Analysis of Covid-19 Textual Data using NLP

Ajinkya Ashok Jawale[1]

[1]*Graduate Student University of Pune, Pune, India*

---***---

Abstract - *Covid at first known as Corona Virus Disease of 2019, has been announced as a pandemic by World Health Organization (WHO) on 11th March 2020. Astounding squeezing factors have mounted on each country to make persuading necessities for controlling the general population by assessing the cases and properly utilizing open resources. The speedy number of emotional cases from one side of the planet to the other has become the fear of furor, fear and pressure among people. The mental and genuine sufficiency of the overall people is found to be directly comparative with this pandemic sickness. The current condition has uncovered more than 24 million people being attempted positively worldwide as of 27th August, 2020. Thus, it is critical to execute different measures to secure the countries by demystifying the fitting real factors and information. This paper intends to draw out the way that tweets containing all handles related to COVID-19 and WHO have been unproductive in coordinating people around this pandemic erupt appropriately.*

The spread of Covid-19 has achieved general prosperity concerns. Online media is continuously used to bestow news and experiences about it. A commonsense examination of the situation is essential to utilize resources preferably and appropriately. In this assessment, we perform Covid-19 tweets determination examination using an oversaw AI approach. Recognizing confirmation of Covid-19 ends from tweets would allow instructed decisions for better dealing with the current pandemic condition.

Tweets are eliminated by an in-house built crawler that uses the Tweepy library. The dataset is cleaned using the preprocessing techniques and ends are removed using the TextBlob library. The responsibility of this work is the display appraisal of various AI classifiers using our proposed feature set. This set is formed by connecting the pack-of-words and the term repeat speak report repeat. Tweets are appointed positive, fair-minded, or negative. Execution of classifiers is surveyed on the precision, exactness, audit, and F1 score.

Key Words: Covid-19, natural language processing, sentimental analysis,, medical imaging, deep learning, information extraction

1. INTRODUCTION

Initiated in Wuhan, China, the exponential spread of Coronavirus disease (COVID-19) has caused a public health crisis regionally and internationally as well [1]. The Centers for Disease Control and Prevention (CDC) operated its Emergency Operations Center (EOC), and the World Health Organization (WHO) published its first report regarding the situation about Coronavirus disease 2019 (COVID-19) on January 20, 2020 [2]. The WHO recognized and gave the name “2019-nCOV” to the novel Coronavirus. The seriousness of the COVID-19 pandemic had been underrated until the National Health Commission (NHC) categorized it as a B-type infectious disease officially and took measures to fight against this pandemic on January 20, 2020 [3]. WHO later announced it as a health emergency internationally on January 30, 2020. One and a half months later, on March 11, 2020 COVID-19 is classified as pandemic [4]. Coronavirus disease (COVID-19) is an infective pandemic caused by a newly discovered virus named Corona. Most of the formidable diseases are generated from unhygienic habits. Hygiene measures and sanitation, such as hand washing, could play an important and cost-effective role in reducing the spread of pandemics, such as the COVID-19 [5].

The feeling classification can be carried out by utilizing different methodologies. We can significantly characterize these methodologies in the accompanying three sorts: 1) dictionary based methodology; 2) AI/profound learning approach, and 3) mixture approach. We have utilized an AI way to deal with the assessment of India's normal public in this work. To utilize AI for assumption examination, information pre-handling on crude information is essential since the effectiveness of the calculation utilized is straightforwardly relative to the nature of preparing and testing dataset. In assessment examination, the preprocessing of text is known as normal language handling (NLP).

2. LITERATURE Review:-

Assumption investigation from web-based media information is one of the profoundly arising research fields. It could assume a basic part if there should be an occurrence of health related crises like the COVID-19 pandemic, and subsequently it is more critical. However a great deal of examination from different points on slant arrangement and NLP is as yet in progress, a portion of the finished works are as per the following.

Wu et al. [1] utilized information from December 31, 2019, to January 28, 2020, on the tally of irresistible people traded from Wuhan to deduct the quantity of cases in Wuhan from December 1, 2019, to January 25, 2020. Cases traded intra-nation were then anticipated. They anticipated the COVID cases the nation over by utilizing the flight booking information and COVID positive people went through flight and along these lines guessed the public and global spread of COVID-19 subsequent to ascertaining the effect of the metropolitan-wide isolate of Wuhan and encompassing urban communities which was begun in China from Jan 23–24.

