Applied Soft Computing Journal 97 (2020) 106754

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media

Koyel Chakraborty [a], Surbhi Bhatia [b], Siddhartha Bhattacharyya [c,*], Jan Platos [d], Rajib Bag [a], Aboul Ella Hassanien [e]

[a] Department of Computer Science and Engineering, Supreme Knowledge Foundation Group of Institutions, Mankundu, India
[b] Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Saudi Arabia
[c] Department of Computer Science and Engineering, CHRIST (Deemed to be University), Bangalore, India
[d] Faculty of Electrical Engineering and Computer Science, VSB Technical University of Ostrava, Czech Republic
[e] Faculty of Computers and Artificial Intelligence, Cairo University, Egypt

ARTICLE INFO

Article history:
Received 23 July 2020
Received in revised form 14 September 2020
Accepted 21 September 2020
Available online 28 September 2020

MSC:
00-01
99-00

Keywords:
Tweets
COVID-19
Sentiment analysis
Deep learning
Emotional intelligence
WHO
Fuzzy rule
Gaussian membership function

ABSTRACT

COVID-19 originally known as Corona VIrus Disease of 2019, has been declared as a pandemic by World Health Organization (WHO) on 11th March 2020. Unprecedented pressures have mounted on each country to make compelling requisites for controlling the population by assessing the cases and properly utilizing available resources. The rapid number of exponential cases globally has become the apprehension of panic, fear and anxiety among people. The mental and physical health of the global population is found to be directly proportional to this pandemic disease. The current situation has reported more than twenty four million people being tested positive worldwide as of 27th August, 2020. Therefore, it is the need of the hour to implement different measures to safeguard the countries by demystifying the pertinent facts and information. This paper aims to bring out the fact that tweets containing all handles related to COVID-19 and WHO have been unsuccessful in guiding people around this pandemic outbreak appositely. This study analyzes two types of tweets gathered during the pandemic times. In one case, around twenty three thousand most re-tweeted tweets within the time span from 1st Jan 2019 to 23rd March 2020 have been analyzed and observation says that the maximum number of the tweets portrays neutral or negative sentiments. On the other hand, a dataset containing 226,668 tweets collected within the time span between December 2019 and May 2020 have been analyzed which contrastingly show that there were a maximum number of positive and neutral tweets tweeted by netizens. The research demonstrates that though people have tweeted mostly positive regarding COVID-19, yet netizens were busy engrossed in re-tweeting the negative tweets and that no useful words could be found in WordCloud or computations using word frequency in tweets. The claims have been validated through a proposed model using deep learning classifiers with admissible accuracy up to 81%. Apart from these the authors have proposed the implementation of a Gaussian membership function based fuzzy rule base to correctly identify sentiments from tweets. The accuracy for the said model yields up to a permissible rate of 79%.

1. Introduction

COVID-19 is not just an infectious disease which is transmitted through contact and by small droplets produced when people cough, sneeze or talk, it is now becoming a source of depression, stress and anxiety because of misleading information posted on social media. The mental health is directly affected because of the rapid spread of false information on social media. With the current situation of lockdown and social distancing, the prime dependency of individuals is on Internet and the highest activity has been reported [1] on social media. The statistics clearly shows the graph in Fig. 1 depicting the increase in data usage of internet globally.

Social media has become a huge part of our life. It connects people to the outer world. Social media provides a way to showcase our lives, discretely, conveniently and on our own terms. People rely more on the posts and tweets shared on the social networking sites like Twitter®, Facebook®, and Instagram®. It

* Corresponding author.
E-mail address: siddhartha.bhattacharyya@christuniversity.in
(S. Bhattacharyya).

https://doi.org/10.1016/j.asoc.2020.106754
1568-4946/

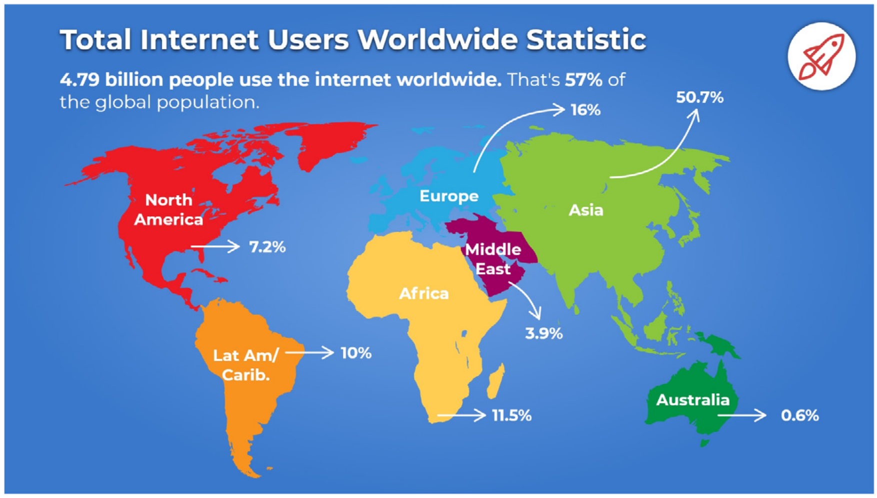

Fig. 1. Internet users in the World: 2020 [1].

is anticipated that social media should guide people in getting correct and authentic information on Corona cases, but analyzing the posts, it has been seen that most of them have misled people by posting false data and figures.

Social media is not allowing people to get through this disaster; rather the tweets and opinions on COVID-19 are becoming dangerous and a cause of concern which needs to be raised to handle misleading information from different sources. This paper focuses on the fact that people should stop sharing popular data in social media as they might harm the overall impact during emergencies. People should be sensible enough to share those data which might be of help to anybody in general. Premium agencies should take special care and attention in deploying fact checkers through their media so that huge amount of unwanted information can be barred from spreading in crisis periods.

This paper deals with an extensive analysis of the sentiments emoted through the tweets since the beginning of this year pertaining to the novel COVID-19. The research begins with the collection of tweets from various COVID-19 related handles followed by extensive cleaning of the same. Two sets of datasets have been used in this work. While one dataset contains all the tweets that have been published during the time span between December 2019 to May 2020, priority has been given to the second set of tweets that have been retweeted the most. Further, the sentiments of the tweets have been labeled and WordCloud is presented for every sentiment to show the ineffectiveness of social media in this time of distress. The most frequent words used in the texts have also been found out mathematically to portray how people are being led to nowhere following tweets doing their rounds in Twitter®.

The motivation of the paper lies in alerting the society that the extensive usage of social media worldwide needs to be restricted as it is becoming instrumental in disseminating useless information just as the pandemic had spread into humanity. The novelty of the work lies in proposing a model to deploy a Gaussian fuzzy rule based technique to evaluate the sentiments expressed in the tweets. As a contribution to the society, this paper has demonstrated with facts that how it is shocking that though people are initially sharing positive and neutral data but further users are re-tweeting those tweets which are negative in nature. This paper emphasizes with substantial proof the need to employ "monitoring mechanisms" to prevent negative psychology from being disseminated within the minds of the social media users. The work done in the paper also relates to the title as how instead of the magnanimous popularity of social media, it is leading to the dwindling of tweets and for which the same cannot be considered as a trusted source. State-of-the-art deep learning classifiers have been applied on the word vectors and doc2vec models and have been tested to find out the model yielding the best accuracy. The implemented model gives an accuracy of above 75% on both the datasets used for this purpose.

Other than the above, this paper proposes the implementation of fuzzy logic for taming the fuzziness of sentiments. As fuzzy sets are ideally suited to counter the ambiguities in life, the authors have proposed the initial integration of fuzzy logic in effectively handling the sentiment identification of tweets. Fuzzy logic has proved to be efficient in handling intricate problems efficiently in diverse fields of business.

To be precise, the present work is relevant in the present context as regards to the following approaches incorporated.

(a) While labeling the sentiments, real values have been considered while classifying them to positive, negative and neutral sentiment scores. Through this, the raw data is being handled rationally to unearth the sentiments of the social media users.

(b) A fuzzy rule based model has been proposed to handle the uncertainty prevalent in the raw sentiments which would otherwise get truncated while rounding up the values during labeling into various classes of positive, negative and neutral sentiments. Gaussian membership function has been used to characterize the fuzziness of the model and has been compared to the popular triangular membership function.

(c) We have implemented various state-of-the-art deep learning classifiers to extract the actual sentiment of social media users during this pandemic.

The rest of the paper is organized as follows. Section 2 provides a background of the work selected elucidating the basics of social media, the motivation behind this work, a brief introduction on sentiment analysis and emotional intelligence and related works. Section 3 details the results obtained by implementation of the state-of-the-art classifiers along with those obtained with the proposed fuzzy rule based technique for determining sentiments. The experimental results have been described in Section 4. Section 5 presents the discussions on the methods implemented. Finally, Section 6 draws a conclusion to the paper with future directions of research.

2. Background

With the Corona Virus cases multiplying day by day, researchers are putting tremendous efforts by developing novel rapid point-of-care diagnostics to control the spread. The unknown nature and the volatility of the situation keep on edge, wondering individuals what will come next. This situation can create panic and make individual feel afraid, overwhelmed, and helpless. While the threat is real, fear and having our emotions run amok will make the situation even worse. Uncertainty and anxiety go hand-in-hand, according to experts at the Yale Center [2] for Emotional Intelligence (CEI), and that is why the many unknowns about the Corona Virus pandemic, when cases will peak, when schools will reopen, when it will be safe to visit loved ones, are creating widespread anxiety. In fact, people should adhere to strategies that can help mitigate anxiety as they are socially distancing and are briefed with constant pandemic updates. In a series of webinars beginning March 25, 2020 CEI experts have been addressing ways of maintaining emotional health, regulating emotions, and developing resilience using emotional intelligence strategies.

