



# Smart Simon Bot with Public Sentiment Analysis for Novel Covid-19 Tweets Stratification

B. N. Ramya<sup>1</sup> · Shyleshwari M. Shetty<sup>1</sup> · A. M. Amaresh<sup>1</sup> · R. Rakshitha<sup>2</sup>

Received: 24 December 2020 / Accepted: 29 March 2021  
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2021

## Abstract

In present modern era, the outbreak of COVID-19 pandemic has created informational crisis. The public sentiments collected from different reflexions (hashtags, comments, tweets, posts of twitter) are measured accordingly, ensuring different policy decisions and messaging are incorporated. The implementation demonstrates intuition in to the advancement of fear sentiment eventually as COVID-19 approaches maximum levels in the world, by making use of detailed textual analysis with the help of required text data visualization. In addition, technical outline of machine learning stratification approaches are provided in the frame of text analytics, and comparing their efficiency in stratifying coronavirus tweets of different lengths. Using Naïve Bayes method, 91% accuracy is achieved for short tweets and using logistic regression classification method, 74% accuracy is achieved for short tweets.

**Keywords** COVID-19 · Tweet · Sentiment analysis

## Introduction

The outbreak of the coronavirus disease (COVID-19) was first reported at Wuhan, China, 31 December 2019. In this paper, we collected tweets from twitter during from February 2020 to July 2020 to analyse sentiments of people about COVID-19. This analysis determines how the pandemic has created detrimental impact like fear, terror, anxiety and stress amidst people. The censorious issues discussed in this

paper are (i) leverage of twitter data precisely tweet for sentiment analysis. (ii) community sentiment correlated through the advancement of virus and COVID-19. (iii) detailed textual analysis and information visualization. (iv) collation of text stratification procedure used in machine learning. In current pandemic scenario because of the geometric growth of virus and COVID-19 infections, there is vigorous demand for determining effective analytic methods in formulating the informational crisis and advancement of public sentiment. Recent pivotal facet of COVID-19 management and revitalization of economic strategy was identified by McKinsey [1]. Certain critical elements like information management, monitoring and data splashboards are emphasized in industry oriented report.

Research and application development includes rapid development in the harnessing of natural language processing, use of textual analytics and artificial intelligence techniques. The recent most NLP mechanisms are also vulnerable to adversarial text—this was demonstrated by scientists at CSAIL, MIT (Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology) [2]. This necessitates the recognition of intimate constraints of text stratification techniques and significant machine learning algorithms. The stratification techniques support the principle of “Whole is greater than sum of its parts” in recreating artificial intelligence

---

This article is part of the topical collection “Data Science and Communication” guest edited by Kamesh Namudri, Naveen Chilamkurti, Sushma S J and S. Padmashree.

---

✉ Shyleshwari M. Shetty  
shyleshwari@gsss.edu.in

B. N. Ramya  
ramyabn@gsss.edu.in

A. M. Amaresh  
amaresham@gsss.edu.in

R. Rakshitha  
rakshitha.cse.vviet@gmail.com

<sup>1</sup> Department of Computer Science and Engineering, GSSSIETW, Mysore, India

<sup>2</sup> Department of Computer Science and Engineering, VVIET, Mysore, India

guided perception from social human conveyance. Study of e-markets depicted the effect of information crisis on human behaviour along with machine learning modelling [3].

The underlying core data for all tweet information analysis, COVID-19 sentiment curve in Fig. 1 shows the availability of community tweet information, intentionally pulled down for the purpose of the current research study. The *x*-axis shows the date from February 2020 to July 2020 on which our study focuses. The *y*-axis shows the scale of mapping of tweets. The curve represents the fear sentiments expressed in twitter API was almost zero in the month of February and it grew exponentially by the month of July 2020. The immense elevation in public reliance on social media has created an urge in the growth of artificial intelligence methods rather than conventional news agents [4–6]. People convey their opinion, moods, and activities on public network about several social areas like cultural dynamics, natural hazards health, and social trends owing to network effects, individual connectivity, confined costs and fast access [7].

In this paper, we put forward the text analysis of twitter information to recognize community sentiment, especially by investigating progression of anxiety connected with the speedy outspread of coronavirus infections and reopening of the economy. So we use machine learning algorithms for detection of sentiment, keyword association, world economy and slew for disaster scenario corresponding to the present COVID-19 sensational situation.

Basically fetching of the data from Twitter, cleansing the data and formatting the wordcloud forms the three prime phases in the context of synthesizing the wordcloud. The fetching of the data from Twitter includes downloading tweet history for users which constituted around 1890 recent tweets along with metadata for each tweet. The metadata encompasses information such as location, date, links, time, abusive words and text. Cleansing the data includes removal of line breaks, hyperlinks, text like ‘the’, ‘a’, ‘an’ etc. After cleansing of dataset the below formatted Table 1 is considered for building word cloud using wordcloud2 package in Python. Table 1 gives the

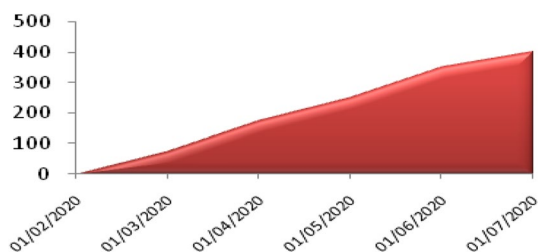


Fig. 1 COVID-19 sentiment curve

Table 1 Commonly used words to build word cloud

Word	Frequency
Corona	410
Virus	223
People	153
Trump	112
Outbreak	102

most commonly used words along with their frequencies to build word cloud.

