Vol 12 Issue 03 2023

ISSN NO: 2230-5807

Unveiling Women Safety: A Twitter Data Analysis

Dr. A. Rama Swamy Reddy[1], Dr. Patnala S. R. Chandra Murty[2], Sathwika Mikkili[3]
[1]Professor, Department of CSE, Malla Reddy Engineering College, Hyderabad, Telangana, India.
[2]Professor, Department of CSE, Malla Reddy Engineering College, Hyderabad, Telangana, India.
[3]PG Scholar, Department of CSE, Malla Reddy Engineering College, Hyderabad,Telangana,India.

ABSTRACT

Stalking is defined as following and harassing a woman or girl in public, which may escalate to verbal or physical assault. This research looks at how Twitter, Facebook, and Instagram might help women feel safer in India's major cities. This research also looks at methods to assist ordinary Indians to develop a sense of social responsibility, which is critical for safeguarding Indian women in their everyday lives. Tweets, which generally feature photographs and text, might be utilized to educate Indian youth culture about the need of taking immediate action against female harassers in Indian cities. Twitter and other Twitter handles that include hash tag messages that are widely distributed around the world, sir, as a platform for women to express their views about how they feel while going to work or traveling in public transportation, how they feel when surrounded by unknown men, and whether they feel safe. These messages are being broadcast all throughout the planet.

Keywords: Women sefety, tweets, sentimental, Ml.

INTRODUCTION

Users of Twitter will tweet their responses. under keep tweets under 140 characters, users must use acronyms, slang, shot forms, emoticons, and other tactics. Many people use sarcasm and polysemy to convey their point. As a result, Twitter language is unstructured. The significance of the tweet is revealed by studying its wording. Sentiment analysis collects critical data. Sentiment analysis may be used to investigate public opinion on government policies, women's perspectives, and other topics. Twitter data has been thoroughly examined in order to categorize tweets and measure consequences. This paper also examines works on machine learning and emotional analysis of Twitter data. This article just introduces a few machine learning techniques and models.

The main cause of harassment of girls is a lack of safety and security or actual consequences in the lives of women. In some cases, women were harassed by their neighbours as they were walking to school or there was a lack of safety that caused young girls to feel anxious. As a result, they suffer throughout their lives as a result of that one incident in which they were forced to do something unacceptable or were sexually harassed by one of their own next-door neighbours or any other unidentified person. From the standpoint of women's legal rights to influence the city without fear of physical assault or unwanted sexual advances, the best cities are those that offer women safety and security. It is the responsibility of culture to imprecise the demand of security for women and also recognise that women and girls likewise have a right similar to males have to be secure in the City rather than imposing constraints on women that society typically enforces. The names of individuals and women whose names appear in the analysis of the texts from Twitter who endure unwanted sexual propositions and duplicitous behaviour from males in Indian cities that make them uncomfortable to walk in public are also included. the data collecting on women's health that was obtained through Twitter.

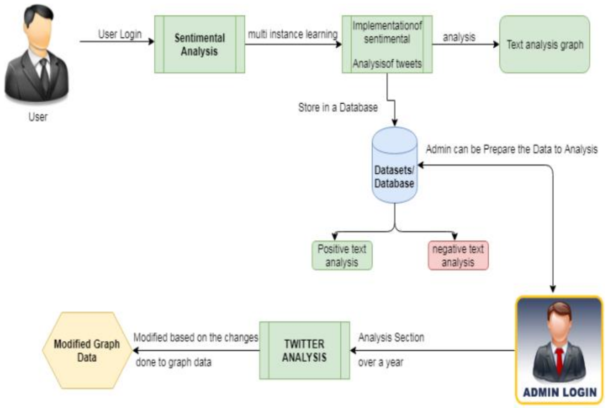


Fig : Architecture

LITERATURE SURVEY

Phrase polarity analysis utilizing lexical affect score and syntactic n-grams: We offer a classifier that, given the context, can predict the polarity of subjective assertions in a phrase. Our method uses the Dictionary of Affect in Language (DAL) and WordNet to score the majority of our input words without the need for human intervention. This is possible using our strategy. To account for context, lexical scoring involves n-gram analysis. DAL scores are used with syntactic components to derive n-grams from the whole corpus of utterances. The syntactic polarity of the statement is also evaluated. Our findings outperform both a majority-class baseline and a more difficult lexical n-grams baseline.