2.1. Social media vs. misleading information

It is to admit that pandemics such as Corona Virus or COVID-19 happen once in a lifetime, and hence the methods or measures to tackle the same are yet undefined. While some countries have been successful in combating the outbreak, there are some countries which have failed miserably to tackle the grave situation. Pertinent to the times that we are residing in, it is inevitable that social media will play an important part in our lives [3]. In the serious times of social distancing, where the governments have imposed serious lockdowns and have made it mandatory that no people should move out of their houses even for the slightest of needs, it would have been better if social platforms would guide us in the right direction in these times of need. Contrary to expectations, it is being perceived that people have often indulged in sharing inappropriate content or misinformation through social media platforms.

Instead, it would be right if meaningful data could be shared so that people, who are turning into social media for appropriate content, could stay updated with the latest trends in developments regarding this dangerous disease that has grasped the entire world. It is alarming that these inaccurate materials are being shared in even wider circles leading to mental harassment of the people in general. It can be said that the advent of this novel Corona Virus has opened doors for an entire new set of problems posing social media to be a nightmare. Dangerous techniques of spreading cures for the disease, deceptive declarations, oily trading pitches and various plots and schemes are doing the rounds of social media at a risky pace.

2.2. Motivation and justification of the proposed work

The social media platform at this moment which could prioritize the sharing of meaningful content is Twitter®. Twitter® is one of the most trendy micro blogging sites, which is considered as a crucial depository of sentiment analysis [4]. Netizens tweet their expressions within allotted 140 characters. This work is conducted with two different datasets, the first one comprising all the unique tweets that have been tweeted during the phase of the pandemic from December 2019 to May 2020. A total of 2, 26, 688 tweets have been collected from IEEE Dataport [5]. For the second case, data from Twitter® have been collected for this work between the span of 1st January, 2020 to 23rd March, 2020. For this, a total of 31, 50, 26, 574 tweets was downloaded using Tweet Binder [6]. The share of tweets per day has been amassing during the mentioned period as is evident from various internet sources. While initially there had been less engagement in social media evident by the lower number of tweets that had surfaced, but eventually the rise has been alarming. It is anticipated from the smartest of all beings on earth, that people will obviously be uploading or sharing information that is of meaning and that which is accurate, so that it comes of help to the people following them on social media. The unique dataset has hence been analyzed and presented to support the claim that though people uniquely are spreading positive vibes pertaining to the pandemic but the tweets those are negative or neutral in nature are being re-shared most.

Hence, complete unique tweets and the tweets that have at least 1,000 re-tweets based on the logic that accurate and not popular messages are being seen, shared or followed at par with outbreak of Corona Virus in the world, have been considered in this paper.

The first dataset contains distinctive 2, 26, 668 tweets crawled through ‘corona’, ‘covid’, ‘sarscov2’, ‘covid19’ and ‘coronavirus’ keywords. Coming to the second dataset, 23, 000 tweets have been hence considered which has a minimum of 1, 000 re-tweets. The handles used to extract the tweets have been #covid19 OR #coronavirus OR coronavirus OR #covid-19. It has been seen that the most retweeted users in Twitter® are @realDonalTrump, @WHO etc.

The motivation behind this work has been to portray the fact that how irrationally people are behaving in the grim times of a pandemic. This particular pandemic seems to have claimed millions of lives and is leading the world to an utter recession in terms of economy. It would have been easier for the victims at this time to have gathered some structured information from social media. This call in for a strict quality check of the tweets to ensure that meaningful content are shared in these most used social networking sites. The outcome of this paper will lie in the initiation of fact checking implied on social sites before its wide sharing, so that false and inappropriate news can be prevented from being disseminated within netizens.

2.3. Sentiment analysis and emotional intelligence

Sentiment analysis is the gathering of people’s views regarding any event happening in real life. In such situations in which the world is currently going through, understanding the emotions of the people stands extremely important. The grave scenario wherein people cannot go out of their houses demands exploring what the people is actually being thinking about the whole scenario. Hence, the authors have planned this work on understanding the demanding situation especially on social media [7].

Emotional Intelligence (EQ or can be also be abbreviated as EI) on the other hand is the skill to appreciate, supervise and perceive the mental situation of fellow humans so that it can be entangled with your own to be able to converse efficiently, alleviate anxiety and resolve clashes [8]. EQ brings self control in oneself, awareness of strengths and weaknesses, empathizing with co-workers and helps to manage healthy relationships.

Although sentiment analysis and emotional intelligence are used synonymously, they are not the same. While sentiment analysis depends primarily on the data to categorize expressions as positive, negative or neutral, EQ further probes into the subtleness of the emotions articulated through the comments. EQ is much more difficult and multifaceted than sentiment analysis. For example, for a particular comment, sentiment analysis will check whether the same is positive, negative or neutral, but emotional intelligence will further check whether the comment leads to sadness, dissatisfaction, or sarcasm if the comment is found out to be negative. Thus, EQ dives deep into the categorization done by sentiment analysis tools

2.4. Related works

As Novel Corona Pneumonia (NCP) reports a major warning to the international population health, cooperation is required by all the countries to combat it [9]. The media and other social media channels should ethically present the relevant and correct reports to increase motivation among the general public instead of presenting biased information; such coverage may only serve to divide individuals and stoke fear. Sural et al. [10] demonstrated that Trait Emotional Intelligence (TEI) is directly associated with Problematic Social Media Use (PSMU) and indirectly associated with motives of presenting a popular side and passing time. This result interpreted as those individuals who are lower on TEI use PSMU as a coping strategy to deal with their real life troubles. A sample of 444 individuals aged between 18 to 43 years and having an active social media account were taken. The statistical analysis was carried out with SPSS 23 and AMOS 23 software and for path analysis, maximum likelihood estimation was used. Direct and Indirect relationships were analyzed by using the bootstrapping method with 5000 bootstrap samples and 95% bias-corrected confidence intervals. Hornung et al. [11] presented the relationship between EI and Facebook® use. A sample of 105 individuals (60 female and 44 male participants) was taken. The sample's average age is 25.53 years and average number of Facebook® friends is 365.97. First, the measurement model of the directly observed indicators was assessed from the questionnaire (such as Use and four EI dimensions) and second, the latent variable scores for the four EI dimensions were assessed by the higher-order constructs as well as the structural model. The research stated that the relationship between EI and Facebook® use is very positive for the younger group and very negative for the older group. A younger group uses Facebook® more and is accustomed to it as they grow up with social media. They possibly develop their EI through or along with social media use and through social media networks. Herodotou et al. [12] examined the role of trait Emotional Intelligence (trait EI) for play and frequency of gaming over a sample of 1051 young adult US/European gamers (96% males and 4% females). Chen et al. [13] investigated the influencing factors of the three types of user social media intentions. A sample of 502 social media users including 52% females and 48% males aged between 21 to 35 years was taken for data analysis through online survey. Partial Least Squares (PLS) analysis and estimation were performed in two phases. The first phase conducted reliability and validity analysis, whereas the second phase estimated and verified the path coefficients of the structural model which concluded that social media marketing activities have a significant influence on three types of intentions i.e. continuance intention, participation intention and purchase intention. Kim et al. [14] examined the relationships between narcissism, the Big 5 personality traits, the need for popularity, the need to belong, and various types of selfie posting behaviors—posting solo selfies, selfies with a group, and editing selfies. A sample of 260 participants was taken. The study found that selfie behaviors are associated with various personality traits and psychological needs. The results of this study suggested that a range of interpersonal motivations as well as egocentric traits underlie individuals' selfie activities. Depoux et al. [15] discussed the necessary measures to be taken in order to combat the pandemic of social media panic. The paper emphasized to use social media wisely so that proper use of digital technologies can overcome the problem and help people balance their mental health using supportive measures from health ministry and solidarity during the stressful period of quarantine. Merchant et al. [16] explained the importance of disseminating information through various social media platforms during the outbreak of Corona Virus 2019. It discussed that transferring correct and reliable information is rather important for the people in case of crises as social media is the diagnostic tool and the utmost referral system. People believe more on posts on when to get tested, what to do with the results, and where to receive care. So, navigating wrong and misleading information can sometimes be hazardous to someone's life. It discussed the importance of communication as it is directly proportional to people's belief and reaction accordingly. Merchant et al. [17] highlighted the epicenter of infodemic with social media news. The study discussed the picture of the COVID-19 epidemic and COVID-19 infodemic with the main objectives to create a mathematical model at the intersection of the above listed major studies. Li et al. [18] explored the impacts of COVID-19 on mental health of people by conducting various tests on sentiment analysis using social networking sites. It was demonstrated by experiments that the knowledge gaps of short-term individuals change in psychological conditions after the outbreak. The work used Online Ecological Recognition (OER), which referred to the automatic recognition of psychological profile (e.g., anxiety, well-being, etc.) by using predictive models [19] based on ecological behavioral data from Weibo. Mohammad et al. [20] discussed the Arabic tweets and conducted the Naïve Bayes algorithm to perform sentiment analysis. The results showed positive responses towards the disease and this will further help in overcoming the situation. Talwar et al. [21] discussed the dark side of social media and linked the association between fake news and social media. The conceptual model was proposed based on ten hypotheses and powerful constructs were listed behind rational theories from psychology. The implications will help the society and researchers to get aware of the factors that are positively and negatively associated with fake news sharing behavior. The study was limited to data collection from one country and enhancement on asymmetrical relations on constructs is the future work of the paper. Sharma et al. [22] conducted an in-depth analysis on varied topics consisting of sentiments and trends with identifying false information and checkpoints in spreading false information on pandemic COVID-19 through Twitter®. The paper leaves open marking annotation for misinformation to improve classification techniques. Ghafarian et al. [23] discussed on identifying informative tweets by using distributional assumption. Each crisis-related tweet is considered as a "distribution". Remarkable results were achieved in identifying informative tweets about a crisis incident. For considerate social interventions, even the government has taken mandatory steps and has created awareness in response to COVID-19: (1) To implement social distancing to mitigate the spread by deploying effective social distancing strategies. A non-pharmaceutical intervention reduces human contact within the population [24] and therefore constrains the spread of COVID-19. Data science can support contact tracing and monitoring of social distancing, for instance by extracting social media data and using language processing techniques [25]. (2) To control misleading information and Online Harms. Different strategies and measures are taken by public to control the spread of misinformation [26] which has potentially dangerous outcomes. Instances are there where researchers have implemented models to track the development trend of the pandemic as discussed by Ahmad et al.in [27]. For example, online rumors accusing 5G deployments for causing COVID-19 led to mobile phone masts being attacked in the UK [28]. Wikipedia maintains an up-to-date list of misinformation surrounding COVID-19. This confirms the spread of a number of dangerous forms of misinformation, e.g., that vinegar is more effective than hand sanitizer against the Corona Virus. Naturally, users who believe in such misinformation could proceed to undermine public health. One important use case would therefore be to develop classifiers and techniques to stem this flow. For example, the study in [29] demonstrates testing simple interventions to reduce the spread of COVID-19 misinformation.