Medford et al. [2] brought tweets identified with COVID-19 and determined the recurrence of watchwords subject to disease avoidance practices, immunization, and racial favoritism. They executed an assessment examination to notice nostalgic valence and prevailing feelings. They performed subjects demonstrating to extricate and investigate hot conversation points over the long run. They removed 126 049 tweets posted by 53 196 unique clients. The recurrence of COVID-19-related tweets unexpectedly expanded from January 21, 2020 forward. The notions are ordered from around half (49.5%) of all presents communicated dread on about 30% communicated shock. The quantity of racial posts firmly coordinated with the recurrence of new instances of COVID-19 positives. The monetary and political impacts of the COVID-19 were the most widely recognized points in conversation.

Li et al. [3] gathered and broke down the Weibo posts from 17 865 dynamic Weibo clients utilizing on the web biological acknowledgment (OER) in view of some AI prognostic models. They assessed word recurrence, supposition marker scores (e.g., despondency, uneasiness, resentment, and joy), and psychological pointers (e.g., social danger appraisal, and life delight) from the extricated posts. To assess the distinctions in a similar gathering they have executed the assessment mining and the matched example t-test previously, then after the fact the affirmation of COVID-19 on January 20, 2020. The outcomes assessed by them mirror that the affectability to social dangers and the negative notions expanded while the scores of positive assumptions and life satisfaction have been diminished.

Pandey et al. [5] tended to the hole between the data and hazard of falsehood by fostering a deep rooted learning model that gives certified data in Hindi, the most regularly utilized nearby language in India. They coordinated with the wellsprings of valid and veritable data, for example, the news given by WHO by utilizing AI and NLP. They noticed a Cohen's Kappa of 0.54 with the best performing mix and was sent in their application.

Kayes et al. [6] gathered 100 000 tweets with the watchword #coronavirus inside Australia. Among these 1 lakh tweets, 3076 contain the watchword "social removing" and #socialdistancing. They utilize 8000 tweets for preparing and approval and 2000 tweets for testing the model. They accomplished an exactness of 83.70% and a F1-Score of 81.62% on the test information. They applied the prepared model on the 3076 tweets that contain the watchword "social separating." They saw that 80% in addition to tweets talking about "social removing" have a positive assessment, as outlined. They inferred that individuals in Australia upheld social separation just as acknowledged it.

Minister et al. [7] played out an investigation to hear the point of view of the understudies in regards to the online method of conveyance of guidelines in light of outrageous local area isolate during the COVID-19 pandemic. They played out this investigation by taking understudies' conclusions in the College of Business and Public Administration (CBPA) of Pangasinan State University, Lingayen Campus. First and foremost, they welcomed every one of the understudies to respond to certain inquiries in regards to the issues they might look at during the online examination. They tracked down that the vast majority of the understudies feel that they may confront a few issues, and a significant number of them were stressed over Internet availability nearby. They presumed that greatest understudies are not ready for online conveyance of guidelines, and in this manner they recommended that an elective method of informative conveyance ought to be given by the establishments so instructive greatness could be kept up with.

Dubey et al. [8] played out an investigation to think about the assessments and feelings present in Indian and US public tweets while they referenced Narendra Modi and Donald Trump, separately. The tweets removed for the assessment mining were presented from April 1 on April 9, 2020. NRC Emotion Lexicon has been utilized to examine feelings and assessments in these tweets. They inferred that 64.53% of tweets referencing Narendra Modi are containing positive assessments, while for Donald Trump, it was 48.71%.

Chen et al. [9] have shown an investigation of the theme with respect to the notice of questionable and non-disputable words identified with COVID-19 on Twitter during the pandemic. They utilized LDA to gather subjects from the disputable and non-dubious tweets removed from Twitter and afterward thought about them through the two arrangements of tweets subjectively. They found that subjects in the questionable tweets are for the most part connected with China, even in the wake of eliminating the watchwords associated with the "Chinese infection" before the investigation, though conversations present in the non-dubious tweets are turning around and battling with COVID-19 in the USA.