Existing system: safety evaluation of women's social media discussions using machine learning. A woman may rant online after a negative encounter. Her postings, tweets, and messages might disclose which nations have the highest rate of female abuse.

Reduce your speed:

Despite this, many women are terrified and uneasy as they go through malls and retail centers on their way to work. These are high-traffic areas.

System conceptualization:

We collected MEETOO tweets about women's safety and security and stored them in a Python dataset folder using TWEEPY since the author of the planned research used it to gather tweets from Twitter whenever it was accessible. We accomplished this since we use TWEEPY to download tweets. This program analyzes tweets to determine the emotions of women.

Natural language toolkit (NLTK) eliminates superfluous symbols and stop words from tweets.

The author uses TEXTBLOB corpora and language to determine tweet polarity. Tweets with a value less than 0 have a negative polarity, those with a value between 0.1 and 0.5 are neutral, and those with a value more than 0.5 are positive.

Advantages:

1. The postings on Twitter include the names of men and women who have spoken out against male abuse, harassment, and unethical conduct in Indian cities, making it dangerous for them to move freely. These activities limit women's freedom of movement.

The second Twitter-gathered statistic regarding the safety of Indian women is:

Qualities of High-Quality Software

Because of its close link with the database, the program is simple to maintain. i. Fewer forms represent a less complicated onboarding process for new users.

iv) Flexibility It is simple to modify software to suit new capabilities.

METHODOLOGY

TWITTER ANALYSIS

Social media can be thought of as an excellent platform to find people's viewpoints and also sensations regarding various events because people actively connect and also share their viewpoint on social networks like Facebook and Twitter. There are a number of opinion-oriented data gathering and analytics tools that try to glean people's opinions on various topics. People often use different phrases and acronyms on Twitter since tweets are so brief. Existing NLP systems struggle to quickly extract the sentiment from these phrases. In order to extract the polarity of the phrases, many scientists have used deep learning and artificial intelligence techniques. Due to the fact that so many people are using social media sites platforms like Facebook, Twitter, and Instagram, a lot of people are using them to express their feelings and also opinions regarding what they think of Indian cities and also Indian society. Utilising Twitter analytics for business is similar to receiving a monthly Twitter analytics progress report. To help you track effectiveness, Twitter analytics compile all the actions that users perform after discovering your content or profile, including clicks, follows, likes, and expands.

SVM:

‘A supervised machine learning approach called "Support Vector Device" (SVM) can be used to overcome both classification and regression challenges. However, classification concerns are where it is most frequently utilised. In this algorithm, each data item is represented as a point in n-dimensional space (where n is a collection of functions you have), with each function's value having a specific coordinate. Then, using a hyper-plane to effectively separate the two classes, we do category analysis (see the picture below). Simply put, support vectors are individual observation's coordinates. The frontier that best separates the two classes (hyper-plane/line) is support vector equipment. A support vector machine, which can be used for tasks like outliers finding or category, regression, generates an active plane or set of active planes in a high- or infinite-dimensional space. The active plane that has the greatest distance to the closest training-data element of any kind of class (supposed functional margin) can easily achieve a huge splitting up since, generally speaking, the larger the margin, the smaller the classifier's generalisation error. The sets to differentiate are frequently not linearly separable in that area, despite the fact that the initial problem may be addressed in a finite dimensional space. In order to make the splitting apart easier due to space, it was suggested that the initial finite-dimensional space be mapped into a much higher-dimensional space.