An infodemic observatory analyzing digital response in online social media to COVID-19 has been created by the CoMuNe lab at Fondazione Bruno Kessler (FBK) institute in Italy and is available online. The observatory uses ML techniques based on Twitter® data to quantify collective sentiment, social bot pollution, and news reliability and displays them visually. The end note comes out that the rapid spread of misinformation is undermining trust in vaccines and is crucial to public health [30]. Researchers are trying hard to come up with measures to detect the disease through various methods. One work has been reported in [31], where the authors have proposed architecture to easily identify the infected state of patient from chest X-ray images. The model aims to classify a patient as a positive or a negative COVID-19 affected person. Table 1 presents a tabular format of some of the recent works done in to COVID-19.

3. Proposed methodologies

3.1. State-of-art classifier based proposed model

To make sense of the innumerable tweets being posted in social media per second, a model has been implemented that will successfully recognize the sentiments emoted through the tweets. Sentiment analysis is one of the best possible methods to be able to derive expressed emotions from unstructured texts by transforming the data into a structured format. The detailed model is illustrated in Fig. 2.

The model aims to classify sentiments into positive, negative and neutral scores. Natural language Toolkit (NLTK) library has been used which acts as an appropriate text processor for language dealings. For clarification purposes, the first dataset that contains approximate 2 lakh distinct tweets from December 2019 to May 2020 is named DATA_SET 1 and the second dataset that contains the most retweeted tweets from January 2020 to March 2020 is named DATA_SET 2. Irrespective of the dataset, the first step post crawling of data from Twitter® is to clean the tweets for processing. Initially, duplicate rows or similar tweets are eliminated from the Comma Separated Values (CSV) file containing the tweets. The tweets are then cleaned to eradicate redundant symbols which are generally associated with tweets. Symbols like @, RT, #, URLs, numeric values and punctuation marks are cleaned by using the "re" python module.

It is to keep in mind that building a model for classification initially needs finding of relevant features that are present from the text available in the tweets. Hence, while training the model, the tweets can be broken down into words and be appended to the feature vector. If one word is added, the approach is termed as "unigram", for two words it is "bigram" and for three words it is "trigram", respectively.

Post the essential cleaning of the tweets, SentiWordNet (SWN) lexical resource has been used which is responsible for assigning sentiment scores to words. The Parts_Of_Speech tagging is done first to make it compatible to be implemented by SWN. The entire sentiment score of the sentence is calculated by summing up the total polarity based on positive and negative scores. A rise in the value of positive_score indicates the high level of positivity of the cleaned tweet. In this case, a cleaned tweet is identified positive if the summation of the sentiment total is more than 0, negative if it is less than 0 and neutral if the value is equal to 0 as shown in Table 2.

Then the total numbers of positive, negative and neutral tweets are calculated by two methods, viz., TextBlob and Afinn [37]. While TextBlob is an efficient word library to carry out Natural Language Processing tasks, Afinn is a word list which has been specially designed for microblogs like tweets. It is generally observed that TextBlob produces normalized scales in contrast

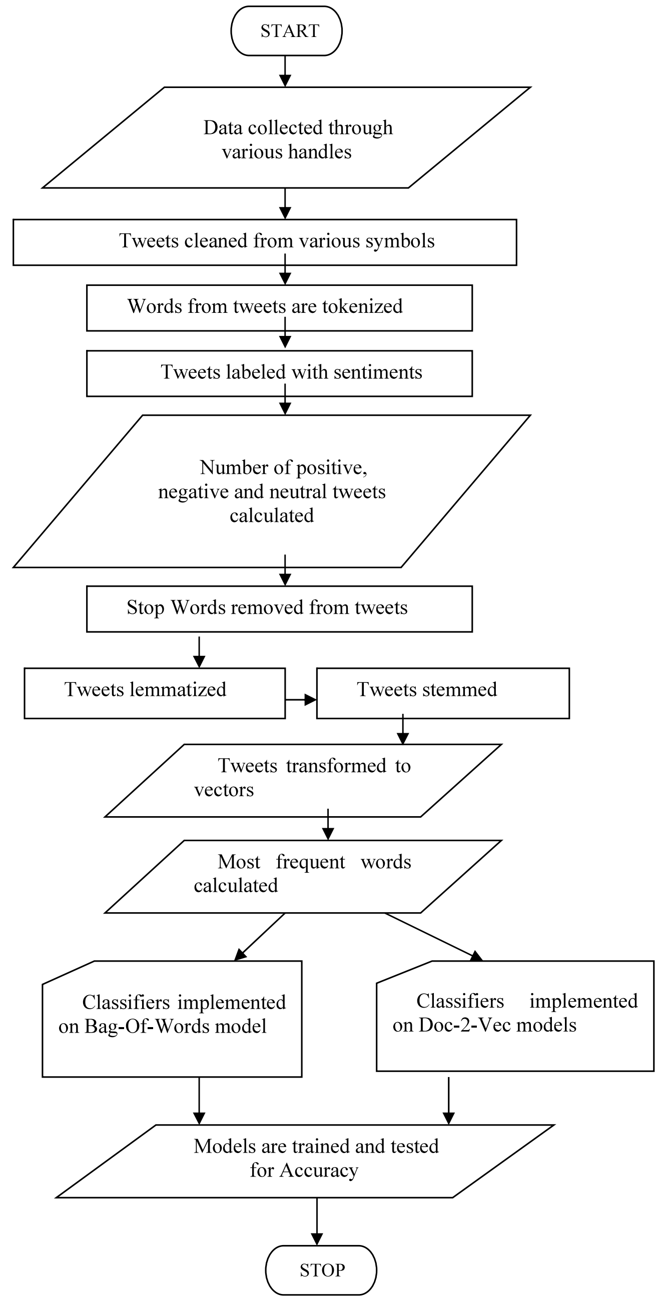

Fig. 2. Overall structure of the state-of-art classifier based proposed model.

to Afinn [38]. Table 3 shows the number of positive, negative and neutral tweets labeled by TEXTBLOB and AFINN methods for both the datasets. For DATA_SET 1, there is a maximum number of positive and negative tweets, but to our utter shock, we found out that the number of negative or neutral tweets is remarkably more in the re-tweeted DATA_SET 2 by both the processes that have been used for the same. Fig. 4 shows the categories of positive, negative and neutral tweets labeled by TextBlob and Afinn methods for DATA_SET 1 and DATA_SET 2, respectively.

The figures clearly depict that people showed confused behavior in social media during the pandemic outbreak and portrayed emotions likewise. While the unique tweets result in positive, but in both the cases, the number of neutral tweets are remarkably high, hence emoting uncertainty over the state of the disease. After the polarity of the tweets has been determined, the stop-words have been cleared in the tweets. Post this, the tweets have

Table 1
Comparison of few of the relevant works pertaining to the pandemic.

Ref no.	Motivation and aim of the work	Datasets used	Methods used in the work and Obtained Accuracy	Limitations
[32]	To trail the growth of panic amongst Twitter® users based on a specific keyword	Tweets from Twitter® API was extracted using R language. Around 9 lakhs tweets were analyzed in this work.	Naïve Bayes and Logistic Regression classifiers have been used. The accuracy for shorter tweets was obtained as 91% and 74%, respectively.	The paper analyzed the sentiments based on a single keyword tracking based on only the fear of the people based in USA. Further aspects barring geographical constraints could also be explored.
[33]	To analyze the sentiments of Indians post lockdown imposed by the government	Datasets was collected using Twitter® application interface by R. Approximately 24 thousand tweets were used by extracting from the handles #IndiaLockdown and #IndiafightsCorona within the time span of 25th to 28th March, 2020.	Analysis was done by the help of software R and by using WordCloud only.	Very few tweets were considered of a particular country. The study portrayed that Indians took the strategy of the government positively on imposing the lockdown.
[34]	To trail the out the economic, political and health related impact on the people as envisaged through CoronaTracker website.	Data was collected from the CoronaTracker website between 22nd January and 3rd March, 2020.	Susceptible-Exposed-Infectious-Recovered (SIER) model was used to predict the outbreak of the disease as the main idea of the work was to guess and estimate the death and recovery rates using analytical modeling.	The news extracted for analyzing dates up to the beginning of March, 2020. Data was collected from a website only.
[35]	To extract an exact idea by detecting the primary topics tweeted by netizens related to COVID-19 pandemic.	Around 1,68,000 tweets were considered for grouping them into various topics.	Tweets were analyzed by unigrams and bigrams and were influenced by Dirichlet allocation for helping in topic modeling.	Though twelve topics were identified to be posted by the users during the span of February '20 to March '20, but they emphasized on the sentiments of the topics. These were mostly related to health care issues.
[36]	To study how the Chinese Weibo users were affected emotionally on and after 20th January, 2020.	Weibo messages were used from available 17,865 active users of this platform within the time span of 13th to 26th January, 2020.	Paired sample *t*-test by SPSS (Statistical Product and Service Solutions) was used to measure the emotional features of the Weibo users.	As Weibo users are mostly youth, the results were anticipated to be biased. The study shows how a rise in negative sentiments occur post the outbreak of the pandemic in the minds of the youth.