Barkur et al. [10] managed Indian residents' conclusions after the Indian government reported the lockdown. For examination, they utilized the web-based media stage Twitter. They inspected the tweets to separate the conclusions of the Indians with respect to lockdown. They removed the tweets utilizing the two as often as possible utilized hashtags: #IndiaLockdown and #IndiafightsCorona from March 25, 2020 to March 28, 2020. They analyzed 24000 tweets for the examination utilizing programming R and created a word cloud that assesses the feelings of the tweets. They found that there was bitterness, dread, cynicism, and nausea about the lockdown; still the positive feelings were available conspicuously in the tweets. They inferred that not set in stone that they needed to lessen the spreading pace of COVID-19 and were dedicated to it.

Alhajji et al. [11] dissected an aggregate of 53 127 tweets from the Saudi residents with respect to COVID-19 and tracked down that the quantities of positive tweets are more noteworthy than negative tweets for practically every one of the actions. They tracked down that the best feelings were available in the strict practices-related advances. They inferred that Saudi Twitter clients have positive conclusions and backing toward the disease control steps in battling with COVID-19, and this steady disposition of Saudi residents brings about the general certainty of the Saudi government. As indicated by them, now and again of pandemic, strict convictions may likewise assume a significant part in planning adherents. They gathered separate tweets on different advances taken by the Saudi government. They separated 9924 tweets after the declaration of the Grand Mosque conclusion and tracked down that 76.72% of the tweets were positive. Also, they gathered tweets for Qatif lockdown, conclusion of schools and colleges, shopping centers, parks, and eatery conclusion measures, sports rivalry suspension hashtags, for the congregational and week by week Friday supplications suspension measure, lastly for cross country time limit measure.

Samuel et al. [12] showed some information about the movement of dread slants over the long run as COVID-19 drew nearer to peak in the USA, utilizing amazing printed examination helped by fundamental text based information representations. They have broken down issues with respect to public suppositions ruminating profound worries about infection and COVID-19, guiding to the distinguishing proof of an expansion in dread and negative estimation. They additionally introduced the utilization of exploratory and expanded literary investigation and text based information representation ways to deal with discovering beginning experiences. At long last, they gave a relative investigation of printed order components utilized in AI applications and showed their significance for tweets of various lengths.

Abd-Alrazaq et al. [15] included around 2.8 million tweets for their investigation. Out of them, 167 073 tweets from 160 829 distinct clients met the consideration models. They broke down the tweets on 12 subjects and afterward gathered them into four primary topics: wellsprings of the infection; its starting point; its impact on populace, nations, and the economy; and strategies for palliating the risk of disease. They tracked down the positive mean conclusion for all aside from two themes: one is casualties brought about by COVID-19, and second is expanded bigotry. They noticed the base mean of 2722 tweets for expanded bigotry and a limit of 13 413 for monetary misfortunes. They additionally tracked down the most noteworthy mean for preferences of 15.4 for monetary misfortunes and the least for movement boycotts and alerts of 3.94.

Burnap et al. [16] developed models to gauge the data stream size and endurance utilizing information recovered from the popular microblogging website Twitter by following the fear monger occasion in Woolwich, London in 2013. They clarified the information stream as the proliferation after some time of data presented on Twitter by means of retweets. They utilized zero-shortened negative binomial and Cox relative risks relapse techniques to ascertain the assessed worth of social, content, and fleeting components of the tweet.

Naiknaware et al. [17] utilized the opinion investigation score technique to anticipate the prevalence of various plans offered by India's administration. They utilized the under seven stages interaction to discover the outcomes: 1) removing significant tweets utilizing Twitter API; 2) tweets preprocessing; 3) putting away the prepared tweets in CSV File design; 4) apply score. The assessment () technique; 5) produce sentence score; 6) dissecting the assumption extremity of each tweet; and 7) get ready outcomes. In light of the sentence score found in the fifth step, they partition the sentence's extremity into positive, negative, and impartial and, subsequently, expressed their expectation.

Wu et al. [18] fostered a clever choice of emotionally supportive network utilizing assessment investigation, support vector machine, and summed up autoregressive restrictive heteroscedasticity (GARCH) demonstrating. This model has been generally used to conjecture time series containing highlights of autocorrelation and heteroscedasticity. They apply GARCH displaying and utilize the outcomes into the SVM model to oblige confounded nonlinear and hilter kilter relations engrafted in whimsical determining. To start with, they physically name polarities for the postings of the informational index. Then, at that point they use feeling examination with a physically commented on informational index to remove highlights from the content composed on the stock gathering and accordingly to anticipate the extremity of different posts naturally. From that point onward, they coordinate the postings of each stock day by day. Then, at that point they utilized the resultant GARCH-SVM model to anticipate arbitrariness in future stock costs. They additionally think about this model exactness, which was 81.82% with the dictionary approach, which produces a precision of 75.58%.