RECOMMENDED SYSTEM

Women have the right to the city, which entitles them to travel wherever they like, including to educational institutions and other destinations that they choose. But because of the numerous unidentified Eyes body shaming and pestering these women, women feel unsafe in places like malls and shopping malls on their way to their workplace. The main cause of harassment of girls is security or a lack of tangible impacts in women's lives. There have been instances where women have been harassed by their neighbours as they commuted to school, or when a lack of safety and security caused young girls to feel psychologically anxious, causing them to carry that anxiety throughout their lives. These situations have either required the women to perform inappropriate tasks or exposed them to abuse from their neighbours or other unidentified people. The majority of secure cities address female safety from the perspective of women's legal rights to influence the city without fear of physical assault or harassment. It is the responsibility of society to recognise that women and also women also have a right to the same degree that men need to be secure in the city, rather than imposing the restrictions on women that culture typically imposes. The advantages of the proposed system include The names of people and women who speak out against sexual harassment and dishonest behaviour by males in Indian cities that prevents them from moving around freely are also included in the analysis of the twitter texts collection. The data set about women's safety and security in Indian society that was obtained via Twitter

Modules

1) Belief Analysis: The data is ready for the sentimental evaluation process once the classifier has cleaned up the dataset. A few methods of nostalgic analysis include machine learning, vocabulary-based learning, and hybrid learning. Nero Linguistic Programmes and Natural Language Processing are two other techniques. The equipment learning approach involves educating a dataset and then testing that taught dataset. Data education and data evaluation are necessary for the classifier to apply the formula. A few of the algorithms that can be used to train the classifier include Maximum Degeneration, Naives Bayes classification, Bayesian Networks, and Network Support Vector Maker. To assess the sentiment classifier's effectiveness, information is evaluated. Lexicon-based learning does not utilise a training dataset. This tactic makes use of a built-in vocabulary that includes words that are related to human beliefs. The third method, called hybrid knowledge, combines the device learning and vocabulary learning approaches in order to improve classifier performance.

2) Sentiment Category: The dataset is prepared for categorization at this point. Each and every tweet's language will be studied, and opinions will be formed in accordance with subjectivity. Sentences with subjective expression are kept, whereas sentences with objective expression are rejected. Various levels of nostalgic analysis use tools like unigrams, negation, lemmas, and more. Positive and negative emotions can be differentiated broadly into 2 kinds. 3) EXECUTION OF SENTIMENTAL EVALUATIONS OF TWEETS Each of the subjective sentences that will undoubtedly be kept is classified into good, bad, like, dislike, or positive and adverse.

Keep a record of the tweets you retrieved through Twitter's API. The existence of the Twitter API has made a variety of methods for the examination of information from social networks with nostalgia quite accessible. A selection of available libraries have been used in this project.

Four) CHART A In order to reduce the distance between the real and clinically depressed interaction graphs, a social chart version is used to create the clinically depressed communication graph G_.The information from the input (true) social networking sites is used to create an interaction graph G. The way social media celebrities interact with one another is depicted in a communication graph [25, 26]. After identifying the entities and the communications they have on social media, an interaction graph is created with a vertex collection V for the entities, an edge set E for the communications, and a quality collection A for the vertex (entity) attributes as well as edge (interaction) connections.

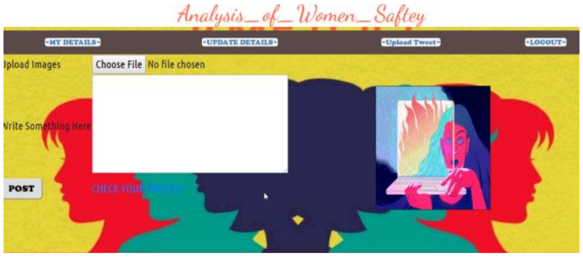
Final Report

If the number of neutral tweets is disproportionately high, it indicates that people are less passionate about the subject and do not have a strong opinion about it. It's also crucial to note that based on the experiment's data, we can get different results because people's perspectives might alter depending on the circumstances. For instance, in 2017, rape news became one of the most popular stories of the year. The fact that the percentage of neutral tweets on several queries is higher than 60% makes the limitations of the sights very clear. By using the analysis we just performed, it is evident that Delhi is the riskiest city, while Chennai is the best.