Table 2
Samples showing sentiment scores for each tweet from DATA_SET 2.

S No.	TWEETS from DATA_SET 2	Positive_score	Negative_score	Neutral_score
0	UNITED STATES: Man in his sixties dies after trying to self-medicate against corona virus by taking chloroquine phosphate in the form of fish tank cleaner	0	0.125	−1
1	LIVE: Press Briefing with Corona virus Task Force https://t.co/Ciwe6T4Dr2	0	0	0
2	The U.S. Chamber of Commerce and the heads of major corporations have lobbied the administration against using the Defense Production Act. Trump and Larry Kudlow, as well as Jared Kushner, were persuaded by those arguments, administration officials said. https://t.co/3YbYFnvcB8	2	0.25	1
3	This priest, Suffering from the corona virus, gave up his ventilator to give to a younger patient and has died. His name was Don Giuseppe Berardelli. He was 72 years old, from Bergamo. https://t.co/3N1ONtxFjO	0.625	0.5	1
4	JUST IN: More than 100,000 people have now recovered from the corona virus	0.25	0	1

been stemmed and lemmatized to obtain the original forms of the words that have been retrieved from the tweets. While stemming performs its activity by eliminating frequently used prefixes and suffixes, lemmatization works by changing the words based on its morphological information. For example, 'studies' and 'studying' will be 'studi' and 'study' for stemming respectively, but for lemmatization, both will be changed to 'study'. The changing of cases of all tweets to lower case, removing of stop words, stemming and lemmatizing processes are done to transform the words to vectors to make it compatible for the various classifiers.

The claim of the authors that social media has been unable to provide proper course in which the netizens shall combat a pandemic like COVID-19, has been validated by the WordCloud presented in Figs. 3(a) and 3(b). The majority of the words that have been portrayed in each of the sentiments has been visualized using the WordCloud modules. These too display words

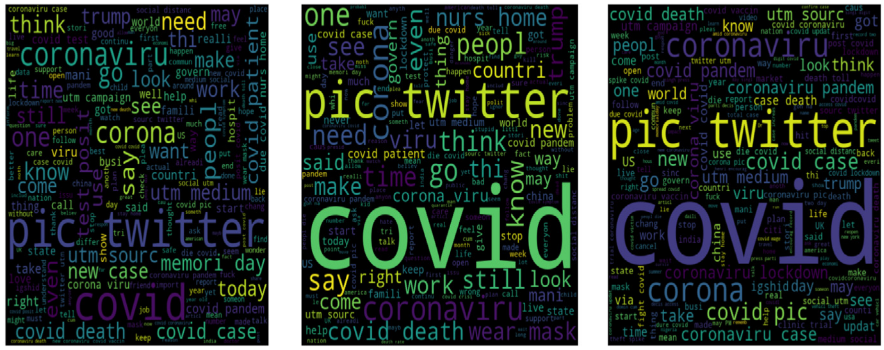

Fig. 3(a). Most used words in Positive, Negative and Neutral sentiments have been represented by these three Word Clouds in DATA_SET 1.

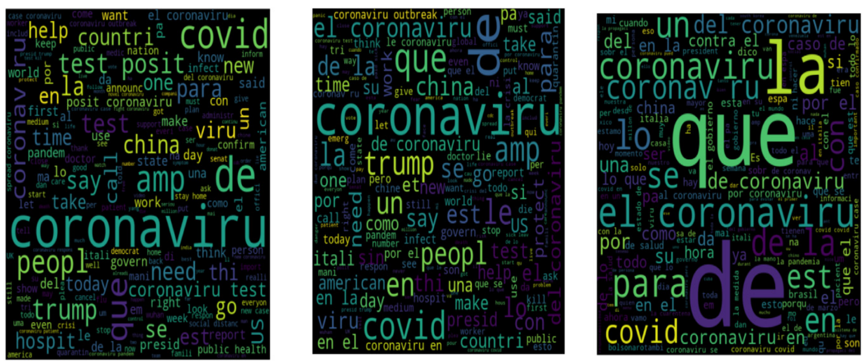

Fig. 3(b). Most used words in Positive, Negative and Neutral sentiments have been represented by these three Word Clouds in DATA_SET 2.

Table 3
Comparative table representing positive, negative and neutral tweets labeled by TEXTBLOB and AFINN methods for both the datasets.

Dataset	DATA_SET 1		DATA_SET 2	
Methods used	TEXTBLOB SENTIMENT LABELING	AFINN SENTIMENT LABELING	TEXTBLOB SENTIMENT LABELING	AFINN SENTIMENT LABELING
Total tweets in the dataset	226 673	226 673	22 985	22 985
Total tweets with sentiment	226 673	226 673	22 985	22 985
No. of positive tweets	98 995	66 481	6728	4604
No. of negative tweets	53 414	101 607	3597	9273
No. of neutral tweets	74 264	58 585	12 660	9108

that do not prove any efficiency in representing a viable solution during emergencies.

To embed a word into the semantic region, the Bag-of-Words vectorizers, namely Count Vectorizer and Tfidf Vectorizer from the sklearn library were used providing the result shown in Figs. 4(a) and 4(b). This process is crucial as data to be fed in the different models has to be in a mathematical format.

3.2. Proposed fuzzy rule based model

Mathematically, fuzzy logic is a type of multiple valued logic in which the truth values range from 0 and 1 [39]. The values might include in themselves both 0 and 1. It had been envisaged to cater to cases which involve partial truths emerging in uncertain situations. In an analogy, the same uncertainty which arises while determining sentiments from texts has been mapped in this approach. Fuzzy rule based approaches depend on the selection of membership functions and its intervals to depict the inherent system fuzziness. It is to be kept in mind that the range of the values of the membership functions should always be within 0 and 1. Though there is a plethora of techniques by which membership functions can be envisaged [40], the most widespread use is that of the triangular membership function. Apart from the triangular membership function, another membership function that is preferred for its efficiency to model human reasoning is the Gaussian membership function. Gaussian membership function (MF) [41] depends on two of its parameters, the mean of the data and its standard deviation. This method has been implied to present an alternative to the much used triangular membership function. The model implements a fuzzy rule based system to determine the sentiment of a tweet with the help of Gaussian MF.

The proposed model models the uncertainty in the sentiment analysis system as a fuzzy system, which is used to predict the nature of sentiments depending on the fuzziness in the positive and negative scores. The fuzzy inputs to the model, viz., positive score and negative score are characterized by the Gaussian membership functions (LOW, MEDIUM and HIGH), whereas, the fuzzy output sentiment is characterized by the Gaussian membership functions (NEGATIVE, NEUTRAL and POSITIVE). The model is guided by a set of seven disjunctive fuzzy rules to determine the

