iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59199

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Analysis of Women Safety in Indian Cities Using Machine Learning

Dr. G. Ravi Kumar[1], K. Venkata Giridhar[2], M. Vaishnavi[3], N. Yesumani[4]
[1]*Associate Professor, Dept of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*
[2, 3, 4]*UG Student, Dept of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*

Abstract: ***In later times, ladies have been confronting an disturbing rise in viciousness, counting badgering, totally different cities. This frequently starts with stalking and can raise into injurious ambush. In this dialog, we are centering with in the area of social mediastages like Twitters, FB, and Insta at advancing the security of ladies in India. These stages give a chance at raise mindfulness and teachindividuals approximately taking unequivocal activities against damaging behavior. By utilizing hashtags and spreading messages all inclusive, ladies are enabled to openly express their conclusions and feelings. This empowers us to pick up experiences into their state ofintellect while ladies out for working or travel, utilizing open transportation, or within the nearness of new people. It makes a difference us decide whether they feel secure or not in these circumstances.These times, badgering and other shapes of savagery against ladies are common in numerous cities! Stalking is the primary step towards injurious badgering, which is now and then alluded to as manhandle attack? With specific reference to the utilization of various social media stages, counting Twitter, Facebook, and Instagram, our essential center in this consider at with in the task with social media in advancing women's security in India. This exposition moreover emphasizes the require for common individuals to assume more duty in several Indian city neighborhoods in arrange to ensure the security of ladies in their quick region. Inside the Twitter app, a tweet comprises content messages, photographs, recordings, sound records, emoticons, and hashtags! These regions, require more consideration in regards with women's security, and communalengagement is crucial for improvement.***
Keywords: Twitter, Civic, Facebook, Instagram, Sentimental Analysis, Data Preprocessing.

I. INTRODCUTION

In These Advanced times, bullying and other shapes of brutality against women are common in various cities and all over the world. Stalking is the essential step towards harmful bullying, which is presently and after that implied to as abuse assault. With particular reference to the utilization of differentsocial media stages, checking Twitter Tweets, Facebook and Instagram posts, our fundamental center in this consider is on the work of social media Information in progressing women's security in India. This piece in addition emphasizes the require for common people to expect more obligation in a few Indian city neighborhoods in orchestrate to guarantee the security of women in their speedy locale. Interior the Twitter app, The Tweets comprises of substance messages, photos, recordings, sound records, emoticons, and hashtags. These locales, require more thought in terms of women's security, and communal engagement is vital for enhancement. Opposite to common convictions, women have the proper to feel secure, though utilizing social media stages and have to be be secured from Cyber threats. The government's part it to stay in most extreme centrality in actualizing strict laws to ensure the online securitywith lady at these times. Furthermore, the cooperation of non-profit organizations is imperative to create social mindfulness on these issues and offer back to casualties of online viciousness and badgering towards woman.The progression of social media has driven to both positive and negative impacts in society, especially concerning women's safety. In conclusion, collaboration among government, non profit organizations, and people is vital in making a more secure environment for ladies in India through analyzing the information like messages through Twitter organize with the assistance of a few computerized strategies like wistful investigation, Normal Dialect Preparing at that point Actualizing and processing them with the assistance of Different algorithms like For case XGBoosting Machine Learning Calculation.To Extent quick preparing time and Tall exactness, it is signify to select calculation.

II. LITERATURE SURVEY

1) Apoorva Agarwal, Fadi Biadsy and KathleenR. Mckeown: Within the cutting edge day, Twitter has risen as a communal microblogging arrange that can match other shapesof media. As individuals connected, they watch hundreds of millions of stoners and create over 5 hundred carefully on open media stages like FB and millions of "Tweets" each day. One way to degree the Twitter social framework is as a smooth way to memorize almost people's suppositions.