Figure 2 describes about the word cloud and sentiment maps used for text analytics and information visualization of virus and Covid-19.

## Literature Survey

### Coronavirus Disease (COVID-19)

The coronavirus disease is mainly consequence of a virus called SARS-CoV 2 (severe acute respiratory syndrome Coronavirus 2). This new coronavirus had its origin from the wet market of China’s city called Wuhan [8]. This virus has created devastation by affirming 523,011 lives of human across the globe as reported by world health organization [9]. The coronavirus disseminates very rapidly among the people through tiny droplets released by humans when they cough, sneeze or talk to others who are in close contact [10]. These droplets falls onto the ground or surface rather than going long distances. If any human touches these surfaces there is high probability that the virus can enter the human body. The primary symptoms include cough, high fever, body pain, breathing difficulties, fatigue and loss of sense [11].

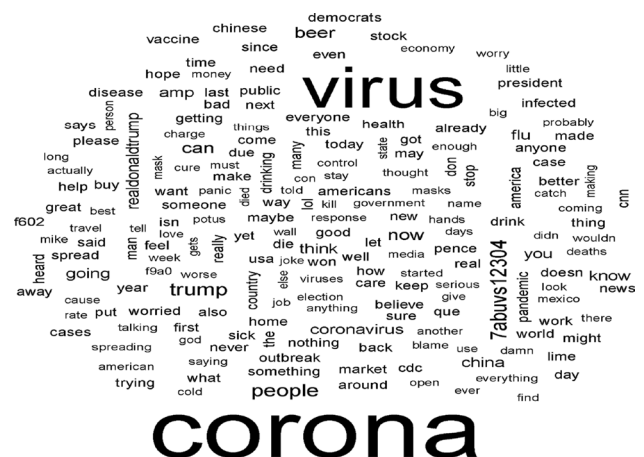


Fig. 2 Illustration of the expression cloud in twitter data

The new coronavirus has an immense effect on public as well as country's economy. The world health organization predicted that number of corona cases would rise to 15,380,200 by September 2020 across the globe. The deceased cases till September 2020 due to COVID-19 are 926,716 [12]. The government of India states the highly afflicted states and Union territories are Maharashtra with 230,599 cases, Tamil Nadu with 126,581 cases and Delhi with 107,051 cases as on 15 September 2020 [12]. Indian government declared lockdown to prevent the exponential growth of the pandemic within the country [13]. This nationwide lockdown brought the decrease in the infection curve.

## Multiple Regression Model

Sentiment analysis revolves around finding the weighed combination variables that best predicts the response. The focal point of the paper revolves around the stratification of the sentiments from twitter data using regression model. Regression model ensures customization of weight for observed data. Logistic regression limits the text polarization to positive and negative. Using multiple regression method we can analyze the sentiments of the public and categorize tweets based on polarities such as strongly positive, positive, weakly positive, strongly negative, negative, weakly negative and neutral. Hence, multiple regression method is considered. Multiple regression method is also called as multiple linear regression (MLR) and is proven to be a strong statistical technique to forecast the result of a response variable. This model maps the independent (explanatory) variables and dependent (response) variables in a linear fashion.

Figure 3 shows model representation of multiple regression.  $x$ -Axis represents number of tweets and  $y$ -axis represents scale (0–5) to which the sentiments are mapped. The regression equation involving multiple variables is:

$$Y = a + bx_1 + cx_2 + dx_3 + \epsilon. \quad (1)$$

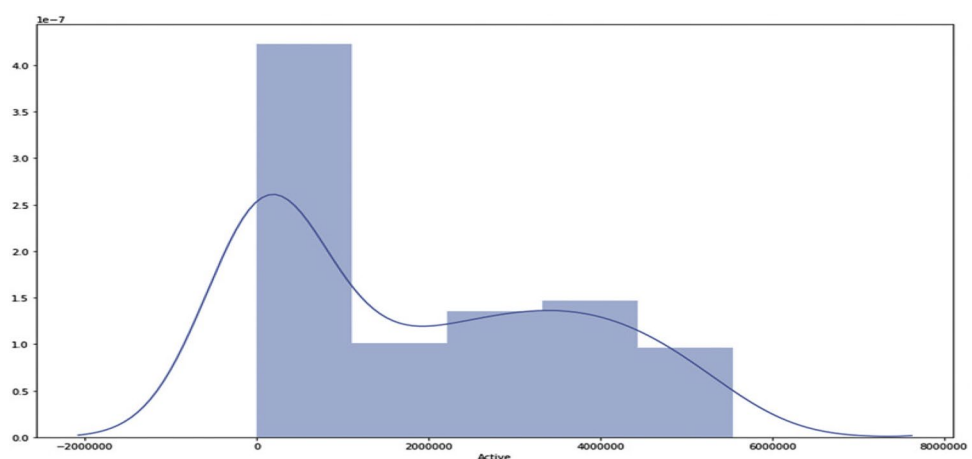
The model is configured according to Eq. (1) where 'x' illustrates independent variable and 'Y' illustrates dependent variable. The intercept is represented by 'a' which is the expected mean value of  $Y$  when all  $x=0$ . 'b', 'c', 'd' illustrates regression parameter which describes the relationship between predictor variable and the response and 'ε' represents random error which represents the disparity among the hypothetical value of the model and the actual results is the predictor or target variable and  $\times 1, \times 2, \times 3$  are the independent variables.  $a$  is the intercept and  $b, c, d$  are the slopes and  $\epsilon$  is the error term [14]. The main advantages of multiple regression is to analyze data using many variables. This method has the capability to identify the proportional influence of many predictor variables to the value of criterion. Also the anomalies or the outliers can be recognized.

