Twitter Sentiment Analysis On Coronavirus Outbreak Using Machine Learning Algorithms

Harshita Tandale

Computer engineering

Modern Education Society's College of Engineering.

Pune, india.

harshitatandale14@gmail.com

Pratiksha Shitole

Computer engineering

Modern Education Society's College of Engineering.

Pune, India.

pratikshashitole2000@gmail.com

Prof. Sradhha khonde

Computer engineering

Modern Education Society's College of Engineering.

Pune, India.

Shraddha.khonde@mescoepune.org

Vaishnavi Rubde
Computer Engineering

Modern Education Society's College of Engineering.

Pune, India

rubdevaishnavi@gmail.com

Mahima Bachhav
Computer Engineering
Modern Education Society's College of Engineering

Pune,India

mahimabachhav@gmail.com

Abstract— Social media may be a supply that produces huge quantity of knowledge on Associate in nursing unprecedented scale. It is a platform for each person to share their views, opinions and experiences except for simply being a platform that offers data to the general public UN agency search for data on the sickness. As surprising because the prevalence of corona virus disease 2019 (COVID-19) was, it's been radically touching folks everywhere the globe, there is a requirement to analyze the opinion of individuals on the pandemic COVID-19. This paper focuses on the sentiment analysis of COVID-19 victimization twitter knowledge. The analyses are primarily based on the machine learning algorithms. this text provides Associate in Nursing analysis on however folks react to an epidemic an endemic scourge deadly disease virus virulent disease pestilence} outbreak, what quantity they're attentive to the sickness and its symptoms, what precautionary measures they are taking and whether or not or not folks are following government's pointers etc. Understanding the posts on social media pages throughout a pandemic natural event permits health agencies and volunteers to higher assess and perceive the public's insolences, sentiments and wishes so as to deliver acceptable and effective information.

Keywords—: Twitter, Corona Virus, Machine Language

Introduction

In the end of 2019, the COVID-19, on-going corona virus disease originated in Wuhan, China. The novel virus is believed to have created from an animal-to-human spill over event linked to seafood and live-animal markets like butcher shops. The virus has spread and communicated locally in Wuhan and other places in China, despite strict intervention measures and efforts implemented in the region. It is affecting 203 countries and territories around the world as on 2 April 2020. Corona virus affected 936,725 people,

Claimed more than 47260 lives as of 2 April 2020. According to WHO, the fatality rate is around 2% as discussed in the press conference that was held on the 29th of January,2020. The World Health Organization declared the corona virus pandemic outbreak as a Global Public Health Emergency. The on-going outbreak of corona virus disease, has taken 58 lives, along with 2032 confirmed cases in India, as of 2nd April 2020. Due to COVID-19, several people had to lose their lives and it is surprising to know that the number of deaths associated with COVID-19 surpassed the other corona viruses SARS-Cove, and MERS-Cove, which stood a highest threat to the world's public health.

All countries square measure taking numerous steps to regulate the pandemic like Janata curfew, nation lockdown, cancelling transport facilities, impose social distancing

restrictions etc. Twitter is one of the quickest data sharing platform among all on-line social networking media.

Messages or tweets on twitter vary from personal data to world news or events. Analysing this incessantly generated knowledge is extremely fascinating and informative sanctionative users or organisations to accumulate data. This helps the govt. or organisations to understand how way the general public is alert to the illness happening, its symptoms and precaution measures. Sentiment analysis is well studied exploitation Twitter knowledge in recent days to predict and/or monitor health connected problems. Twitter contains Brobdingnagian range of unimportant messages and unwanted or contaminated content that negatively affects the perception analysis performance. The Traditional techniques don't seem to be compatible attributable to the short length of tweets, writing system and grammatical errors, and also the frequent use of informal languages. During this effort, data about diseases and diseases is extracted from Twitter with spatiotemporal restraints throughout a given illness happening amount. Sentiment analysis is employed to know the perception of the people concerning corona virus illness and conjointly to understand what extent individuals' keep is affected. The analysis results can facilitate quicker response to and preparation for epidemics and even be terribly helpful for each public and governments to create additional well-read selections.

I. LITERATURE REVIEW

Two most typical features in informatics TF-IDF and N-Grams [1] were used on the SS-Tweet dataset. Six completely different algorithms are used for classification and that they found that TF-IDF feature is giving higher results (3-4%) as associated with N-Gram options. Categorization of individual's views into positive, negative and neutral was done on 393,869 static twitter information that was taken from world web site. [2] Authors came out the with the result that most and supreme people's" read is neutral. A hybrid approach [3] is developed for sentiment analysis that makes use of machine learning algorithms like Naïve mathematician and support vector machines (SVM). Comparison study on political opinions was created exploitation sentiment analysis. [4] The paper discovers the numerous sentiments applied to Twitter information and their outcomes. Varied techniques for Twitter sentiment analysis strategies are mentioned, which incorporates machine learning, ensemble approaches and lexicon(dictionary) primarily based approaches. Twitter Sentiment analysis supported ensemble technique and hybrid sentiment analysis techniques were looked into. [5] Compare the sentiment with vote information to envision however ample correlation is shared.