©IJRASET: All Rights are Reserved

2) Barbosa, L., & Feng, J:In this paper, we propose a 2- step opinion examination classification strategy for Twitter, which to begin with classifies messages as subjective and objective, and encourage recognizes the subjective tweets as positive or negative. To diminish the labeling exertion in making these classifiers, instep of utilizing physically explained information to compose the preparing information, as normal administered learning approaches, we use sources of boisterous names as our preparing information. These loud names were given by many assumption location websites over twitter information. To superior utilize these sources, we confirm the potential esteem of utilizing and combining them, giving an investigation of the given names, look at diverse methodologies of combining these sources in arrange to get the finest result; and, propose a more vigorous include set that captures a a lot of theoretical showing of tweet, composed by meta-information related to words and particular characteristic with what tweet can be composed. By utilizing it, we point to handle way better: the issue of need of data on tweets, making a difference on the generalization prepare of the classification calculations; and the loud and one-sidednames given by those websites.

3) Gupta, B. ,Negi, M. ,Vishwakarma, K. ,Rawat, G., Badhani, P.: Twitter has developed as a major micro blogging site, having over 100 million clients creating over 500 million tweets each day. With such huge gathering of people, Twitter has reliably pulled in clients to communicate their conclusions and point of view almost any issue, brand, company orelse various different point of intrigued. Due tothis reason, Twitter is utilized as an enlightening source bynumerous organizations, educate and companies. On Twitter, clients are permitted to share their conclusions within the shape of tweets, utilizing as it were 140 characters. This leads to individuals compacting their explanations by using slang, shortened forms, emoticons, brief shapes etc. In conjunction with this, individuals pass on their conclusions by utilizingmockery and polysemy.

4) Mamgain,N., Mehta,E., Mital,A., & Bhat, G: In today's world, conclusions and reviews accessible to us are one of the most basic components in defining our sees and impacting the success of a brand, item or benefit. With the coming and growth of social media within the world, partners regularly take to expressing their conclusions on prevalent social media, to be specific Twitter. Whereas Twitter information is greatly instructive, it presents a challenge for investigation since of itshumongous and disorganized nature. This paper may be a intensive exertion to plunge into novels space of perform the sentimental analysis of audience suppositions as respecting to the beat collage in Indian. Other than taking extra preprocessing measures just like the extension of Netlingo & expulsion of copy information, a probabilitys show based on the Bayes hypothesis was utilized for spelling adjustment, which is neglected in other investigate ponders. This paper moreover highlightning a difference of the comes about gotten by misusing the taking after machine learning calculations: Naïve Bayes and Bolster Vector Machine and an Fake Neural Organize show: Multilayer Perceptron. Besides, a differentiate has beendisplayed between four diverse parts of SVM: RBF, straight, polynomial and sigmoid.

5) Sahayak, V., Shete, V., & Pathan, A: Twitter site is the popular small scale blogging administrations where client can study and post messages which are 148 characters in length. Twitter messages are also called as Tweets. We are going utilize these tweets as crude information. We'll utilize a strategy that consequently extricates tweets into positive, negative or unbiased assumptions. By utilizing the assumption investigation the client can know the criticism almost the item or administrations some time recently making a buy. Thecompany can utilize assumption investigation to know the supposition of clients approximately their items, with it we are able to analyze client fulfillment and concurring to with they are able to process the item. Estimation examination has gotten to be one of well known inquire about zone in computational phonetics, since of the blast of assumption data from social web locales (i.e., Twitter and Facebook), online gatherings, and blogs as in paper.

6) Charniak, E., & Johnson, M: Twitter could be a stage broadly utilized by individuals to precise their suppositions and show opinions on diverse events. Estimation examination which is approaching with analyzing information & recover idea that with epitomizes. Twitter estimation investigation which an advantage with assumption investigation on information through Twitters(tweeting), with arrange to extricate opinions passed on by the client. Within the previous times, as investigate with the work was reliably developed. The reasons below at this can be the challenging arrange with the tweet which make the handling troublesome.