## Naïve Bayes Classifier

Naïve Bayes classifier is a distinguished machine learning stratifier having huge array of applications like text classification, spam filtering, and sentiment analysis. This paper applies the Naïve Bayes algorithm for text stratification, label assignment of the polarity of sentiments for the tweet data collected from the social media. The focus is mainly towards the positive, negative and neutral sense of direction towards analysis of coronavirus and Covid-19 tweets.

Naïve Bayes is generally used stratification methodology in accordance with Bayes theory. The conditional density evaluation and class prior probability decide the posterior class probability for the given information set. Then we derive a specific test data point which shall be given to the class with the maximum posterior class probability. The class conditional density estimation is one of the major issues in naïve Bayes stratification methodology.

**Fig. 3** Illustration of the multiple regression models with independent and dependent variables



Conventionally the class conditional density is evaluated with reference to the data points.

If there is any uncertainty in stratification problems then data objects are denoted by probability distributions which enables us to determine the class conditional density. With the purpose of processing uncertain data, the Naïve Bayes is classified into three approaches. (i) Averaging (AVG): average point of each uncertain information object is acquired which is then passed to Naïve Bayes. (ii) Specimen-based method (SBC): this method uses arbitrary probability distributions comprising of kernel function which is re-engineered to take into account the values obtained from specimens of uncertain data. (iii) Formula-housed method (FHC): it is a specific case of specimen based methods which comprises of formula based kernel function. Usually the formula deployed is based on Gaussian distribution function [15].

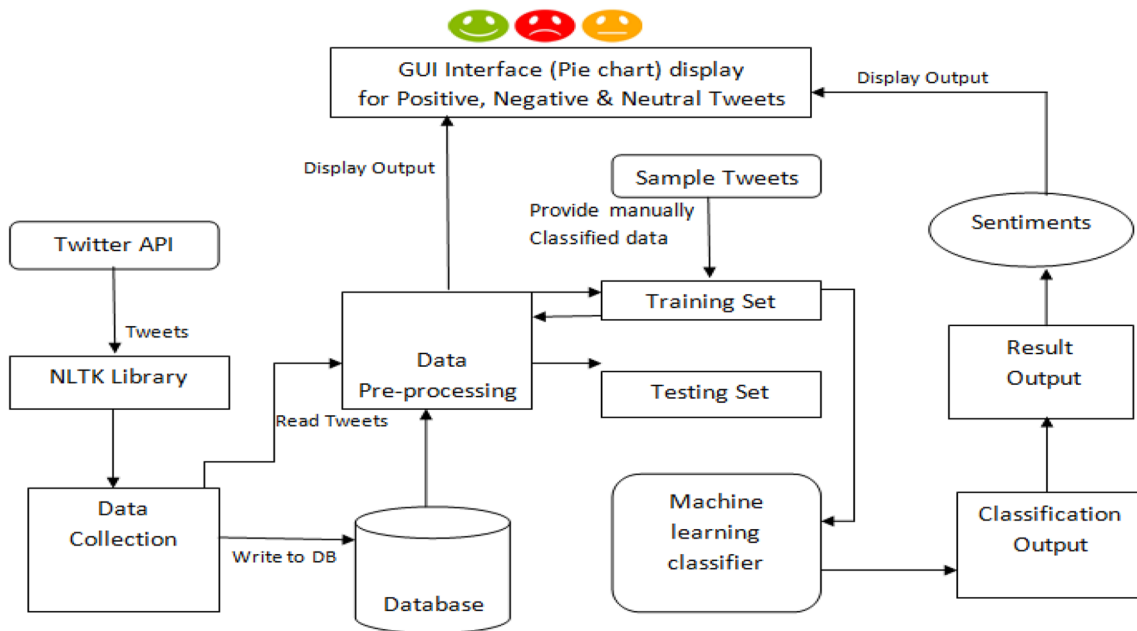
Table 2 shows the comparison study between different machine learning algorithms based on size of datasets, computational performance, storage capacity and formation of word-cloud that is training corpus. Here Excellent depicts percentage range from 80 to 100, Good percentage range means 60–80, average percentage range means 40–60.