Ding et al. [19] developed a substance level assessment investigation instrument named SentiSW for issue remarks comprising of feeling order and element acknowledgment. The target of the created instrument is to group given remarks into three polarities: positive, negative, and nonpartisan, and to perceive the substance of remarks. They construct informational index physically by clarifying 3000 remarks chose from 231 732 gave remarks taken from various GitHub projects. They utilized a ten times cross-approval method to assess the SentiSW instrument and got 63.98% normal review, 68.71% normal exactness, and 77.19% precision. They physically name 660 remarks by "Individual" and "Task" element to assess element acknowledgment and accomplished an in general 88.73% review, 76.58% exactness, and 75.15% precision.

Pota et al. [20] applied the opinion examination on political tweets utilizing a neural-network-based methodology. They utilized the SemEval-2017 informational collection, which contains 20 633 tweets for preparing and 12 284 tweets for testing the prepared model. Their technique addresses the content by thick vectors, including data of subwords to notice likenesses between words by utilizing morphology and semantics both. Then, at that point, they utilized the convolutional neural organization method and prepared the model dependent on the marked informational index. They applied that model on an assortment of tweets accumulated during the residency of some prior days U.K. General Elections. As per noticed outcomes, they inferred that the CNN approach is better when contrasted with vocabulary based methodologies for order of sentences into positive and negative polarities.

Tomar et al. [21] performed assessment mining on GST utilizing information from Twitter to discover the popular assessment on GST. They develop the informational collection by physically clarifying 1000 tweets in addition to 2000 surveys from the Internet Movie DataBase (IMDB). They train and test the classifier by applying the informational collection in three diverse manners. Model-1 utilized the IMDB informational index for preparing and testing with the K-crease approval measure utilizing K = 10, Model-2 utilized IMDB informational index for preparing and physically clarified informational collection for testing. In conclusion, Model-3 utilized a full informational collection for preparing and testing with the K-overlay approval measure utilizing K = 10. They accomplished a precision of 74.75%, 67.1%, and 72.3% for Model-1, Model-2, and Model-3, separately. Noticing the outcomes, they expressed that if a preparation informational index is developed utilizing various sources rather than single-source, better precision can be accomplished.

Das et al. [22] developed the informational index with the assembled 20 000 tweets utilizing Twitter API and NodeXL programming about GST during its execution in India. The informational collection contains 2006 positive words, 4782 negative words, and words to recognize incorrectly spelled words that as often as possible show up in web-based media. They likewise proffered some thought regarding how the presence of any wistful word in a sentence can change its extremity. Subsequent to executing, they reasoned that the Naïve Bayes classifier is the most mainstream since it is moderately simpler to carry out, yet it is somewhat confounded and beats numerous other convoluted calculations in exhibition.

Li et al. [23] addressed the inquiries like how to sort out and group the circumstance put together data with respect to web-based media and what are the distinctive consistency of features of the spreading size of different kinds of circumstance based information utilizing COVID-19-related discussions on Sina Weibo, the generally utilized microblogging webpage in China (like Twitter). They ordered the information into seven circumstances: alert and counsel, warnings, gifts, passionate help, help-chasing, condemning, and counter bits of gossip. They reasoned that the picked highlights for different sorts of situational information could likewise help the organization in getting sorted out their COVID-19-related data to increment or diminish the reposting of their posts.

Luo [24] utilized a powerless contaminated recuperated (SIR) model to anticipate the existence of COVID-19 in numerous nations and the world. To anticipate the COVID-19 life cycle, he took every day refreshed information about COVID-19 from "Our World in Data" site and retrogressed the numerical SIR model utilizing publically accessible codes from the Milan Batista site. He ran relapse for nations separately and changed it every day with the fresher information. The model then, at that point framed is utilized to anticipate the existence pattern of a full pandemic and afterward build the existence cycle bend. He fitted the information to plot an underlying section of the bend, and the excess fragment is assessed. He anticipated 97%, almost 100%, and totally finishing dates of individual nations and the entire world. They reasoned that this pandemic may end before the finish of November 2020 from the world.