Lexicon and Naive Bayes Machine Learning algorithmic rule was wont to analyze the sentiment and label or tag tweets supported hash tag content, [7] The analysis has been carried by classifying Associate in Nursing opinion/view/belief within the kind of commentaries into 2 categories, which is positive and negative with the extent of accurateness that's prejudiced by the coaching procedure. Public sentiment knowledge to the traveler lures comprised within the positive sentiment. [8]This work acquainted with the sentence-level approaches to run on automaton OS and measures their performance in terms of memory usage, CPU usage, and battery consumption. Their findings reveal sentence-level approaches that require nearly no editions and run relatively fast.

[10] Examine the results of sentiment analysis options in locating ADR mentions. Methods. Results show that

sentiment analysis options slightly advance ADR credentials in tweets and well-being connected forum posts. This study shows that adding sentiment analysis options will slightly advance the performance of even a progressive ADR identification technique. [11] paper suggests a method of sentiment lexicon implanting that signifies sentiment word's linguistics relations. The define was that combined secret writing morphemes and their POS tags, and dealing out solely vital lexical morphemes within the embedding house. As a result, the revised embedding approach increased the performance of sentiment classification.

[12] verified that AD-related tweets accustomed uphold public stigma, that compact negative expectations of people with the illness. [13] analyses text sentiment in social media mistreatment lexical-opinion technique. [14] analyses Twitter datasets in NLTK Corpora mistreatment a feature extraction technique. varied machine learning classifiers like Multinomial, Bernoulli, Logistic Regression square measure mentioned. Experimental results demonstrate that Bernoulli, Logistic Regression, and SGD classifier reached accuracy as high as seventy fifth.[15] designed a rule classifier with a voting-based ensemble of supervised classifiers. A set of rules supported the occurrences of emoticons and sentiment-bearing words square measure framed. Experimental results demonstrate the effectiveness of the tactic. [16] focused on Natural Language Toolkit techniques for process knowledge from Twitter. Feature choice is extracted by Chi sq. check and Naïve Bayes classifier is employed for coaching.[17] focuses on sentiment analysis at entity level for Twitter. A lexicon based mostly approach is adopted to perform entity level sentiment analysis. Recall is improved by inclusion of seemingly tweets. The classifier helps to coach and assign polarities to the entities within the new tweets.

II. PROPOSED SYSTEM MACHINE LEARNING APPROCH

Machine learning strategies square measure trained on datasets and a model is made for analysis.
Based on the accuracy of the model, the machine learning technique is appropriate. The three methods in machine learning algorithms square measure supervised learning, unattended learning and reinforcement learning. In supervised learning, the model is trained victimization labeled information that contains each input and results. The sections of process square measure coaching section and testing phase.
Unsupervised learning strategies don't use coaching information or labeled information. It finds the hidden structures or patterns from unlabeled information.

Supervised Learning

Supervised learning needs a well-labelled dataset to coach. supervised learning is of 2 types particularly regression and classification. Classification techniques facilitate to seek out the appropriate category labels which may predict the positive, negative and neutral sentiments. A machine learning model is developed that uses the tagged knowledge to coach, classify the tweets and predict the emotions of the tweets. call Tree, Random Forest, Bayesian belief network, Naive Thomas Bayes and KNN classifiers ar a number of the algorithms that ar employed in this method.

Unsupervised Learning

Unsupervised ways ar supported machine learning or lexicon. the necessity of the labelled datasets isn't needed in unsupervised learning. Sentiment analysis once done using unsupervised learning; it's typically supported a Sentiment Lexicon. Text classification helps to extract phrases that contain adjectives or adverbs to estimate a phrase‟s linguistics orientation. linguistics orientation is then accustomed classify the emotions
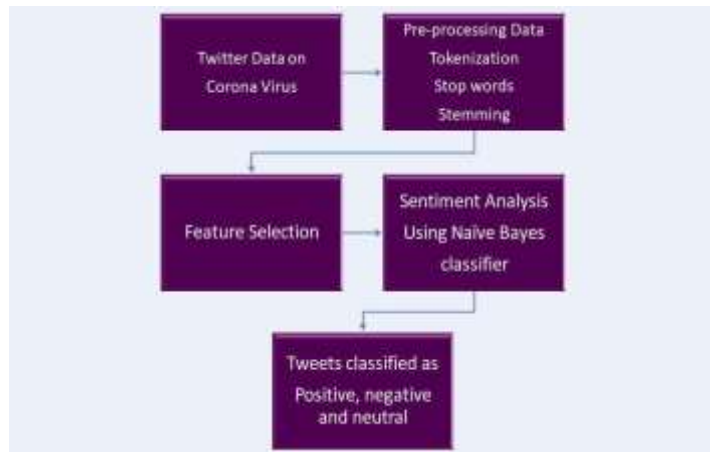

Fig 1 Proposed Model

Algorithm:Twitter_CoronavirusSentimentAnalyzer()