### System Architecture

The system architecture as shown is Fig. 4 comprises of creating a dataset where tweets are streamed with the help of NLTK library in python and the tweets are stored in the database. The tweets are then pre-processed which includes cleansing of data which mainly focuses on removal of abusive words ('abvs') and common words like 'the', 'a', 'an' etc. After pre-processing the data are passed in trained

**Table 2** Comparison of different machine learning algorithms

ML techniques	Comparison parameters					
	Small dataset	Large dataset	Conditional independence	Computationally intensive	Large storage	Training corpus
Naive Bayes	Excellent	Excellent	Excellent	Good	Excellent	Excellent
Multiple (logistic) regression	Excellent	Good	Good	Good	Good	Good
SVM method	Good	Average	Average	Good	Good	Average
KNN method	Good	Average	Average	Good	Average	Average
Feature driven SA	Good	Average	Good	Average	Average	Good
ME classifier	Good	Average	Good	Average	Average	Good



**Fig. 4** System architecture

classifier, which will analyse the tweets as positive, strongly positive, neutral, negative, weakly negative, strongly negative. Around 10,000 tweets were collected as training data and around 1000 tweets were collected as test data. The test data is retrieved from twitter API and it is fed into the trained classifier for sentiment analysis. NLTK is very powerful and most used library in python due to its support of Naïve Bayes classifiers which is based on Gaussian distribution. The tweets collected are converted into csv format so that they can be processed using python. These tweets are used to classify the sentiment into different categories like positive, negative and neutral.

**Methodology Used for Textual Data Analytics**

The methodology comprises with the speculative text analytics by exploring internal and external features of the tweet texts. Information visualizations and the features of the virus tweets data are the prime elements focused here. Because of the inefficiency of the conventional statistical summary for quantitative and ordinal data, machine learning algorithms are used to exploit and coalesce text information by focusing on parts of the tweet, sentiment scores, # tags and assessing the utilization of characters into convenient traits which provides helpful incursions. This analysis develops text analytics particular information visualizations for the benefit and provides quick incursions with the utilization of code-words correlated with COVID-19 and coronavirus. Machine learning techniques are used to detect the polarities of the sentiment categories. It is therefore implicitly recognized that information analytics additionally contains sentiment analysis of text element of twitter data. In order to classify the Tweets for sentiment classes such as fear, sadness we have used a package named NLTK. This sentiment scores are divided into train and test data and machine learning

stratification methods are applied along with discussion of results.

**Preparatory Text Analytics**

Preparatory text analytics manages with creation of contour for text characteristics in information along with textual variables, and the feasible association of similar text characteristics with additional non-text variables in the information. For instance, a fundamental factor which is frequently used in tweet analysis is the total number of characters in the tweet, which is augmented by estimating the number of words per tweet [16].

The primitive stage in textual analytics is building a “Word Cloud” is appealing information visualization, comprising the size and visual focusing of words being weighted by their frequency of occurrence in the textual aggregation, and is used to depict notable words in a text aggregation diagrammatically [17]. The Fig. 5 shows different current representations of word cloud in different colours and shapes to increase aesthetic impact.

**Data Formation and Procurement**

This paper emphasizes on tweet collection, data formation process which are described below. Tweets are downloaded using Twitter API, the NLTK Tweet packages in Python, are used to cluster around ninety hundred thousand tweets from March to July of 2020, imposing the keyword “Covid-19” (ignoring the case). The phenomenon assures text aggregation dedicated on the virus, COVID-19 and associated manifestations for contemporary data procurement [18, 19]. The unprocessed information with 100 variables were processed and created for analysis using the Python programming language and associated packages. The main intension of using the tweets was academic research and so all the identifiable abusing words were replaced by text “abvs” followed by

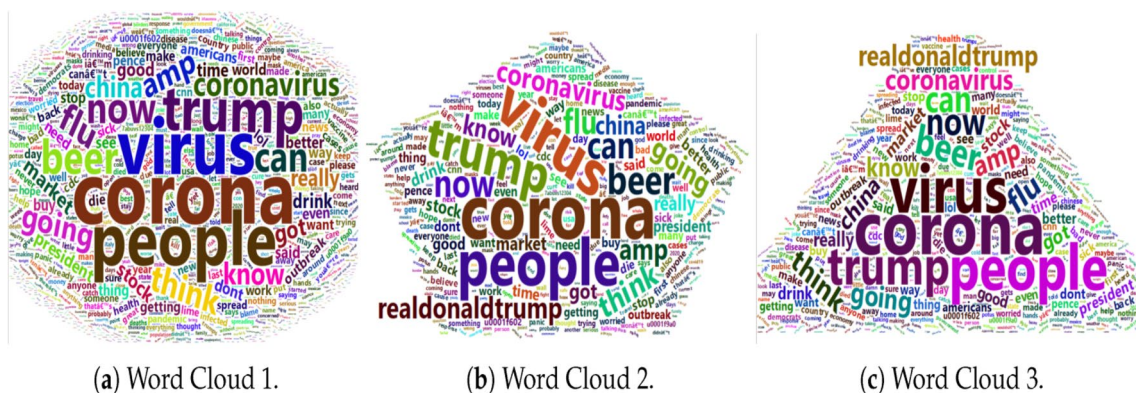


Fig. 5 A few word cloud illustrations

some numbers. This customized substitute was an add on to cleaning process [20, 21].