Rao et al. [25] proposed a way to deal with direct the screening of individuals effectively. As indicated by him, accumulate the movement history for certain more broad signs utilizing a versatile based online review. The information consequently gathered can help in the essential screening and early acknowledgment of COVID-19-positive people. Information focuses can be accumulated and refined through the man-made brainpower (AI) model, which can at last assess people who might be positive and order them into four classes: no danger, negligible danger, moderate danger, and high danger of being septic with the Coronavirus. The acknowledgment of the great danger cases would then be able to be isolated on need, subsequently decreasing the likelihood of spread.

Dutta et al. [26] played out an investigation to decide if AI could be utilized to assess how much forecasts about affirmed, negative, delivered, and demise cases are near genuine qualities. They utilized profound learning neural organizations, long transient memory (LSTM), and gated intermittent unit (GRU) for preparing the informational index. Expectation results are then cross-checked by genuine information. They inferred that the consolidated LSTM-GRU model gave similarly better outcomes in anticipating affirmed, negative, delivered, and passing cases.

From the above discussion, undeniably assessment examination is one of the prominent unique strategies. In an especially pandemic situation, feeling assessment may expect an essential part in pushing toward controlling and managing the pandemic by the Indian government. Very few examinations have been represented on suspicion assessment of Indian people over COVID-19. Henceforth, the essential objective of this investigation is to get the extent of people for lockdown in India and people against this. Such advances taken by the public authority may be productive in case people are supporting it. Along these lines, we prescribed a system to get the viewpoints on normal people quickly, or we can say constantly, which can help the public to be unique.

4. PROPOSED APPROACH

In this segment, we have proposed a structure for assessment examination during the COVID-19 pandemic. The structure is displayed in Fig. 2. The different periods of the structure to play out the conclusion examination of lockdown during crown upheavals are as per the following.

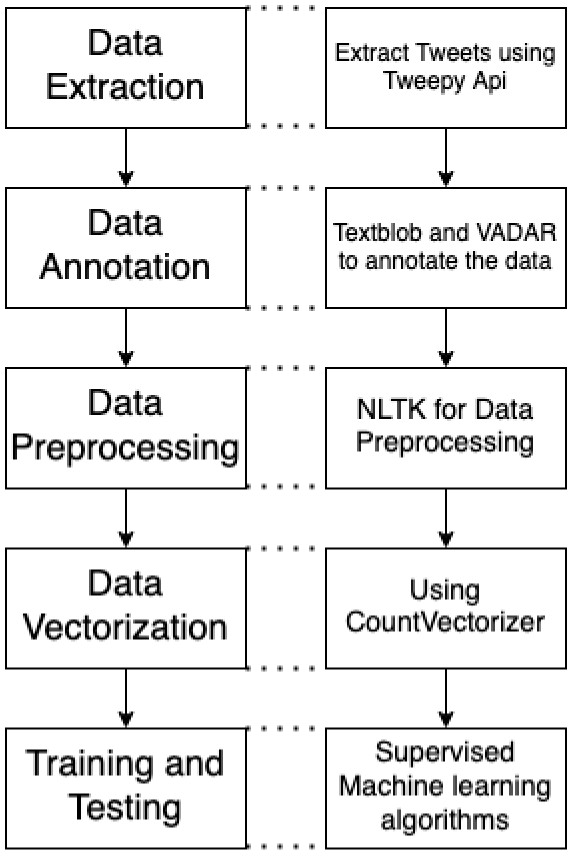

fig: Proposed Approach Process

A. DATA EXTRACTION

Assumption investigation is performed on the lockdown forced by the Indian government from March 25, 2020 to April 14, 2020, during COVID-19. In this manner, the target explicit information are not accessible, and subsequently we have arranged the informational collection physically. Considering the weakening circumstance, conversations of the pandemic via online media have definitely expanded since March 2020 [9]. We have removed 12 741 tweets having the catchphrase "Indialockdown" from April 5, 2020 to April 17, 2020 utilizing Tweepy.

B. DATA LABELING

After the tweets assortment, we have utilized the accompanying methodology displayed in Fig. 3 to name the tweets as certain, impartial, and negative. We have created each tweet's extremity utilizing the TextBlob library and VADER (Valence Aware Dictionary for sEntiment Reasoning) apparatus of the Python. Then, we have taken the crossing point of TextBlob and VADER results to solidify the polarities. After this progression, we are left with 7284 tweets having 3545 with positive extremity, 2097 with unbiased extremity, and 1642 with negative extremity.