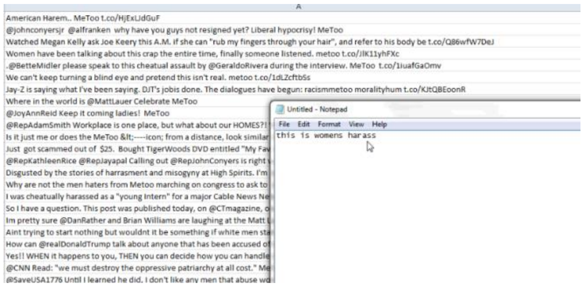
Retrospective analysis plays a crucial role in turning raw data into information that is valuable and relevant. Once the process is complete, numerous sorts of charts can be generated to show the analysis's final outcome. A few examples of how the result can be displayed include pie charts, time series, and bar charts. Bar graphs can be used to determine how positively and negatively people felt about the tweets. Similarly, Time collection can be used to measure in terms of sort, disapproval, and average size of tweet for a specific period. Pie charts can be used to determine the tweet's original source.

SCREENS

Every user data such as credentials, new tweets, re-tweets and tweet score will be stored in the database for the admin to monitor and perform the analysis. The sentiment analysis is applied on the user data in order to monitor and confirm whether any tweets are abusive to women or not. Admin performs this analysis on each and every user tweets to provide safety for the women. Sentimental analysis will be implemented on the tweets of user that are stored in the database. Admin can now prepare the data to perform the analysis. The tweets made by every user of the application will be called as the initial input for the sentiment analysis and hence they will be the dataset. Along with this, text analysis graph can also be shown. Admin will store the filters in the database. Filters are the keywords for which the tweet context will be searched for in order to declare as abusive or not. There can be two types of filters – positive keyword and negative keyword. Positive keywords are those words which are abusive or disrespect the women by any means. Negative keywords are the words which are normal and will not abuse the women..



To add tweets to the page above, choose the dataset and click "Upload Tweets."

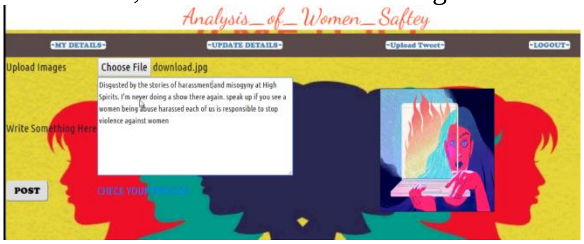


American Harem.. MeToo t.co/HjExLJdGuF
@johnconyersjr @alfranken why have you guys not resigned yet? Liberal hypocrisy! MeToo
Watched Megan Kelly ask Joe Keery this A.M. if she can "rub my fingers through your hair", and refer to his body be t.co/Q86wfW7DeJ
Women have been talking about this crap the entire time, finally someone listened. metoo t.co/JIK11yhFXc
.@BetteMidler please speak to this cheatual assault by @GeraldoRivera during the interview. MeToo t.co/1iuafGaOmv
We can't keep turning a blind eye and pretend this isn't real. metoo t.co/1dLZcftbSs
Jay-Z is saying what I've been saying. DJT's jobis done. The dialogues have begun: racismmetoo moralityhum t.co/KJtQBEoonR
Where in the world is @MattLauer Celebrate MeToo
@JoyAnnReid Keep it coming ladies! MeToo
@RepAdamSmith Workplace is one place, but what about our HOMES?!
Is it just me or does the MeToo &lt;----icon; from a distance, look similar
Just got scammed out of $25. Bought TigerWoods DVD entitled "My Fav
@RepKathleenRice @RepJayapal Calling out @RepJohnConyers is right
Disgusted by the stories of harrasment and misogyny at High Spirits. I'm
Why are not the men haters from Metoo marching on congress to ask to
I was cheatually harassed as a "young Intern" for a major Cable News Ne
So I have a question. This post was published today, on @CTmagazine, o
Im pretty sure @DanRather and Brian Williams are laughing at the Matt
Aint trying to start nothing but wouldnt it be something if white men sta
How can @realDonaldTrump talk about anyone that has been accused of
Yes!! WHEN it happens to you, THEN you can decide how you can handle
@CNN Read: "we must destroy the oppressive patriarchy at all cost." Me
@SaveUSA1776 Until I learned he did. I don't like any men that abuse wo