7) Kim, S. M., & Hovy, E: Distinguishing assumptions (the full of feeling parts of conclusions) may be a challenging issue. We display a framework that, given a theme, naturallyfinds the individuals who hold conclusions around that theme and the assumption of each supposition. The framework contains a module for deciding word estimation and another for combining opinions inside a sentence. We explore with different models of classifying and combining assumption at word and sentence levels, with promising comes about.

III. RELATED WORKS

In the quest for innovation and efficiency, modern projects frequently rely on existing solutions as fundamental building blocks for development. That method doesnt only reminds the expertise and advancements of those who came before us but also nurtures a collaborative ecosystem where ideas can evolve and confront new challenges. In our project, we wholeheartedly embrace this ethos, conscientiously integrating elements from existing solutions to enrich our endeavor. These existing solutions serve as guiding lights, offering insights and frameworks that shape the direction of our project.

A. NLP(Natural Language Processing)

Computational methods are exceptionally associated in Ordinary Lingo Taking care of (NLP) to assessment, get, and make information on human lingos. It incorporates a large values with errands like machine interpretation, named substance assertion, doubt testing, and substanance classification. NLP calculations distill structure and meaning from substance data by applying neural frameworks, quantifiable models, and etymlogical rules. Its livelihoods consolidate a wide amplify of zones, checking chatbots, virtual arsssitstants, data recovery systems, and estimation testing devices, progressing robotized planning and tongue comprehension!

B. A. Sentimental Analysis

The tweets picked up from Twitter API given by Twitter itself. With the nearness of Twitter API, its has number of methods available for wistful examination of information on social media. A few accessible libraries were utilized in this venture. Here's how to extricate feelings from a tweet:

1) Beginning with downloading the thoughtful word reference
2) In it the position download the Twitter testing enlightening collections and include them as a commitment to the program.
3) Clean tweets by dispensing with the halt words and repetitive letters.
4) Tokenize each word and apportion solidarity to the words within the dataset and bolster it to the program.
5) For each word, differentiate it and positive conclusions and negative sentiments word reference and a short time later increment the great tally or critical number of the common expression.
6) At long last, given the positive check and negative tally, the result rate around feeling to conclude the limit which is orchestrated in Great, Critical, and Fair-minded. Engineers have done diverse wistful investigations .

C. XGBoosting Algoritum

Extraordinary Angle Boosting, or XGBoost for brief, could be a effective machine learning strategy that exceeds expectations at taking care of both organized and unstructured information. It is based on the slope boosting design and finds utilize in gathering learning procedures. The adaptability of the approach makes it fitting for dealing with expansive datasets, and regularization procedures are coordinates into the calculation to anticipate overfitting issues. Iteratively progressing expectation precision, XGBoost makes utilize of an gathering of choice trees. Due to its flexibility and steadfastness, it has been broadly utilized for a assortment of ML applications, counting relapse, positioning examination, and classification.

1) *Step1:* Pre-processing.
2) *Step2:* Features extraction utilizing counting vectorization, word level TF-IDF and n-gram level TF-IDF. Then Converting the Text into Vectors.

$$TF(t) = \frac{\text{(Number of times term t appears in a document)}}{\text{(Total number of terms in the documents)}}$$

$$IDF(T) = \log\left(\frac{\text{Total number of docs}}{\text{Number of docs with term t in it}}\right)$$

3) *Step3:* Fit the training dataset, it learns the model
4) *Step4:* XGBoosting classifier is utilized to predict the values as the XGBoosting initializtion is done, the accurate calculation of the model will be done inlcuding all the three vectorization techniques.