C. DATA PREPROCESSING

The information we have gathered may hold some unsought and assessment less words like connections, Twitter-explicit words, for example, hashtags (begins with #) and labels (begins with @), single letter words, numbers, and so forth These sorts of words can assume the part of commotion in our classifier preparing and testing. To correct classifier productivity, it is important to eliminate commotion from the named informational collection prior to taking care of the classifier.

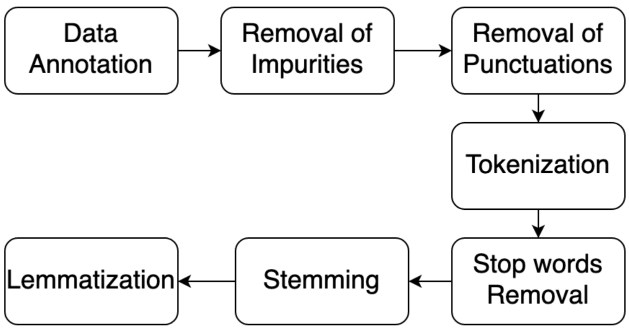

fig: Data preprocessing.

Our pre-handling module isolates clamor from the named informational collection [21]. The means of pre-handling are displayed in Fig. 4. In this progression, we carried out a module to eliminate the above-determined contaminations, changed over the informational index into an information edge, and afterward executed expulsion of string accentuations, tokenization, and evacuation of English stop words, stemming, and lemmatization.

D. VECTORIZATION

The AI classifiers can't take the information written in any language aside from numbers. Consequently, prior to utilizing the content information for prescient displaying, it is needed to change over it into highlights. We have utilized the CountVectorizer highlight extractor to figure word frequencies. CountVectorizer tallies the recurrence of each word present in the report and makes a meager lattice, as displayed in Table I. For instance, Doc1: "She was youthful the manner in which a genuine youngster is youthful." CountVectorizer will change over this content into the accompanying inadequate network with a file of the words in sequential request as follows:{“she”: 4, “was”: 6, “young”: 8, “the”: 5, “way”: 7, “an”: 1, “actual”: 0, “person”: 3, “is”: 2}.

E. TRAINING & TESTING THE CLASSIFIERS

After feature extraction of the preprocessed data set, we have passed the data to machine learning classifiers. We have used eight classifiers (Multinomial NaiveBayes, Bernoulli NaiveBayes, LogisticRegression, LinearSVC, AdaBoostClassifier, RidgeClassifier, PassiveAggressiveClassifier, and Perceptron) for this purpose. We have used 80% data for training and 20% data for testing the classifiers.

In this part, the outcome examination of the relative multitude of classifiers dependent on exactness, accuracy, review, F1-Score, and collector working qualities (ROC) bends with various grams has been talked about. Alongside that, we have approved our informational index for each model with unigrams, bigrams, and trigrams utilizing k-overlap cross-approval strategy with k=10 . Exactness can be determined utilizing (1) and (2) which is the one out of numerous metric units for assessing classifiers. It is characterized as the quantity of right forecasts over the complete number of expectations.

Accuracy=number of correct predictions/total number of predictions.(1)

Technically, we can understand accuracy in terms of positives and negatives as

Accuracy= True Positives+True Negatives/total number of predictions.(2)

where the all out number of expectations is the amount of genuine positives, genuine negatives, bogus positives, and bogus negatives. Genuine positives are the quantity of right forecasts of the positive class. Also, the quantity of right expectations of the negative class is known as evident negatives, the quantity of mistaken forecasts of the positive class is known as bogus positives, and the quantity of wrong forecasts of the negative class is known as bogus negatives. Fig. 5 portrays the pictorial portrayal of positives and negatives.

		Actual Class	
		Positive	**Negative**
Predicted Class	**Positive**	True Positive	False Positive
	Negative	False Negative	True Negative

Fig. Representation of positives and negatives.

K-fold cross-validation is the best method to test the effectivity of the AI model. The re-testing approach of the k-overlay cross-approval procedure is a lot of valuable in estimating the productivity of any AI model with a restricted measure of info information. In this method, the informational collection is partitioned into k equivalent parts or overlap and afterward any one overlay is utilized as a testing set and rest k−1 folds are utilized for preparing the model and the cross-approval score for this specific change is recorded. This cycle gets rehashed for k occasions having another overlap as a testing set and rest as a preparation set. Then, at that point the mean of the scores of the relative multitude of changes is determined which is the last cross-approval score of a model. We have determined the cross-approval score of each model with unigrams, bigrams, and trigrams.