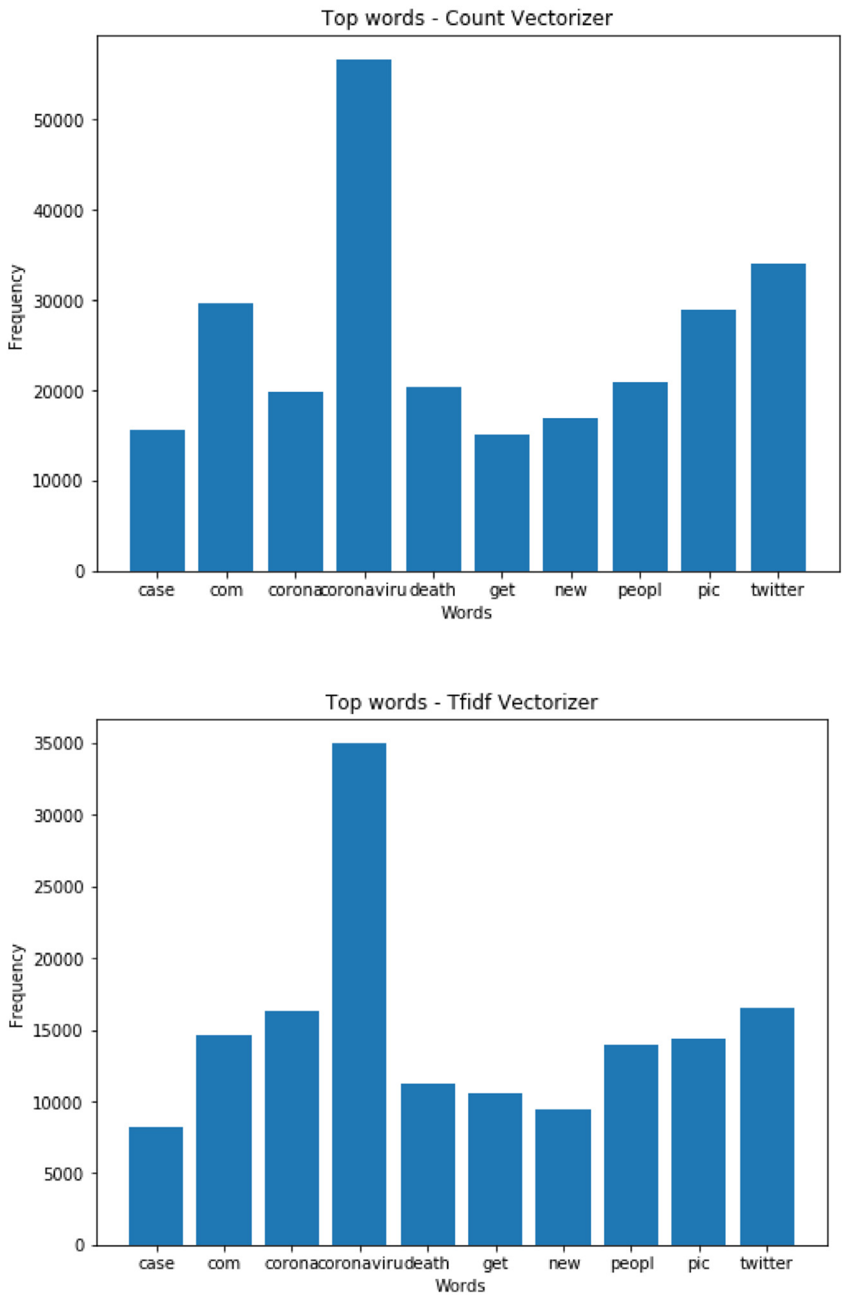

Fig. 4(a). Top words of DATA_SET 1 as shown in both the vectorizer methods fail to show some impactful direction towards the disease or its combating methods.

output sentiment. The following steps describe the operation of the proposed fuzzy rule based model.

Step 1: Limited sized social media texts are retrieved and preprocessed.

Step 2: VADER sentiment lexicon is used to label the data into three classes, viz., positive, negative and neutral based on their polarity scores.

Step 3: Mamdani [42] style fuzzy inference technique is deployed to process each text.

(a) The input variables are fuzzified.

(b) Inference rules are evaluated.

The following seven proposed inference rules based on Mamdani fuzzy inference mechanism characterize the model.

(i) Rule 1 — IF positive score is LOW AND negative score are LOW, THEN sentiment is NEUTRAL

(ii) Rule 2 — IF positive score is HIGH AND negative score is LOW, THEN sentiment is POSITIVE

(iii) Rule 3 — IF positive score is MEDIUM AND negative score is MEDIUM, THEN sentiment is NEUTRAL

(iv) Rule 4 — IF positive score is LOW AND negative score is MEDIUM, THEN sentiment is NEGATIVE

(v) Rule 5 — IF positive score is HIGH AND negative score is MEDIUM, THEN sentiment is POSITIVE

(vi) Rule 6 — IF positive score is LOW AND negative score is HIGH, THEN sentiment is NEGATIVE

(vii) Rule 7 — IF positive score is HIGH AND negative score is HIGH, THEN sentiment is NEUTRAL

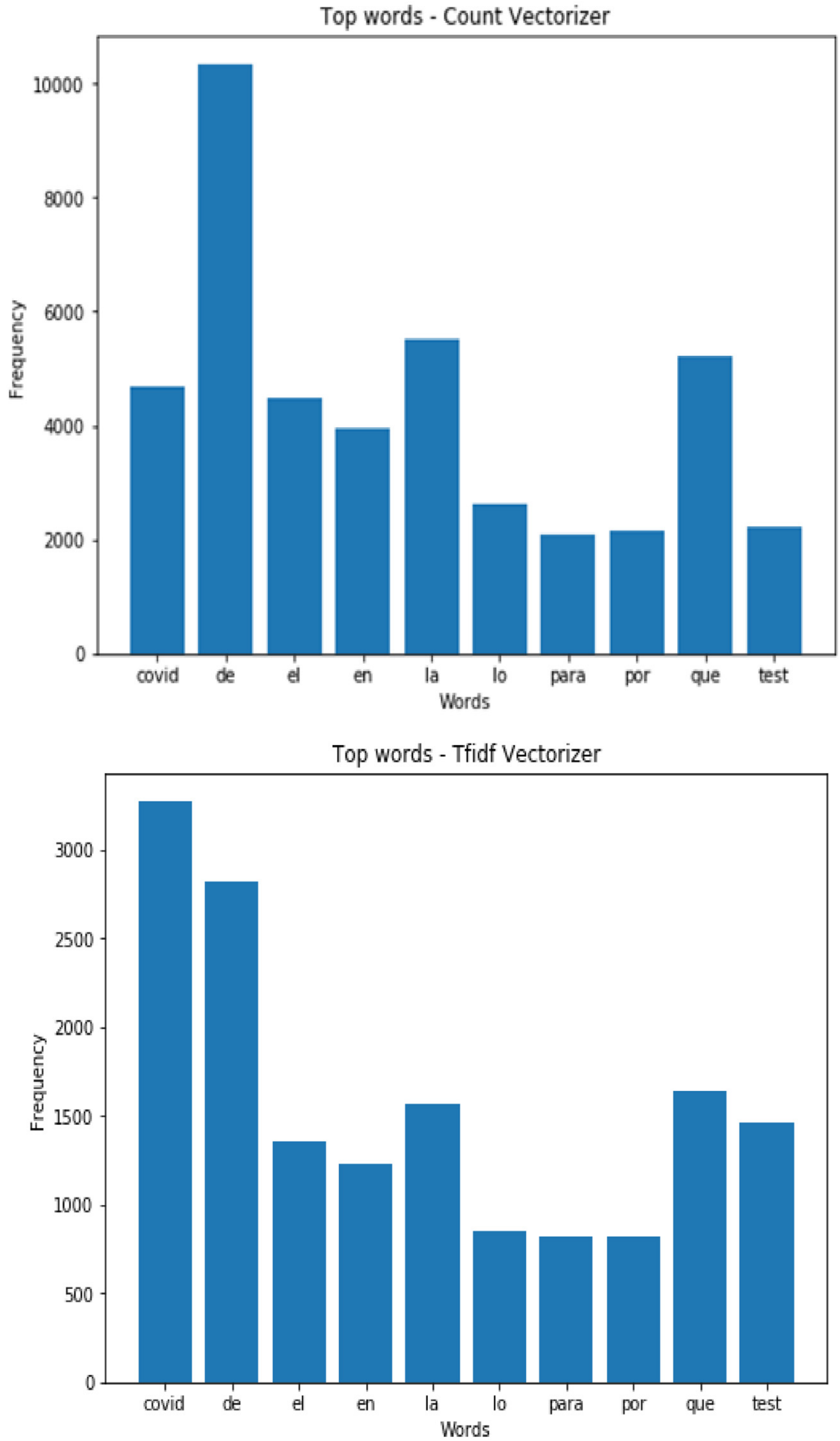

Fig. 4(b). Top words of DATA_SET 2 as shown in both the vectorizer methods fail to show some impactful direction towards the disease or its combating methods.

(c) Finally the crisp output sentiment is obtained by means of defuzzification of the fuzzy sentiment output of the model using the centroid defuzzification method [43].

Following are the samples of some individual sentiment scores of each tweet.

TWEET - 22 945 texas a&m confirms possible case of coronavirus: https://t.co/xb7277snxz #tamu https://t.co/klawqovih2 {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

Positive Score for each tweet: 0.0

Negative Score for each tweet: 0.0

TWEET - 22 946 please read this and stay safe everyone? #coronavirusoutbreak #coronavirus https://t.co/ek6tbwpziq {'neg': 0.0, 'neu': 0.606, 'pos': 0.394, 'compound': 0.6369}

Positive Score for each tweet: 0.4

Negative Score for each tweet: 0.0

TWEET - 22947 en españa se conoce un caso de "coronavirus". se llaman borbones. {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

Positive Score for each tweet: 0.0

Negative Score for each tweet: 0.0

TWEET - 22948 would you look at that.

Table 4
Comparison of performance of proposed Gaussian membership function based rule based system and Triangular membership function based fuzzy rule based system.

Inference system	F1-Score	Precision	Recall	Log Loss	Mean absolute error	Sensitivity
Gaussian membership based fuzzy inference system	**79%**	**82.13%**	**84.56%**	0.49	**0.16**	**91%**
Triangular membership based fuzzy inference system	78%	80.45%	82.61%	**0.37**	0.19	93%

not only is chlorine dioxide (aka "mms") an effective cancer cell killer, it can wipe out coronavirus too.

no wonder youtube has been censoring basically every single video where i discuss it over the last year.

big pharma wants you ignorant. https://t.co/7cqmyuxcxy {'neg': 0.228, 'neu': 0.718, 'pos': 0.054, 'compound': -0.872}

Positive Score for each tweet: 0.1

Negative Score for each tweet: 0.2

TWEET - 22952 progressive media; forget the fact they carry coronavirus, bats are a valuable source of protein. cows are killing the planet.

it's normal in other countries, and you're just ignorant.

eat the bat, bigot.

https://t.co/iessl2owvm {'neg': 0.207, 'neu': 0.716, 'pos': 0.077, 'compound': -0.6486}

Positive Score for each tweet: 0.1

Negative Score for each tweet: 0.2

The defuzzified output sentiment comes out to be 4.59 rendering the sentiment of the document to be "Neutral".

The proposed model has been further compared to the much popular Triangular membership function based rule based system [44]. The results shown in Table 4 portray an improvement in the precision, recall and F-score when a comparison is made between the models under consideration. Further improvement in the comparative results may be effected by incorporating more membership grades in order to enhance the performance of the proposed methodology.