D. Algorithms Comparsion Table

ALGORITHUM	ACCURACY	SCALABILITY	ROBUSTNESS	INSIGHTS
Logistic Regression	78%	High	Moder-ate	Identification of key factors influencing women's safety, such as lighting, police presence, transportation options.
Random Forest	85%	High	High	Ensemble method offers improved accuracy by combining multiple decision trees. Identification of complex relation between safety features and actionable insight.
k-Nearest Neighbors	82%	Low	High	Instance-based learning method identifying local safety patterns. Recommendations for immediate safety improvements in proximity to high-risk locations.
Gradient Boosting	87%	High	High	Ensemble method similar to XGBoost, offering improved accuracy by boosting weak learners.Insights into areas requiring targeted safety interventions for maximum impact.

Table-1:Algorithm Comparison Table

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

IV. METHODOLOGY

Within the Estimation Investigation the taking after steps are major to distinguish the positive, negative or unbiased of the twitter post.

They are:

1) Collecting the Dataset: The information is assembled from the twitter utilizing API. Application program interface (API) is utilized to accumulate the information. Twitter site could be a source which comprises of clients tweets.

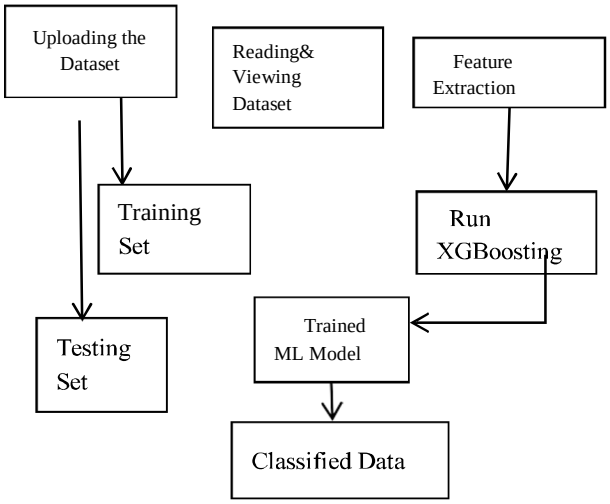


Fig 1 : System Architecture

2) *Pre-Processing the Data-set:* In Information Pre- processing evacuating boisterous, disconnected information, inconsistent and incomplete info from dataset. For the most part, in twitter we have to be evacuate URLs, extraordinary characters, Re-tweets, hash labels.

3) *Feature Extraction:* In this work, we utilized Sack of Words to extricate highlights from content archives. After extraction, these highlights used at training the ML calculations. It makes a language of the clear large number of novel words happening in all the reports within the arrangement set. Sack of words highlights containing term frequencies of each word in each record, i.e. the number ofevent and not sequence or arrange of words things. This will be done by Tally Vectorizer strategy in Python.

4) *Classification:* A classification issue is associated within the occasion that the abandon variable may be a title or category,in our work “Positive” or “Negative”.

V. PROPOSED SYSTEM

The method known as "Sentimental Analysis CLASSIFICATION" reports the tweets that are chosen andgiven by the Twitter program. According to the existence of the Twitter app, there are a few strategies accessible for emotive investigation of social media information.

1) A Downturn in Exchange The crevice between the genuine and sad commerce charts is decreased when a social organize demonstrate makes a G-graph.

2) A commerce chart is developed, and substances andtheir connections in social mass media are associated.

3) The chart appears the likelihood with regard to different topics. Individuals tend to utilize diverse terms and a prepared collection of information that includes condensation since Twitter contains tweets that are unbiased, negative, and positive in expansion to the brief handbooks.

4) Agreeing to advanced NLP, these prompts are touchysufficient to esteem their sentiments.

This expansion proposes a strategy for evaluating women'ssecurity through the examination of social media organizing messages, particularly tweets almost women's security issues gotten through the utilize of the MEETOO hashtag.