The dataset was furthermore assessed to determine the maximum used variables, 100 variables with inadequate, irrelevant and blank values were eliminated to produce a cleaned information set with 56 variables. This data set is additionally processed on the basis of requirements of every analytical segment of analysis by the use of tokenization. This method transforms text to relevant word tokens. Tokenization also comprises of parsing which recognizes the underlying texture between text elements. Parsing involves stemming that abandons prefixes and suffixes by the use of parsing rules to generate basic forms of base words and lemmatization which targets to modify words to simple forms.

### Word and Phrase Interrelation

The word pairs and word chains forms a crucial and definite aspect of text analytics which usually includes determination of maximum used words. This aspect is called N-grams determination in a text aggregation which is vital concept in computational linguistics and NLP. The tweets were changed into a text aggregation comprising of more often used words,

the more often used Bigrams (sequence of two words), the more often used trigrams (sequence of three words) and the more often “Quadgrams”.

Referring to Fig. 6, it can be observed that in some cases, with the instance of usage of popular words like “people”, “abvus” (the tag used to substitute distinguishable abusive words), “beer” and bigrams and trigrams like “stock market”, “Covid beer”, “Corona virus outbreak”, and “confirmed cases (of) Corona virus” (a Quadgram) shows a various mass response to the coronavirus in its preliminary stages.

### Experimental Setup

The setup included the python NLTK package in support with Naive Bayes classification. The training sets were labelled to identify the category of sentiment. The tweets which were labelled is collected in database. The collection of data comprises of couple of steps like logging into twitter, registering on twitter API and fetching tokens and save the log data in the database in the form of csv files. Then pre-processing and cleansing of data is done to obtain

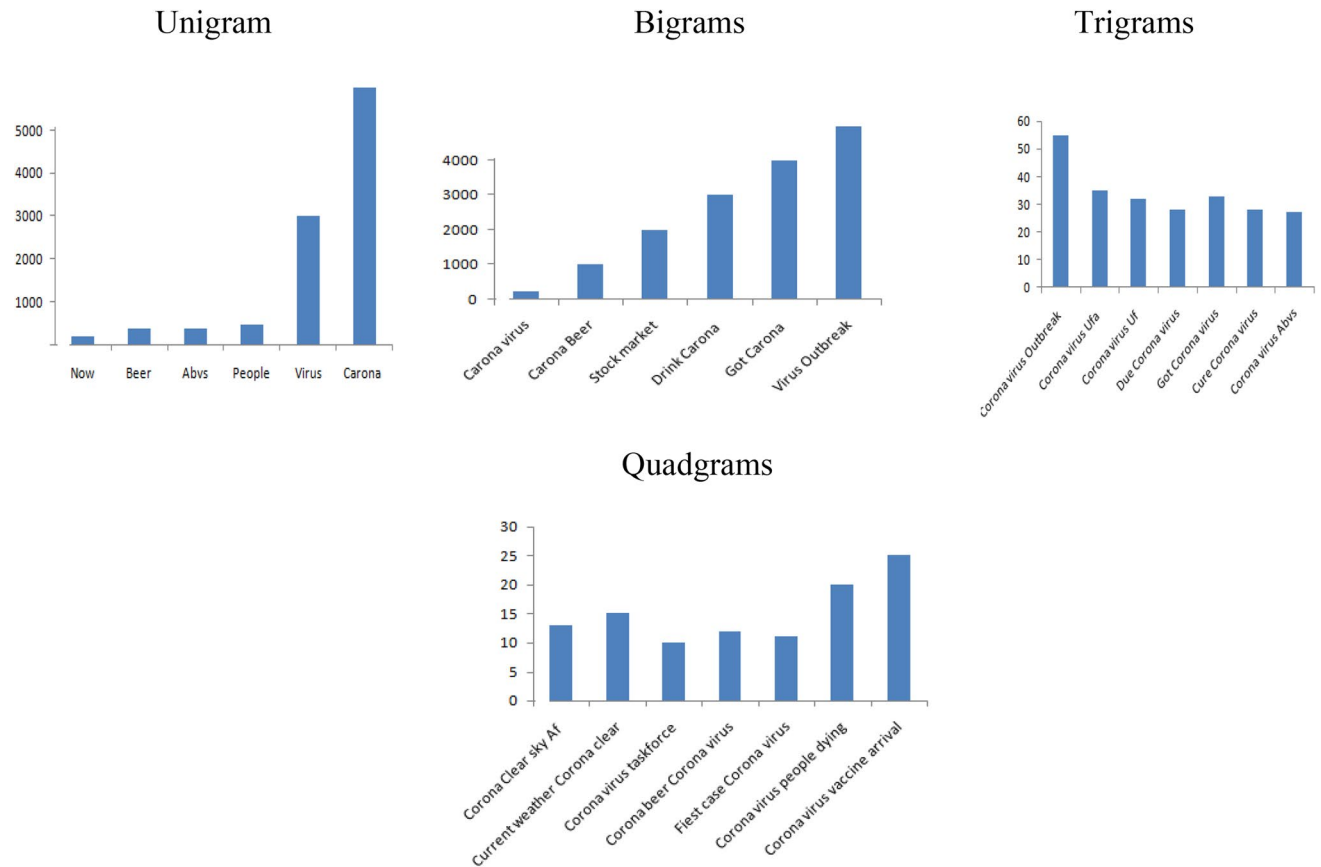


Fig. 6 N-grams

data in structured format. Next, classification is done on the structured data.