this is womens harass

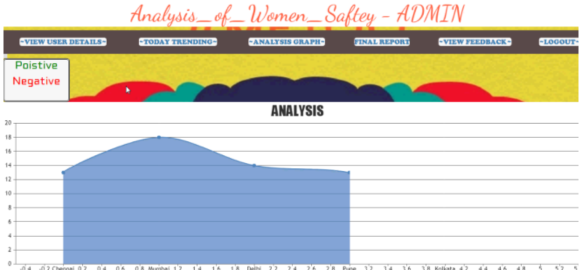
Read tweets from the dataset by clicking "Read Tweets" on the previous screen.

Each line in the window above represents a tweet, so scroll down to see them all. To delete stop words and special symbols from your timeline, choose "Tweets Cleaning."

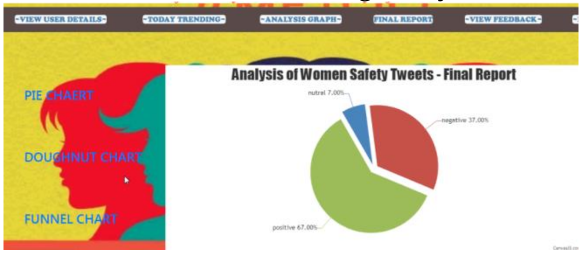


Disgusted by the stories of harassment and misogyny at High Spirits. I'm never doing a show there again. speak up if you see a women being abuse harassed each of us is responsible to stop violence against women

Select "Run Machine Learning Algorithm" after deleting punctuation and stop words from the tweet to continue sentiment prediction.


Analysis_of_Women_Saftey - ADMIN
-VIEW USER DETAILS-
-TODAY TRENDING-
-ANALYSIS GRAPH-
FINAL REPORT
-VIEW FEEDBACK-
-LOGOUT-
Poistive
Negative
ANALYSIS
20
18
16
14
12
10
8
6
4
2
0
-0.4 -0.2 Chennai 0.2 0.4 0.6 0.8 Mumbai 1.2 1.4 1.6 1.8 Delhi 2.2 2.4 2.6 2.8 Pune 3.2 3.4 3.6 3.8 Kolkata 4.2 4.4 4.6 4.8 5 5.2 5

Prior to this one, you could see the whole tweet, including its polarity and emotion score. Scroll down to see all of the tweets. The "Women Safety Graph" button will offer the user with information that will assist them in determining if the location is safe for women. More individuals will tweet favorably or neutrally about a safe location, whereas more will tweet negatively.



There can be 'n' number of positive and negative keywords stored in the database. When the admin implements the sentimental analysis, every keyword in the database will be compared with each and every word in the tweet of the user. If any one of the positive keyword is found in the tweet, that tweet will be classified as positive sentimental analysis and these are abusive to women. If negative keyword is found in the tweet, it will be classified as the negative sentimental analysis which is not abusive to women. Hence, by this stage there will be two types of sentimental analysis made based on the filter in the database. Under positive sentimental analysis, there will be a list of all the tweets in the application that are abusive to women. Similarly, under negative sentimental analysis there will be a list that is clean and are not abusive tweets. Along with the tweet context, user details will also be provided at each of the analysis list.

Our experiment just looks at Twitter, but the same machine learning approaches might be used to Facebook and Instagram. Our experiment just looks at Twitter. Twitter's user interface may take on the present mentality. This would allow the program to do sentiment analysis on millions of tweets, increasing its security.

CONCLUSION:

During this research, we looked at a variety of machine learning methods to organize and analyze Twitter's vast data. This data set includes millions of tweets and SMS every day. Machine learning techniques such as SPC and linear algebraic Factor Model are good for analyzing large volumes of data and classifying the findings. Both are factor modeling approaches. Support vector machines may also be used to extract important information from Twitter and evaluate the safety of women in urban India.

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.