From Table 4, it is evident from the values of F1-score, Precision and Recall that the proposed Gaussian membership based fuzzy rule base system for determination of the sentiments outperforms its Triangular counterpart. Moreover, the proposed system exhibits less sensitivity towards change of inputs, thereby indicating the stability of the system. In addition, the Mean Absolute Error value also stands as an indicator of the efficiency of the system, although the Log Loss measure is lower for the Triangular counterpart. It may be mentioned here that Log Loss seems to calculate only a comprehensive measurement of the performance and hence is harder to decipher compared to the accuracy of the model. So based on the other available metrics like recall, precision and F-score it can be claimed that the proposed model provides better result in predicting sentiments from tweets.

Fig. 5 shows the confusion matrix arising out of the proposed model. It is evident from Fig. 5 that neutral sentiments are more predominant compared to positive and negative sentiments.

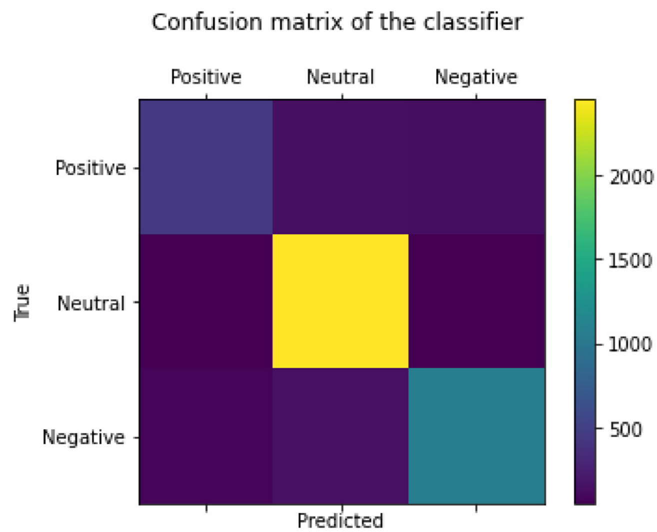

Fig. 5. Confusion matrix arising out of the proposed model.

4. Experimental results

Datasets for this experiment have been obtained through #corona, #covid19, #coronavirus, coronavirus and #covid-19 since the inception of this year 2020. DATA_SET 1 contains around 2, 26, 668 tweets whereas the preliminary tweets which were collected for DATA_SET 2 stood up to 31, 50, 26, 574. But, as mentioned earlier, the tweets with minimum 1000 retweets were considered for this experiment. Post this screening, approximate 23,000 tweets were taken for further processing of DATA_SET 2. Finally to fit the model, the data have been categorized into train, validate and test sets. 90% data from training set, 5% from validation set and rest 5% from the training set have been used. Maximizing the training part is to prioritize the number of tweets in the dataset. To show the accuracy of the implemented model, unigram, bigram and trigram have been performed considering both the vectorizers mentioned in the previous section. N-grams [46] are defined as all the potential combinations of the contiguous words that are present in the tweets. While unigram defines single words, bigrams consider two adjacent words, trigrams consider three adjacent words. After the texts have been transformed into vectors, classification algorithms are executed. It has been observed that for self-created analyser models, there is no perfect classification algorithm that exists. This work contains hyper-parametric classifiers from Naïve Bayes Models [48], Ensemble models [49], Support Vector Machine Models [50], Linear Models [51] viz., Multinomial Classifier, Bernoulli Classifier [52], AdaBoost Classifier [53], LinearSVC Classifier [51] and Logistic Regression Classifier [54]. The Naïve Bayes classifier models are very effective in making predictions for sentiment analysis and are based on Bayes' Theorem [55].

Bayes' Theorem can be described as

$$P(ab) = (P(b|a) * P(a))/P(b) \tag{1}$$

where, $P(a|b)$ is termed as the posterior probability, $P(b|a)$ results in the probability of the data present b, provided that the hypothesis a is correct. $P(a)$ is claimed to be the prior probability of a being correct and is independent of the data and $P(b)$ is the probability of the data independent of the hypothesis a. Examples of Naïve Bayes classifiers dealing with disconnected characteristics include Multinomial Naïve Bayes and Bernoulli Naïve Bayes classifiers. While the former symbolizes occurrences with events produced by a multinomial, the later consists of occurrences which are self-regulating binary variables symbolizing inputs [55]. Ensemble models works on the idea to unite the forecasts of several classifiers in an attempt to choose the optimal solution from multiple classifiers generated for the same problem [56]. The AdaBoost Classifier is one of the most popular boosting enabled ensemble algorithms. The instances within the data are weighted to ultimately classify the predictions which

Table 5
Results of all the classifiers implemented with Bag-of-Words [45] models on DATA_SET 2.

	Vec_Gram	Classifier	Accuracy	crossval_train_ score_mean	crossval_test_ score_mean	crossval_train_ score_std	crossval_test_ score_std	Time taken (s)
1	cv_1	MultinomialNB()	0.614	0.689	0.589	0.002	0.013	11.63
2	cv_1	LinearSVC()	0.723	0.964	0.721	0	0.007	16.68
3	cv_1	AdaBoostClassifier()	0.615	0.623	0.617	0.003	0.010	53.27
4	cv_1	RandomForest Classifier	0.659	0.988	0.637	0	0.007	93.82
5	cv_1	LogisticRegression()	0.701	0.843	0.712	0.001	0.006	12.13
6	cv_2	MultinomialNB()	0.621	0.862	0.618	0.002	0.008	32.65
7	cv_2	LinearSVC()	0.726	0.999	0.728	0	0.008	42.38
8	cv_2	AdaBoostClassifier()	0.597	0.62	0.6	0.001	0.009	146.52
9	cv_2	RandomForest Classifier	0.635	0.988	0.619	0	0.011	412.55
10	cv_2	LogisticRegression()	0.661	0.801	0.601	0.001	0.010	43.13
12	cv_3	MultinomialNB()	0.615	0.91	0.622	0.001	0.007	58.65
13	cv_3	LinearSVC()	0.719	0.999	0.722	0	0.007	74.61
14	cv_3	AdaBoostClassifier()	0.594	0.62	0.584	0.003	0.012	272.64
15	cv_3	RandomForest Classifier	0.611	0.988	0.607	0	0.012	1004.73
16	cv_3	LogisticRegression()	0.619	0.849	0.642	0.002	0.006	92.18
17	tf_1	MultinomialNB()	0.579	0.638	0.555	0.001	0.011	12.14
18	tf_1	LinearSVC()	0.717	0.993	0.71	0	0.008	77.74
19	tf_1	AdaBoostClassifier()	0.642	0.625	0.619	0.001	0.011	38.38
20	tf_1	RandomForest Classifier	0.656	0.987	0.638	0	0.011	126.08
22	tf_1	LogisticRegression()	0.701	0.943	0.712	0.002	0.011	62.15
23	tf_2	MultinomialNB()	0.598	0.844	0.584	0.001	0.009	55.11
24	tf_2	LinearSVC()	0.728	0.999	0.725	0	0.008	288.85
25	tf_2	AdaBoostClassifier()	0.645	0.623	0.616	0.002	0.012	219.44
26	tf_2	RandomForest Classifier	0.641	0.986	0.621	0	0.008	779.30
27	tf_2	LogisticRegression()	**0.75**	0.987	0.712	0	0.015	312.13
28	tf_3	MultinomialNB()	0.601	0.914	0.596	0.001	0.009	93.92
29	tf_3	LinearSVC()	0.729	0.999	0.723	0	0.009	448.69
30	tf_3	AdaBoostClassifier()	0.645	0.623	0.616	0.002	0.012	379.94
31	tf_3	RandomForest Classifier	0.612	0.985	0.615	0.001	0.011	727.36
32	tf_3	LogisticRegression()	0.711	0.973	0.722	0	0.014	500.12

Table 6
Results of all the classifiers implemented with Bag-of-Words [45] models on DATA_SET 1.

	Vec_Gram	Classifier	Accuracy
1	cv_1	MultinomialNB()	0.675
2	cv_1	LinearSVC()	0.764
3	cv_1	AdaBoostClassifier()	0.625
4	cv_1	RandomForest Classifier	0.721
5	cv_1	LogisticRegression()	0.760
6	cv_2	MultinomialNB()	0.680
7	cv_2	LinearSVC()	0.772
8	cv_2	AdaBoostClassifier()	0.640
9	cv_2	RandomForest Classifier	0.710
10	cv_2	LogisticRegression()	0.715
12	cv_3	MultinomialNB()	0.674
13	cv_3	LinearSVC()	0.781
14	cv_3	AdaBoostClassifier()	0.650
15	cv_3	RandomForest Classifier	0.741
16	cv_3	LogisticRegression()	0.698
17	tf_1	MultinomialNB()	0.729
18	tf_1	LinearSVC()	0.793
19	tf_1	AdaBoostClassifier()	0.678
20	tf_1	RandomForest Classifier	0.770
22	tf_1	LogisticRegression()	0.790
23	tf_2	MultinomialNB()	0.760
24	tf_2	LinearSVC()	0.780
25	tf_2	AdaBoostClassifier()	0.620
26	tf_2	RandomForest Classifier	0.750
27	tf_2	LogisticRegression()	0.809
28	tf_3	MultinomialNB()	0.773
29	tf_3	LinearSVC()	0.780
30	tf_3	AdaBoostClassifier()	0.645
31	tf_3	RandomForest Classifier	0.760
32	tf_3	LogisticRegression()	0.814

Table 7
Results of all the classifiers implemented with Doc2Vec [47] models for DATA_SET 2.