Data are collected using Twitter API related to #Covid-19 and saved in the database. The data collected comprised of 10,000 raw tweets in csv format. The GUI interface is executed by interpreting python codes in pycharm-SDK. After the interpretation the GUI is displayed with all functionalities. After several iterations Naive Bayes classifier gave the number of occurrences of each bigram, trigram and quadgram words which indicated positivity, negativity and neutrality towards COVID-19.

### Multiple Regression Application for Virus Tweet Stratification

This paper explores the feasibility of applying polarity based sentiment stratification in frame of virus tweets. The objective was accordingly directional and set to classify positive, negative along with neutral sentiments in virus tweets. A value of 1 was given to positive sentiment tweet, a value of -1 was given to negative sentiment tweet and a value of 0 was given to neutral sentiment tweet. The subsection of information was created on the basis of tweet lengths to assess stratification accuracy. The tweet length was measured by character count for each tweet. The 50% of the tweets data had less than 150 characters in length and the remaining had less than 70 characters. The accuracy of the stratification method using about 80 randomized test values are tabulated in Table 3.

The strong stratification accuracy method classified the short tweets with about nine over every ten tweets appropriately (92.49% accuracy). We noticed an inverse relationship among the tweet length and stratification accuracy, as the stratification accuracy decreased to 60% with increase in the tweet lengths to below 150 characters.

### Results and Discussion

The stratification results obtained indicates an additional validation model establishment with further virus information. Models thus established with auxiliary information and methods, and using Bayes and logistic regression Tweet stratification methods could therefore be utilized as

self-sufficient mechanism for automated stratification of coronavirus sentiment. This paper targeted to create swift perceptions for COVID-19 associated social sentiment using Twitter data that are effectively completed. Besides the viability of machine learning stratification algorithms were exposed and the adequate guiding support utilizing Bayes and logistic stratification for short length tweets to medium length tweets were obtained. The general mass sentiment across the nation or state or local area cannot rely only on Twitter data [22–24]. Nevertheless, the present research gives a clear path for much extensive analysis of multiple text data origins containing other social media platforms, news articles and personal communications data. Altogether, though these restrictions are acknowledged from a general view point, they do not lessen the contributions made by this study, as the generic weaknesses are not related with the primary objective of this research. Between 25th Feb 2020 to 15th Sep 2020 on the sample of 100 we can see raising distance between the two extremes. The Fig. 7 shows the model which estimates tweets within -1 to +1 where the categories are divided accordingly as positive, weakly positive, strongly positive, neutral, negative, weakly negative, strongly negative.

Figure 8 shows part of the tweets extracted using twitter application which is interfaced with the internet. The tweet is #Covid19. For instance the tweet “\$1k baby bonds has nothing to do with #covid19 you are a fear monger using... <https://t.co/I9S2JzOtwG>” has been classified as strong negative by our algorithm. We have extracted 100 tweets from the internet source as on 15 September 2020 and classified the tweets as strongly positive, positive, weakly positive, strongly negative, negative, weakly negative and neutral for #Covid19 hash tag in the tweets (Fig. 9).

The Fig. 9 shows part of the tweets extracted using twitter API. The tweet is #Corona. For example the tweet “Students are not testing kit. Peoples at NTA kindly understand the seriousness of the situation and POSTPONE JEE &... <https://t.co/N7VI7PHMGm>” has been classified as Positive by our algorithm. We have extracted 100 tweets from the electronic highway as on 15 September 2020 and classified the tweets as strongly positive, positive, weakly positive, strongly negative, negative, weakly negative and neutral for #Corona hash tag in the tweets.

**Table 3** Bayes stratification by varying lengths of the tweet

	Tweets (ncharacters < 70)			Tweets (ncharacters < 150)			
	Positive	Neutral	Negative	Positive	Neutral	Negative	
Positive	40	1	5	Positive	6	1	30
Neutral	5	2	1	Neutral	5	2	1
Negative	1	1	43	Negative	1	1	43
Accuracy:	0.9249			Accuracy: 0.6056			