Precision, recall, and F1-Score are other metrics to evaluate the models. The proportion of positive identifications that actually belong to the positive class is known as precision, which is calculated using (3). Recall can be calculated using (4) which shows the number of positive predictions that are identified correctly out of all positive examples.

Precision= True Positives/(True Positives+False Positives)(3)
Recall = True Positives/(True Positives+False Negatives) (4)
F1−Score = 2∗(Precision∗Recall)/(Precision+Recall) (5)

5. LIMITATIONS & CHALLENGES

The most well-known test about the notion investigation of the composed content is that we can't disregard the significance of the preparing of normal dialects. The precision and execution of the investigation are straightforwardly corresponding to the granularity of the informational index, which is built get-togethers on account of assumption examination. We need to handle numerous anomalies, variety, and subjectivity in the information while managing the normal language. The significant restrictions of this work are that we have taken the tweets of a particular stage during the lockdown, yet with the difference in stages, the encompassing conditions may get changed; consequently, the slants of people in general could likewise be changed. We have not thought about the suppositions in emojis and hashtags of the tweets, as we trust it could hamper classifiers' proficiency. Clients can post negative estimation with positive hashtags and the other way around, and they can likewise utilize wrong emojis in mockery. In this work, we have utilized two dictionaries to clarify our informational index, utilizing more could make the informational collection more granular.

6. CONCLUSIONS

Assessment investigation of regular dialects itself contained a huge extension to chip away at, and because of wellbeing crises, this work is additionally showed with a wide scope of future degrees. Future investigations can think about the tweets before the beginning of the principal lockdown and after the finish of the last and can show the progressions in opinions of individuals in the two cases and their outcomes. The variables which can influence mental solidness during pandemics can likewise be considered, and the investigation of the effect of phony news on the general population can likewise assume a significant part in helping the organization and policymakers in controlling the circumstance. According to the specialized perspective, future examinations can hope to work on the precision of the model and can investigate an enormous corpus.

REFERENCES

[1] 1.J. T. Wu, K. Leung, and G. M. Leung, “Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modeling study,” Obstetrical Gynecolo. Surv., vol. 75, no. 7, pp. 399–400, Jul. 2020.

[2] R. J. Medford, S. N. Saleh, A. Sumarsono, T. M. Perl, and C. U. Lehmann, “An ‘infodemic’: Leveraging high-volume Twitter data to understand public sentiment for the COVID-19 outbreak,” medRxiv, Jan. 2020, doi: 10.1101/2020.04.03.20052936.

[3] S. Li, Y. Wang, J. Xue, N. Zhao, and T. Zhu, “The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users,” Int. J. Environ. Res. Public Health, vol. 17, no. 6, p. 2032, Mar. 2020.

[4] WHO Statement Regarding Cluster of Pneumonia Cases, WHO, Wuhan, China, 2020.

[5] R. Pandey, “A machine learning application for raising WASH awareness in the times of COVID-19 pandemic,” 2020, arXiv:2003.07074. [Online]. Available: http://arxiv.org/abs/2003.07074

[6] A. S. M. Kayes, M. S. Islam, P. A. Watters, A. Ng, and H. Kayesh, “Automated measurement of attitudes towards social distancing using social media: A COVID-19 case study,” Tech. Rep., Oct. 2020.

[7] C. K. Pastor, “Sentiment analysis on synchronous online delivery of instruction due to extreme community quarantine in the Philippines caused by Covid-19 pandemic,” Asian J. Multidisciplinary Stud., vol. 3, no. 1, pp. 1–6, Mar. 2020.

[8] A. D. Dubey, “Decoding the Twitter sentiments towards the leadership in the times of COVID-19: A case of

USA and india," SSRN Electron. J., Apr. 2009, doi: 10.2139/ssrn.3588623.

[9] L. Chen, H. Lyu, T. Yang, Y. Wang, and J. Luo, "In the eyes of the beholder: Analyzing social media use of neutral and controversial terms for COVID-19," 2020, arXiv:2004.10225.

[10] G. Barkur, Vibha, and G. B. Kamath, "Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India," Asian J. Psychiatry, vol. 51, Jun. 2020, Art. no. 102089.

