlexicalized parsing”, Association for Computational Linguistics, 2003.

[9] 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.

[10] 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.