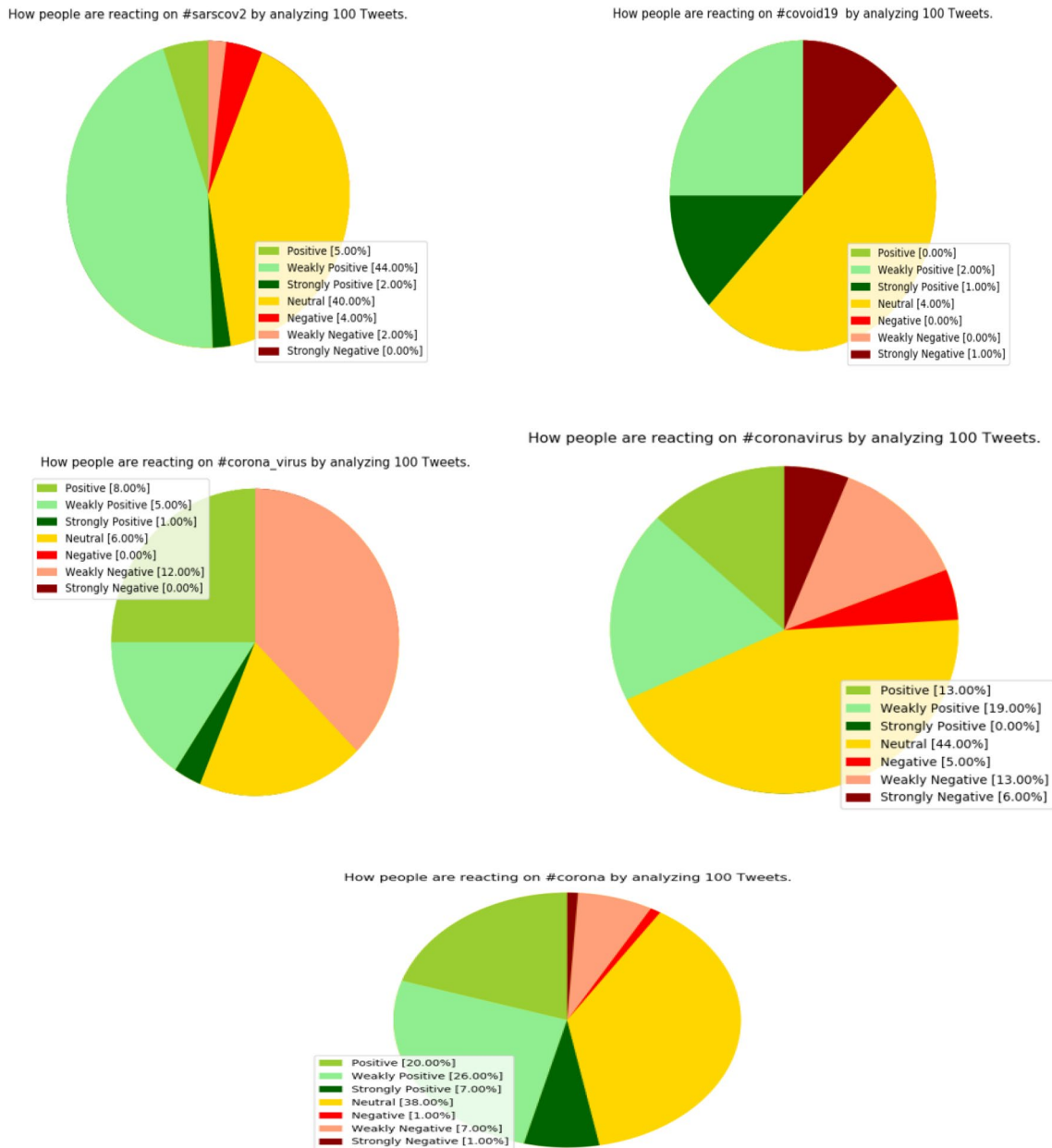


Fig. 7 Polarity detection of 100 tweets

Figure 10 shows part of the tweets extracted using twitter API. The tweet is #Corona\_Virus. For example the tweet “1303582756664877056|@Vipin49437186 only by following the instructions can we combat the #corona\_virus|neutral” is classified as neutral by our algorithm. We have extracted 100 tweets from the electronic highway as on 15 September 2020 and classified the tweets as strongly positive, positive, weakly positive,

strongly negative, negative, weakly negative and neutral for #Corona\_Virus tag in the tweets.



	A	B	C
1	1303940491902287872 #Covid19 coffee or tea.....decisions decisions https://t.co/Bn5jFD31y1	neutral	
2			
3	\$1k baby bonds has nothing to do with #covid19 you are a fear monger usin... https://t.co/I9S2JzOtwG"	Strong Negative	
4			
5	1303802215791026178 Check out this model for how many lives would have been saved if trump had acted on #COVID19 in late January/early... https://t.co/0hg4VMnzhc	Positive	
6			
7	1303795175928532999 Thank you to our Board Member and Kohl's for the recent donation of face mask and hand sanitizer. We appreciate you... https://t.co/WHPOGTuft	neutral	
8			
9	1303143134206058496 Not all heroes wear capes some wear face shields. #faceshield #notallheroeswearcapes #superhero #covid19 #nature https://t.co/LpG6lzhqv	neutral	
10			
11	1301858297914380290 "@ShaktiP26117320 Sir CBSE hum student ka waqt kyu barbbad kar rha hai humhe Admission bhi lena hai .....		
12	Please... https://t.co/M3gjNB51Oh"	neutral	
13			
14	1301800886251839488 "#Patanjali launches #covid19 vaccine and #carryminati uploads his new video..		
15	Both be like.. https://t.co/GPj46TqnDc"	Weakly Positive	
16			
17	1300785591383138319 "RT @YWCA_Canada: Happy #FeministFriday!		
18	This week we are highlighting an op-ed by @YWCAVAN's Director of Advocacy and Communications	Amy..."	Strongly Positive
19			