1. Create a twitter account
2. Get consumer key, consumer secret, access key, access secret from twitter login
3. Initialize twitter API
4. Tweets<- twitterAPI() //tweets downloaded from twitter live data on coronavirus
5. Tweet<-tweet_preprocessing(Tweets) // removes hashtags, usernames, urlsetc
6. Cltweets<-tweets_cleaning(Tweet) // performs tokenization, stemming etc
7. Pos_tweets<-sentiment.positive(Cltweets)
8. Neu_tweets<-sentiment.neutral(Cltweets)
9. Neg_tweets<-sentiment.negative(Cltweets)
10. Coronadataset=Merge(Pos_tweets, Neg_tweets,Neu_tweets)
11. NaiveBayesclassifier<-train(coronadataset)
12. Performance<-test(coronadataset)

III. CONCLUSION

This work was carried on corona virus happening exploitation twitter information from first Apr 2020 to fifth Apr 2020 wherever the virus unfold across many countries and therefore the happening became pandemic. This work helps to know to know perception concerning corona virus and its impact on the general public. the feelings throughout the amount were downloaded and therefore the and therefore the towards the happening was analyzed. Machine learning algorithmic program is applied for information analysis and accuracy of the model is almost seventieth. The folks well understood the govt. policies, safety measures, its symptoms and preventive measures to be taken throughout this era. They well followed and maintained the social distancing and sanitizing ways. This study helps the organizations to know the opinion of individuals throughout the Corona Virus happening. because the virus is spreading smartly, the study has to be dispensed on a weekly basis to possess a more robust understanding on the feelings of the folks.

IV. ACKNOWLEDGMENT

V. REFERENCES

[1] Ahuja, Ravinder& Chug, Aakarsha&Kohli, Shruti& Gupta, Shaurya&Ahuja, Pratyush. (2019). The Impact of Features Extraction on the Sentiment Analysis. Procedia Computer Science. 152. 341-348. 10.1016/j.procs.2019.05.008.

[2] Ajay Bandi and Aziz Fellah. Socio-Analyzer: A Sentiment Analysis Using Social Media Data. Volume 64, 2019, Pages 61–67 Proceedings of 28th International Conference on Software Engineering and Data Engineering.

[3] Ali Hasan, Sana Moin, Ahmad Karim and ShahaboddinShamshirband "Machine learning-based sentiment analysis for twitter accounts", MDPI, 2018.

[4] Alsaeedi, Abdullah & Khan, Mohammad. (2019). A Study on Sentiment Analysis Techniques of Twitter Data. International Journal of Advanced Computer Science and Applications.10.361-374.10.14569/IJACSA.2019.0100248.

[5] Brandon Joyce, Jing Deng. "Sentiment Analysis of Tweets for the 2016 US Presidential Election", in IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA: IEEE (2017).

[6] D T Hermanto et al Twitter Social Media Sentiment Analysis in Tourist Destinations Using Algorithms Naive Bayes Classifier. 2018 J. Phys.: Conf. Ser. 1140 012037

[7] JohnnatanMessias, Joao P. Diniz, Elias Soares, Miller Ferreira,MatheusAraujo, Lucas Bastos, Manoel Miranda, FabricioBenevenuto, Towards Sentiment Analysis for Mobile Devices . 2016

[8] KavyaSuppala and Narasingarao: "Sentiment analysis using Naïve Bayes classifiers", international journal of innovative technology and exploring engineering June 2019.

[9]　I.Korkontzelos,A.Nikfarjam,M.Shardlow,A.Sarker,S. Ananiadou, and G. H. Gonzalez, "Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts," Journal of Biomedical Informatics,vol. 62, pp. 148–158, 2016.

[10] Minchae Song, Hyunjung Park, Kyung-shik Shin"Attention-Based Long Short-Term Memory Network Using Sentiment Lexicon Embedding for Aspect-Level Sentiment Analysis in Korean." Information Processing & Management, 56 (3) (2019), pp. 637- 653

[11] Nels Oscar, Pamela A. Fox, Racheal Croucher, Riana Wernick, Jessica Keune, and Karen Hooker . Machine Learning, Sentiment Analysis, and Tweets: An Examination of Alzheimer"s Disease Stigma on Twitter . J Gerontol B PsycholSciSocSci, 2017, Vol. 72, No. 5, 742–751

[12] Rahman, S. A. El, F. A. AlOtaibi, and W. A. AlShehri. (2019, 3-4 April 2019). "Sentiment Analysis of Twitter Data", in the 2019 International Conference on Computer and Information Sciences (ICCIS).

[13] ShihabElbagir, Jing Yang. Sentiment Analysis of Twitter Data Using Machine Learning Techniques and Scikit-learn.

[14] U. A. Siddiqua, T. Ahsan, and A. N. Chy, "Combining a rule-based classifier with ensemble of feature sets and machine learning techniques for sentiment analysis on microblog," in 2016 19th International Conference on Computer and Information Technology (ICCIT), 2016, pp. 304– 309.

[15] M. Vadivukarassi, N. Puviarasan and P. Aruna. Sentimental Analysis of Tweets Using Naive Bayes Algorithm. World Applied Sciences Journal 35 (1): 54-59, 2017.

[16] Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2015). Combining lexicon- based and learning-based methods for Twitter sentiment analysis. International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE), 89, 1–8.