	Model	Classifier	Accuracy	Time taken (s)
1	DBOW	BernoulliNB()	0.534	0.20
2	DBOW	LinearSVC()	0.616	21.61
3	DBOW	AdaBoostClassifier()	0.576	14.77
4	DBOW	LogisticRegression()	0.621	12.57
5	DBOW+DMM	BernoulliNB()	0.555	0.17
6	DBOW+DMM	LinearSVC()	0.625	41.71
7	DBOW+DMM	AdaBoostClassifier()	0.551	29.68
8	DBOW+DMM	LogisticRegression()	0.637	19.40
9	DBOW+DMC	BernoulliNB()	0.526	0.16
10	DBOW+DMC	LinearSVC()	0.622	36.35
11	DBOW+DMC	AdaBoostClassifier()	0.548	29.34
12	DBOW+DMC	LogisticRegression()	**0.628**	14.12

are finally combined to make the final prediction. Support Vector Machine models are mainly used in classification problems where the appropriate hyper-plane is computed to efficiently divide separate classes of data. It provides excellent results in transforming a non separable problem to a separable one based on the labels that have been characterized. In general, LinearSVC classifiers prove efficient for text data classification cases. Linear models on the other hand formulate a forecast by implementing a linear function of the input characteristics. One example of linear model is the logistic regression, which works on categorical data as its target variable. Amongst all these, the challenge is to create a classifier that provides optimum accuracy in the model. Here, K-fold Cross Validation [57] is also used as a resampling technique to check the steadiness of the model. Initially, the Bag-of-Words [45] models have been considered which has been further extended by the more intricate Doc2Vec models. Table 5 shows the results obtained with all the classifiers by implementing the Bag-of-Words models for DATA_SET 2. Table 6 shows the accuracy obtained by all the state-of-the-art classifiers being implemented on DATA_SET 1. In both the tables, Vec_Gram denotes the combination of the vectorizer used along with the n-gram used. For example, cv_2 represents Count Vectorizer [58] with bigrams and tf_1 represents tfidf vectorizer is implemented with the unigram range. For DATA_SET 2, Logistic Regression Classifier gives the highest accuracy of 75% with bigrams under

Table 8
Comparison of both the datasets implemented with Doc2Vec [47] model.

Dataset	Model	Classifier	Accuracy
DATA_SET 1	DBOW	LinearSVC()	0.57
	DMC	BernoulliNB()	0.47
	DMM	LinearSVC()	0.56
	DBOW+DMM	LinearSVC()	0.60
	DBOW+DMC	LinearSVC()	0.57
	Model	Classifier	Accuracy
DATA_SET 2	DBOW	LogisticRegression()	0.62
	DMC	AdaBoostClassifier()	0.43
	DMM	AdaBoostClassifier()	0.52
	DBOW+DMM	LogisticRegression()	0.64
	DBOW+DMC	LogisticRegression()	0.63

the Tfidf Vectorizer [58]. In case of DATA_SET 1, the highest accuracy of 81% is obtained through Logistic Regression with trigrams under the Tfidf Vectorizer. But, as the datasets has unequal numbers of positive, negative or neutral tweets, a Random Forest Classifier [59] has also been used to create a balance.

A more complicated method which sustains the semantic association between words, the Doc2Vec model is used to check the accuracy of the model. Three algorithms belonging to Doc2Vec [47], viz., Distributed Bag-of-Words (DBOW) [60], Distributed Memory Concatenated (DMC) [61] and Distributed Memory Mean (DMM) [62] are implemented further. Combinations of the different Doc2Vec models have also been made to check for the model with the best accuracy. The detailed accuracy with the models implemented for DATA_SET 2 has been presented in Table 7. It is observed that the Logistic Regression classifier working on the DBOW+DMC model gives the best result as far as time taken for execution is concerned.

A comparative analysis showing the results of implementation of all the classifiers that gives the best accuracy on each of the Doc2Vec models has been shown in Table 8 for assuring that the model works the same for both the datasets. DATA_SET 1 is ten times the size of DATA_SET 2 and yet the model exhibits more or less the same behavior for both the datasets. It is observed that Logistic Regression classifier performs best in all the test cases. Obviously, the time taken for the model to train DATA_SET 1 is much more compared to the time taken for DATA_SET 2 due to its larger size.

Finally, the aforesaid model is estimated on the testing data and the accuracy yields up to 81% and 75% accuracy for DATA_SET 1 and DATA_SET 2, respectively.

To consider whether our assumption of the fact that social media is unable to play an important part during this pandemic, we have performed the non-parametric tests on the dataset to validate if they hold any significant outcomes in this work. The non-parametric independent t-test [62] yields values as t-value = 2.578 and p-value = 0.035. Now, as the p-value is less than the considered threshold of 0.05, then we can claim that there is noteworthy dissimilarity between the two means of positive as well as negative sentiment values. Hence, our claim holds true, that social media has not been useful enough to help people worldwide during COVID-19 outbreak.

5. Discussions

This paper prioritizes the fact that people should be much more aware while spreading information in social media. Precision should prevail over attractiveness. It should be kept in mind that there are other people who are depending on information shared by others to consider as a lead for their well-being. One of the major outcomes of this paper is the establishment of the fact from DATA_SET 1 that people worldwide has shown positive sentiments towards the disease. It may be also mentioned that though the spread of the disease may be gigantic with spanning time, yet people had more or less positive or neutral vibes towards the entire pandemic span until now. From our observation of the re-tweeted DATA_SET 2, it must be stated that people had more negative views while the lockdown were being imposed from March 2020 in most of the countries. The second outcome which can be derived from this work is that people are not sure or specific in the manner in which this disease could be combated which is clearly evident from the huge number of neutral tweets obtained from both the datasets. Tweets from WHO were also analyzed but they also failed to provide precise information which could be retrieved to better deal with the disease. On the other hand, the proposed model in this paper based on fuzzy logic was further implemented by Support Vector Machine (SVM) to yield an accuracy of 79%.

6. Conclusion

Without further delay, all governments should deploy Fact checkers in social media to prevent further sharing of unnecessary information for cases which are of such serious concern. Laws can be designed to impose restrictions on sharing false and useless news during emergencies. This work does not have the features to attend multilingual tweets, which could be considered as a probable future work in this direction. As an extension to this work, researchers can think of incorporating emotional intelligence on the tweets so that the sentiments of the people can be further explored in a fruitful approach. Emotional intelligence applied on the tweets will also be an advantageous source of taking measures to put appropriate filters on these tweets, so that the old aged or sensitive people (living alone) do not get targeted to diseases like depression and anxiety. Obviously, other fuzzy rule based approaches should be explored to yield better results in identifying sentiments. A very immediate and necessary work could be done by collecting all the available resources and creating an all-in-one repository relating to this pandemic, so that it could be easy for statisticians, researchers, doctors and people worldwide to have a one-stop solution of diseases like the dreaded COVID-19 that has kept the world devastated in the year 2020.

CRediT authorship contribution statement

Koyel Chakraborty: Conceptualization, Methodology, Software. **Surbhi Bhatia:** Data curation, Writing - original draft. **Siddhartha Bhattacharyya:** Visualization, Writing - review & editing. **Jan Platos:** Investigation. **Rajib Bag:** Supervision. **Aboul Ella Hassanien:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Internet Users Worldwide Statistic, Available at: https://www.broadbandsearch.net/blog/internet-statistics, Anonymous, retrieved 10th September, 2020.
[2] B. Belli, Yale webinars: Using emotional intelligence to combat COVID-19 anxiety, 2020, Available at:https://news.yale.edu/2020/03/24/yale-webinars-using-emotional-intelligence-combat-covid-19-anxiety.
[3] F. Schultz, S. Utz, A. Göritz, Is the medium the message? Perceptions of and reactions to crisis communication via twitter, blogs and traditional media, Publ. Relat. Rev. 37 (1) (2011) 20–27, Mar 1.