©IJRASET: All Rights are Reserved

The Characteristic Tongue Gadget Pack (NLTK) and the Python programming tongue will be utilized to assemble and plan tweets in arrange to extricate pivotal information. The estimation and substance of the tweets will be assessedutilizing the XG Boost calculation, which is well-known forits capacity to handle both organized and unstructured information.
Extraordinary Angle Boosting, or XGBoost for brief, will bea effective machine learning strategy that exceeds expectations at taking care of both organized and unstructured information. It is related up on slope boosting design and finds utilize in gathering learning procedures. The adaptability of the approach makes it fitting for dealing with expansive datasets, and regularization procedures are coordinates into the calculation to anticipate overfitting issues. Iteratively progressing expectation precision, XGBoost makes utilize of an gathering of choice trees. Dueto its flexibility and steadfastness, it can be used for a assortment of ML applications, counting relapse, positioningexamination, and classification.

VI. RESULTS AND DISCUSSION

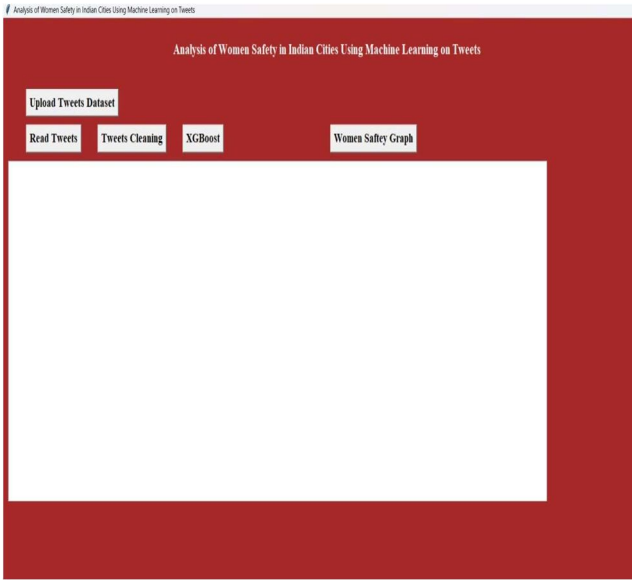


Fig 2: Home Screen

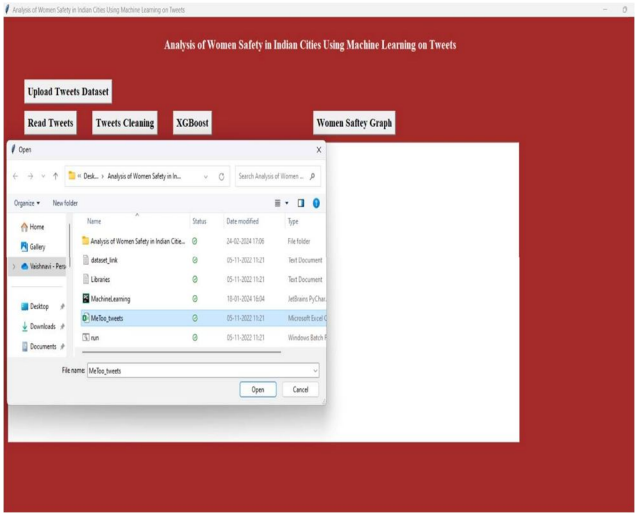


Fig 3: Uploading dataset

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

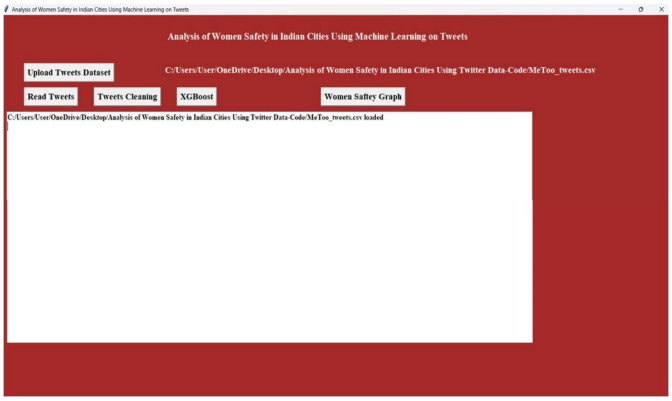


Fig 4: Data Loading

In the above step upload MeeToo_Tweets.csv record and After that press on'Open' button to stack dataset in above activity each line represents single document, will be ablescroll down