[11] M. Alhajji, K. A. Al, M. Aljubran, and M. Alkhalifah, "Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19," Dept. Social Behav. Sci., College Public Health, Temple Univ., Philadelphia, PA, USA, Tech. Rep., doi: 10.20944/preprints202004.0031.v1.

[12] J. Samuel, A. GG, M. Rahman, E. Esawi, and Y. Samuel, "Covid-19 public sentiment insights and machine learning for tweets classification. Nawaz and Rahman, Md. Mokhlesur and Esawi, Ek and Samuel, Yana," Information, vol. 11, no. 6, pp. 1–22, Apr. 2020, doi: 10.3390/info11060314.

[13] R. Liu, Y. Shi, C. Jia, and M. Jia, "A survey of sentiment analysis based on transfer learning," IEEE Access, vol. 7, pp. 85401–85412, 2019.

[14] N. Kaka, "Digital India: Technology to transform a connected nation," McKinsey Global Inst., India, Tech. Rep., Mar. 2019.

[15] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, and Z. Shah, "Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study," J. Med. Internet Res., vol. 22, no. 4, Apr. 2020, Art. no. e19016.

[16] P. Burnap, "Tweeting the terror: Modelling the social media reaction to the woolwich terrorist attack," Social Netw. Anal. Mining, vol. 4, no. 1, p. 206, Dec. 2014.

[17] B. R. Naiknaware and S. S. Kawathekar, "Prediction of 2019 Indian election using sentiment analysis," in Proc. 2nd Int. Conf., Aug. 2018, pp. 660–665.

[18] D. D. Wu, L. Zheng, and D. L. Olson, "A decision support approach for online stock forum sentiment analysis," IEEE Trans. Syst., Man, Cybern. Syst., vol. 44, no. 8, pp. 1077–1087, Aug. 2014.

[19] J. Ding, H. Sun, X. Wang, and X. Liu, "Entity-level sentiment analysis of issue comments," in Proc. 3rd Int. Workshop Emotion Awareness Softw. Eng., Jun. 2018, pp. 7–13.

[20] M. Pota, M. Esposito, M. A. Palomino, and G. L. Masala, "A subword-based deep learning approach for sentiment analysis of political tweets," in Proc. 32nd Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA), May 2018, pp. 651–656.

[21] N. Tomar, R. Srivastava, and J. K. Verma, "Analysing public sentiment on GST implementation in India," in Proc. Int. Conf. Comput., Power Commun. Technol. (GUCON), Sep. 2018, pp. 1101–1106.

[22] S. Das and A. K. Kolya Sense GST:, "Text mining & sentiment analysis of GST tweets by naive Bayes algorithm," in Proc. 3rd Int. Conf. Res. Comput. Intell. Commun. Netw. (ICRCICN), Nov. 2017, pp. 239–244.

[23] L. Li, "Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on Weibo," IEEE Trans. Comput. Social Syst., vol. 7, no. 2, pp. 556–562, Mar. 2020.

[24] J. Luo. (2020). When Will COVID-19 End Data-Driven Prediction.

[25] R. AS and V. JA, "Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantin," Infection Control Hospital Epidemiol., vol. 1, pp. 1–5, Jan. 2020.

[26] B. SK and S. Dutta, "Machine learning approach for confirmation of COVID-19 cases: Positive, negative, death and release," medRxiv, vol. 2, pp. 172–177, Jan. 2020, doi: 10.5281/zenodo.3822623.Shibly KH, Dey SK, Islam MT, Rahman MM. COVID faster R-CNN: A novel framework to Diagnose Novel Coronavirus Disease (COVID-19) in X-Ray images. Inform Med Unlocked. 2020;20:100405. doi:10.1016/j.imu.2020.100405

[27] E.E.D. Hemdan, M.A. Shouman, M.E. Karar COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in x-ray images (2020), p. 11055 arXiv preprint arXiv:2003

[28] Song, Ying, et al. "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images." IEEE/ACM Transactions on Computational Biology and Bioinformatics (2021).

[29] Wang, S., Kang, B., Ma, J. et al. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). Eur Radiol 31, 6096–6104 (2021).

[30] Zheng, Chuansheng, et al. "Deep learning-based detection for COVID-19 from chest CT using weak label." MedRxiv (2020).

[31] Xu, Xiaowei, et al. "A deep learning system to screen novel coronavirus disease 2019 pneumonia." Engineering 6.10 (2020): 1122-1129.