[4] FÅ. Nielsen, A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903. 15 Mar.
[5] Avishek Garain, English language tweets dataset for COVID-19, IEEE Dataport (2020) http://dx.doi.org/10.21227/2zg4-yx02.
[6] https://www.tweetbinder.com/blog/covid-19-coronavirus-twitter/, Anonymous, Retrieved on 8th May, 2020.
[7] K. Chakraborty, S. Bhattacharyya, R. Bag, A survey of sentiment analysis from social media data, IEEE Trans. Comput. Soc. Syst. 7 (2) (2020) 450–464, Jan 7.
[8] R. Madaan, K.K. Bhatia, S. Bhatia, Understanding the role of emotional intelligence in usage of social media, in: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2020, pp. 586–591, Jan 29.
[9] S.I.A. HuiD, T.A. Madani, F. Ntoumi, R. Koch, O. Dar, The continuing 2019-nCoV epidemic threat of novel corona viruses to global health: the latest 2019 novel coronavirus outbreak in Wuhan, China, Int. J. Infect. Dis. 91 (2020) (2020) 264–266.
[10] I. Süral, M.D. Griffiths, K. Kircaburun, E. Emirtekin, Trait emotional intelligence and problematic social media use among adults: The mediating role of social media use motives, Int. J. Mental Health Addict. 17 (2) (2019) 336–345, Apr 15;.
[11] O. Hornung, S. Dittes, S. Smolnik, When Emotions go Social–Understanding the Role of Emotional Intelligence in Social Network use, Research-in-Progress Papers, 40, https://aisel.aisnet.org/ecis2018_rip/40.
[12] C. Herodotou, M. Kambouri, N. Winters, The role of trait emotional intelligence in gamers' preferences for play and frequency of gaming, Comput. Human Behav. 27 (5) (2011) 1815–1819, Sep 1.
[13] S.C. Chen, C.P. Lin, Understanding the effect of social media marketing activities: The mediation of social identification, perceived value, and satisfaction, Technol. Forecast. Soc. Change 140 (2019) 22–32, Mar 1.
[14] J.W. Kim, T.M. Chock, Personality traits and psychological motivations predicting selfie posting behaviors on social networking sites, Telemat. Inform. 34 (5) (2017) 560–571.
[15] A. Depoux, S. Martin, E. Karafillakis, R. Preet, A. Wilder-Smith, H. Larson, The pandemic of social media panic travels faster than the COVID-19 outbreak, J. Travel Med. 27 (3) (2020) taaa031, https://doi.org/10.1093/jtm/taaa031.
[16] R.M. Merchant, N. Lurie, Social media and emergency preparedness in response to novel coronavirus, JAMA 323 (20) (2020) 2011–2012, http://dx.doi.org/10.1001/jama.2020.4469.
[17] Z. Hu, Z. Yang, Q. Li, A. Zhang, Y. Huang, Infodemiological study on COVID-19 epidemic and COVID-19 infodemic. Preprints. 2020 Mar 4.
[18] S. Li, Y. Wang, J. Xue, N. Zhao, T. Zhu, The impact of COVID-19 epidemic declaration on psychological consequences: a study on active weibo users, Int. J. Environ. Res. Publ. Health 17 (6) (2020) 2032, Jan.
[19] M. Liu, J. Xue, N. Zhao, X. Wang, D. Jiao, T. Zhu, Using social media to explore the consequences of domestic violence on mental health, J. Interpers. Viol. (2018) Feb 1:0886260518757756.
[20] M. Alhajji, A. Al Khalifah, M. Aljubran, M. Alkhalifah, Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19, 2020, 2020040031, http://dx.doi.org/10.20944/preprints202004.0031.v1, Preprints.
[21] S. Talwar, A. Dhir, P. Kaur, N. Zafar, M. Alrasheedy, Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior, J. Retail. Consumer Serv. 51 (2019) 72–82, Nov 1.
[22] K. Sharma, S. Seo, C. Meng, S. Rambhatla, A. Dua, Y. Liu, Coronavirus on social media: Analyzing misinformation in Twitter conversations. arXiv preprint arXiv:2003.12309 2020 Mar 26.
[23] S.H. Ghafarian, H.S. Yazdi, Identifying crisis-related informative tweets using learning on distributions, Inform. Process. Manage. 57 (2) (2020) 102145, Mar 1.
[24] A. Wilder-Smith, D.O. Freedman, Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak, J. Travel Med. 27 (2) (2020) taaa020, Mar.
[25] H.J. Larson, The biggest pandemic risk? Viral misinformation, Nature 562 (7726) (2018) 309–310, Oct 1.
[26] J.J. Van Bavel, K. Baicker, P.S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M.J. Crockett, A.J. Crum, K.M. Douglas, J.N. Druckman, J. Drury, Using social and behavioural science to support COVID-19 pandemic response, Nat. Hum. Behav. 30 (2020) 1–2.
[27] M. Dur-e Ahmad, M. Imran, Transmission dynamics model of coronavirus COVID-19 for the outbreak in most affected countries of the world, Int. J. Interact. Multimedia Artif. Intell. 6 (2) (2020) 7–10.
[28] Y. Wang, M. McKee, A. Torbica, D. Stuckler, Systematic literature review on the spread of health-related misinformation on social media, Soc. Sci. Med. 18 (2019) Sep.
[29] G. Pennycook, J. McPhetres, Y. Zhang, J.G. Lu, D.G. Rand, Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention, Psychol. Sci. (2020) 0956797620939054.
[30] Siddique Latif, Muhammad Usman, Sanaullah Manzoor, Waleed Iqbal, Junaid Qadir, Gareth Tyson, et al., Leveraging data science to combat COVID-19: A comprehensive review, 2020, TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.12212516.v1.
[31] F. Saiz, I. Barandiaran, COVID-19 detection in chest x-ray images using a deep learning approach, Int. J. Interact. Multimedia Artif. Intell. 6 (Regular Issue) (2020) 4, http://dx.doi.org/10.9781/ijimai.2020.04.003.
[32] J. Samuel, G.G.M.N. Ali, M.M. Rahman, E. Esawi, Y. Samuel, COVID-19 public sentiment insights and machine learning for tweets classification, Information 11 (2020) 314.
[33] G. N2-Barkur, Vibha, G.B. Kamath, Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India, Asian J. Psychiatry 51 (2020) 102089, Advance online publication. https://doi.org/10.1016/j.ajp.2020.102089.
[34] F.A. Binti Hamzah, C. Lau, H. Nazri, D.V. Ligot, G. Lee, C.L. Tan, et al., Coronatracker: Worldwide COVID-19 outbreak data analysis and prediction. [preprint], Bull World Health Organ. E-pub 19 (2020) http://dx.doi.org/10.2471/BLT.20.255695.
[35] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study, J. Med. Internet Res. 22 (4) (2020) e19016DOI.
[36] S. Li, Y. Wang, J. Xue, N. Zhao, T. Zhu, The impact of COVID-19 epidemic declaration on psychological consequences: a study on active weibo users, Int. J. Environ. Res. Publ. Health 17 (6) (2020) 2032.
[37] Nielsen F.Å., A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, 2011, arXiv preprint arXiv:1103.2903.
[38] E. Kouloumpis, T. Wilson, J. Moore, Twitter sentiment analysis: The good the bad and the omg!, in: Fifth International AAAI conference on weblogs and social media. (2011, July).
[39] https://en.wikipedia.org/wiki/Fuzzy_logic, Anonymous, retrieved 8th August, 2020.
[40] Sadollah Ali, (October 31st 2018). Introductory Chapter: Which membership function is appropriate in fuzzy system?, Fuzzy Logic Based in Optimization Methods and Control Systems and Its Applications, Ali Sadollah, IntechOpen, http://dx.doi.org/10.5772/intechopen.79552.
[41] I.A. Hameed, Using Gaussian membership functions for improving the reliability and robustness of students' evaluation systems, Expert Syst. Appl. 38 (6) (2011) 7135–7142.
[42] E.H. Mamdani, S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, Int. J. Man-Mach. Stud. 7 (1) (1975) 1–13.
[43] K. Wang, Computational intelligence in agile manufacturing engineering, in: Agile Manufacturing: The 21st Century Competitive Strategy, vol. 29, 2001, pp. 7–315, http://dx.doi.org/10.1016/b978-008043567-1/50016-4.
[44] U. Ghani, I.S. Bajwa, A. Ashfaq, A fuzzy logic based intelligent system for measuring customer loyalty and decision making, Symmetry 10 (2018) 761.
[45] S.C. Eshan, M.S. Hasan, An application of machine learning to detect abusive bengali text, in: 2017 20th International Conference of Computer and Information Technology (ICCIT), December, IEEE, 2017, pp. 1–6.
[46] R. Varghese, M. Jayasree, Aspect based sentiment analysis using support vector machine classifier, in: 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), August, IEEE, 2013, pp. 1581–1586.
[47] Q. Le, T. Mikolov, (2014, January). Distributed representations of sentences and documents, in: International Conference on Machine Learning, pp. 1188-1196.
[48] L. Rokach, Ensemble methods for classifiers, in: Data Mining and Knowledge Discovery Handbook, Springer, Boston, MA, 2005, pp. 957–980.
[49] P. Gamallo, M. Garcia, Citius: A naivebayes strategy for sentiment analysis on english tweets, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (2014, August).
[50] I. Rish, An empirical study of the naive Bayes classifier, in: IJCAI 2001 workshop on empirical methods in artificial intelligence, Vol. 3(22) (2001, August), pp. 41-46.
[51] Abhilasha Tyagi, Naresh Sharma, Sentiment analysis using logistic regression and effective word score heuristic, Int. J. Eng. Technol. (UAE) 7 (2018) 20–23, http://dx.doi.org/10.14419/ijet.v7i2.24.11991.
[52] S. Al-Azani, E.S.M. El-Alfy, (2017, January). Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short arabic text, in: ANT/SEIT, pp. 359-366.
[53] T. Fushiki, Estimation of prediction error by using K-fold cross-validation, Stat. Comput. 21 (2) (2011) 137–146.
[54] J. Pete Blair, D.K. Rossmo, Evidence in context: Bayes' theorem and investigations, Police Quart. 13 (2) (2010) 123–135, https://doi.org/10.1177/1098611110365686.
[55] https://en.wikipedia.org/wiki/Naive_Bayes_classifier, Anonymous, retrieved 12th September, 2020.
[56] L. Rokach, Ensemble-based classifiers, Artif. Intell. Rev. 33 (1–2) (2010) 1–39.

[57] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Adv. Neural Inform. Process. Syst. (2013) 3111-3119.
[58] N.F.F. Da Silva, E.R. Hruschka, E.R. Hruschka, Tweet sentiment analysis with classifier ensembles, Decis. Support Syst. 66 (2014) 170–179, http://dx.doi.org/10.1016/j.dss.2014.07.003.
[59] M. Bilgin, İ.F. Şentürk, Sentiment analysis on twitter data with semi-supervised doc2vec, in: 2017 International Conference on Computer Science and Engineering (UBMK) October, IEEE, 2017, pp. 661–666.
[60] M.M. Truşcă, (2019, October). Efficiency of SVM classifier with Word2Vec and Doc2Vec models, in: Proceedings of the International Conference on Applied Statistics, Vol. 1(1), pp. 496-503, Sciendo.
[61] C. Olsen, D.M.M. St George, Cross-sectional study design and data analysis, College Entrance Exam. Board 26 (03) (2004) 2006.
[62] R. Sharma, D. Mondal, P. Bhattacharyya, A comparison among significance tests and other feature building methods for sentiment analysis: A first study, in: A. Gelbukh (Ed.), Comput. Linguist. Intell. Text Process., Springer International Publishing, Cham, 2018, pp. 3–19.