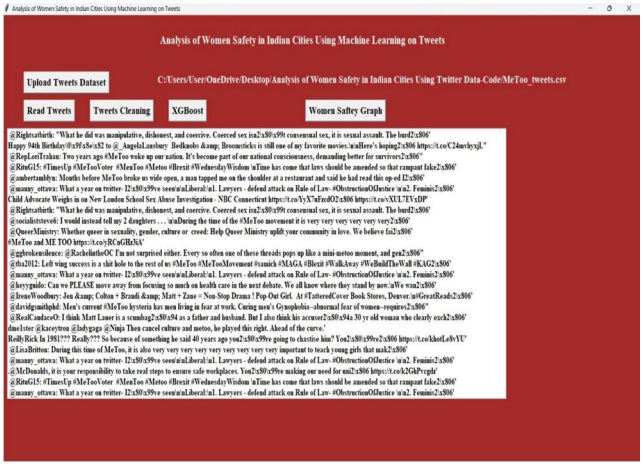


Fig 5: Reading the Tweets

At the above screen we will view every single doc, includes special texts, halt words and to minimize those texts tap on 'Tweets Cleaning button

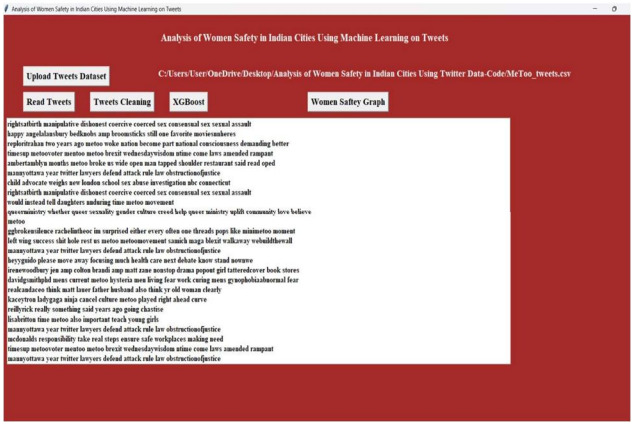


Fig 6: Tweets Cleaning

At above screen we will see all extraordinary images and Halt words and expels through the tweets and as it wereclean wordsare there and presently press on "Run MLalgorithum button to foresee assumptions

©IJRASET: All Rights are Reserved

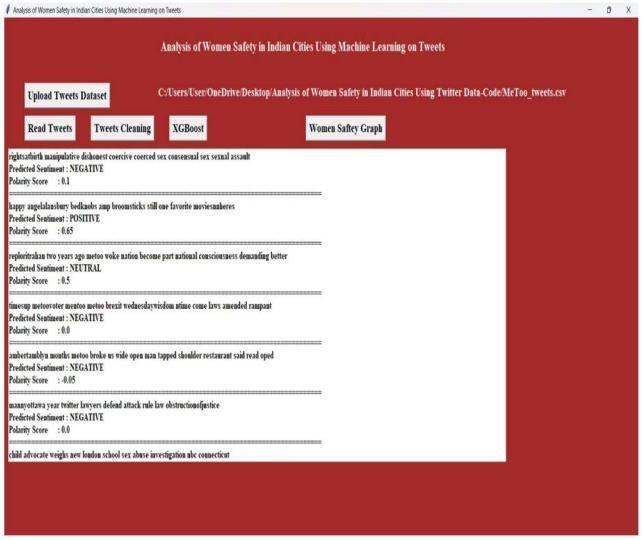


Fig 7: XG Boosting Algorithm

See all extraordinary images and halt words expefrom tweets and as it were clean words are there and presently press at 'Run ML Algorithm' button to foresee assumptions.

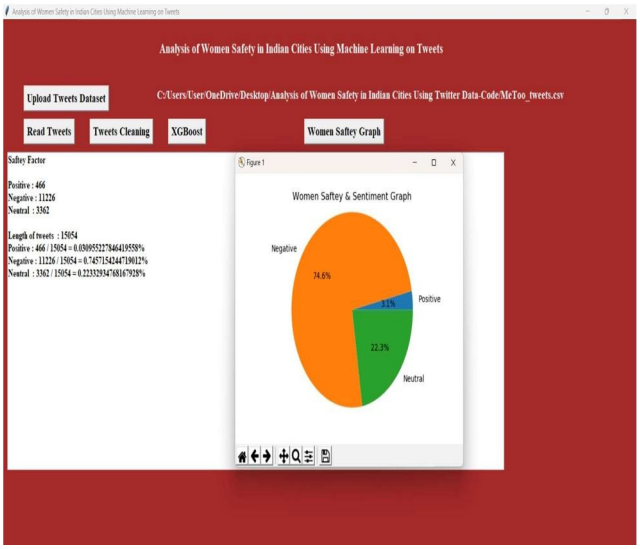


Fig 8: Women Safety Graph

In above activity 0.74 multiplied by 100 which gives 74%that means 74% of audience are talking negatively and area is not safe and only 22 and 3% of audience are talking positively.

VII. CONCLUSION

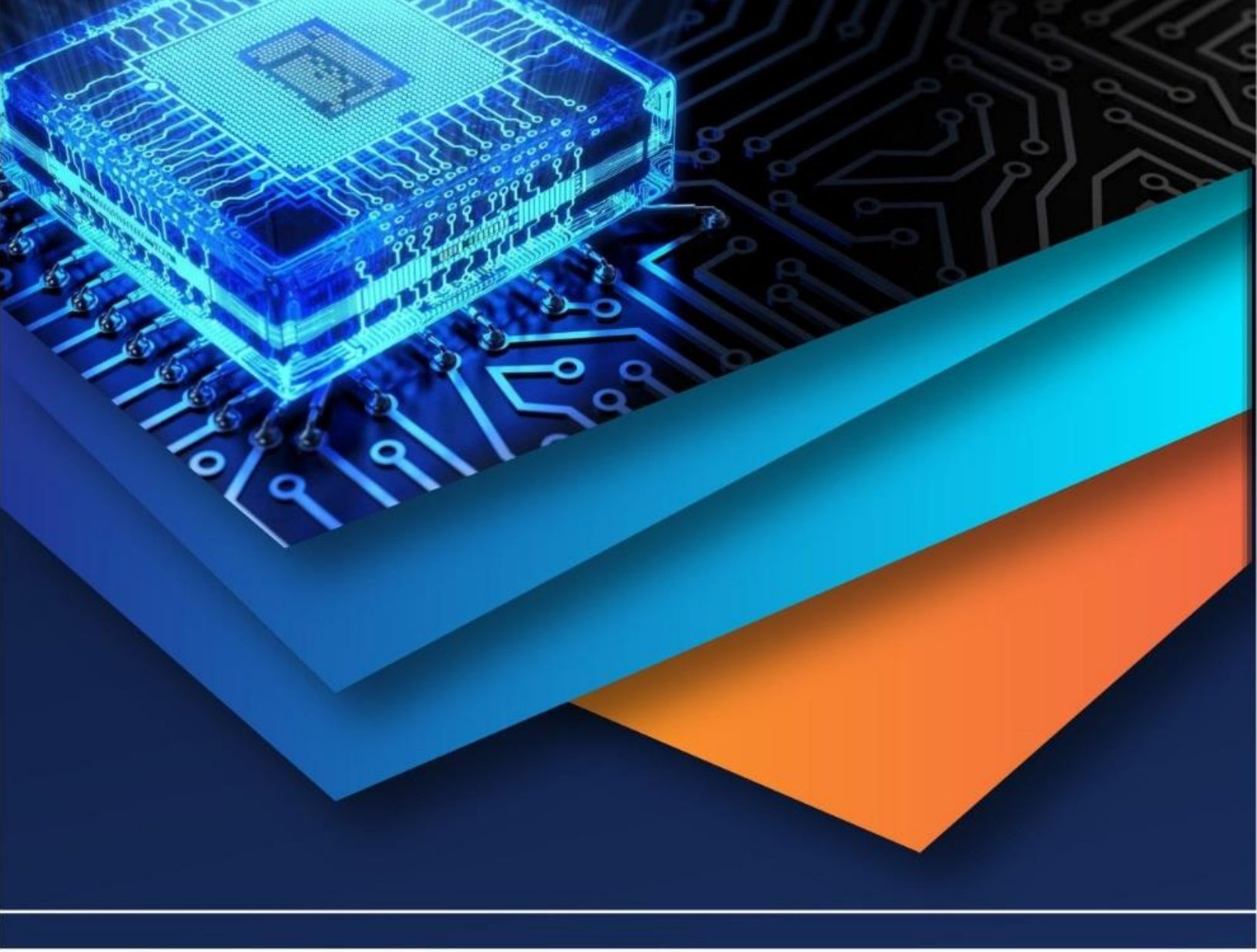
All through the term sheet we utilize examined almost various ML calculations which could offer assistance to organizing and analyzing the colossal sum of Tweeting information gotten counting huge number of Texts and content messages shared each cycle.

This ML calculations are exceptionally compelling and valuable as it arrives to analysis with expansive sum of info include with SPC calculation and direct logarithmic Calculate Show approaches which offer assistance to assist categorize the information into significant bunches.

Back VM with however another shape at ML calculation that's exceptionally prevalent in extricating Valuable data through the Tweeting and receive a thought around at stats of ladies security in India.

REFERENCES

[1] Apoorv Agarwal, Fadi Biadsy,and Kathleen R. Mckeown. "Contextual phrase-level polarity analysis usinglexical affect scoring and syntactic n-grams." Proceedings of 12th Conference of European Chapter of the Association for Computational Linguistics. Association for the ComputationalLinguistics, 2009.
[2] Barbosa, L., & Feng, J. (2010, August). Robusted sentimental detection on twitter from biased and noisy data. InColing 2010: Posters (pp. 36-44).
[3] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G.,Badhani, P., & Tech, B. (2017). Study for Twitter sentiment analysis using ML algorithms on Python. International Journalof Computer Applications, 165(9), 29-34.
[4] Mamgain,N., Mehta, E., Mittal, A., & Bhatt, G. (2016, March). Sentimental analyzing of top colleges in India utilizing Twitter data. In 2016 international conference on computational techniques in information and communication technologies (ICCTICT) (pp. 525-530). IEEE.
[5] Sahayak, V., Shete, V., & Pathan, A. (2015). Sentimentanalysis on twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.
[6] Charniak, E., & Johnson, M. (2005, June). Coarse-to- fine n-best parsing and maxent discriminative reranking. In Proceedings of the 43rd Annual Meeting of the Association forComputational Linguistics (ACL'05) (pp. 173-180).
[7] Kim, S. M., & Hovy, E. (2004). Deterninating the sentiments of opinions. In COLING 2004: Proceedings of the 20th International Conference at Computational Linguistics (pp. 1367-1373).
[8] Bermingham, A., & Smeaton, A. F. (2010, October). Classifying sentiment in microblogs: is brevity an advantage?.In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 1833-1836).
[9] Klein, D., & Manning, C. D. (2003, July). Accurately unlexicalized parsing. In Proceeding with the 41st annually meeting of the association for computational linguistics (pp. 423-430).
[10] Gamon, M. (2004). Sentimental classifications oncustomers feedback data: noise data, largely feature vectors, and the role of linguistics of analysis. In COLING 2004:Proceeding of 20th international conference on computationallinguistics (pp. 841-847).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)