Fig. 8 CSV file of #Covid19

	A	B	C	D	E
1	1303932706439311360 RT @vipul_parihar: Today more than 95 thousand #corona cases have beer	but the media of our country is trying to d...	Positive		
2					
3					
4	1303932706439311595 Just received an UNSOLICITED ballot application from @ohiosec. KEY WO..."		neutral		
5					
6	1303932578471215105 Students are not testing kit. Peoples at NTA kindly understand the seriousness of the situation and POSTPONE JEE &... https://t.co/N7VI7PHMGm		Positive		
7					
8	1303932521269268481 @OfficialDMRC is any report of @ICMRDELHI @MoHFW_INDIA that due to currency #corona spread? Then why you're not acc... https://t.co/EvdQub5o0K		Weakly Negative		
9					
10	1303932488176136192 "@Vishvesh03 12000.....no chance before end of #corona				
11	10800 & Below...will come....US market will crash...we will follow them."		neutral		
12					
13	1303932292126011392 Coronavirus may dice heart muscle fibers into tiny snippets	remove cells' DNA https://t.co/Km3jgSARmf #covid19 #COVID_19 #Corona	neutral		
14					
15	1303932158810247169 "RT @MckScubaSteve: @HMLoeschMck @seanhannity @MikeDeWine				
16	Just received an UNSOLICITED ballot application from @ohiosec. KEY WO..."		neutral		
17					
18	1303931066617479168 RT @vipul_parihar: Today more than 95 thousand #corona cases have beer	but the media of our country is trying to d...	Positive		
19					

Fig. 9 CSV file of #Corona

### Conclusion and Future Work

In the paper, issues revolving around social sentiment depicting extremely disturbing coronavirus and COVID-19 fear sentiment analysis were discussed. We expressed the use of preliminary interpretive text analysis and information visualization methods done by grouping words with different levels. We use Naïve Bayes classifier to stratify

the emotions on tweets as discussed in Table 3. The experimental setup provided improvement in accuracy by incorporating small amount of dataset. The end user need not read all comments to get the insights about the polarity of tweet. This helps in saving the time in understanding the opinion of people in some context. The current study presents a methodology with expensive data and social sentiment perceptions that can be used to develop much awaited inspirational clarification to encounter the quick

	A	B	C	D	E	F	G
1	1303805450752753664 RT @KingSiasi: You Can Completely Defeat #Corona_Virus by Wearing Masks and Adopting all #Precautionary_Measures So Take Care Others & Take...						Weakly Positive
2							
3	1303675988702879746 "#Breaking_News						
4	The National Center for Disease Control records 10 deaths from the #Corona_virus: 82 recoveries		and... https://t.co/fANGjTajny"				Weakly Negative
5							
6	1303604821954420736 "RT @AntiDolos: Direct impact of #Coronavirus on #Crypto.						
7	Analysis of recent events	such as #Corona_Virus	the collapse of #Chinese econom..."				Weakly Positive
8							
9	1303582756664877056 @Vipin49437186 Only by following the instructions can we combat the #corona_virus						neutral
10							
11	1303311337129488385 "#Breaking_News						
12	The National Center for Disease Control records 18 deaths from #Corona_virus	85 recoveries	and 749... https://t.co/lPxefBnb5P"				Weakly Negative
13							
14	1303169839742111745 Yaks carry daily use goods in Nepal mountain region	no matter for this animal with high altitude and they carry re...		https://t.co/kMqsmmgPN9			Weakly Positive
15							
16	1302590203593359365 "#Breaking_News						
17	The National Center for Disease Control announces the registration of 13 deaths from the... https://t.co/rMDxdrU2xj"						Weakly Negative
18							
19	1302293326645858304 "After performing the #Corona_virus test	the result is positive.					
20	Now.. I am in home quarantine.."						Weakly Positive
21							

Fig. 10 CSV file of #Corona\_Virus

spread of coronavirus and Covid-19 [24]. The machine learning algorithms exhibited their effectiveness for variable tweet lengths.

With the effortless accessibility of COVID-19 related information, the state of art machine learning algorithms needs to be developed to handle the complexities done by the pandemic worldwide. From small business to huge corporate can benefit through these kind of sentiment analysis in order to meet the customer sentiments and expectations. Still the research is ongoing and we have the plans to analyze large volumes of new data which will help to build models to endorse the social and economic conditions to progress in the time ahead.

**Funding** No funding was received from any organization.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

## References

1. COVID-19: briefing materials homepage. <https://www.mckinsey.com/~media/mckinsey/business%20functions/risk/our%20insights/covid%2019%20implications%20for%20business/covid%2019%20may%2013/covid-19-facts-and-insights-may-6.ashx>. Accessed 11 June 2020.
2. Jin, D, Jin, Z, Zhou, J.T, Szolovits, P. Is bert really robust? A strong baseline for natural language attack on text classification and entailment. arXiv 2019. [arXiv:1907.11932](https://arxiv.org/abs/1907.11932).
3. Samuel, J. Information Token Driven Machine Learning for Electronic Markets: Performance Effects in Behavioral Financial Big Data Analytics. JISTEM J. Inf. Syst. Technol. Manag. 2017;14:371–83.
4. Shu K, Sliva A, Wang S, Tang J, Liu H. Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor NewsL. 2017;19:22–36.
5. Makris C, Pispirigos G, Rizos IO. A distributed bagging ensemble methodology for community prediction in social networks. Information. 2020;11:199.
6. Heist N, Hertling S, Paulheim H. Language-agnostic relation extraction from abstracts in Wikis. Information. 2018;9:75.
7. Akula V, Shen J, et al. A novel social media competitive analytics framework with sentiment benchmarks. Inf Manag. 2015;52:801–12.
8. Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. J Med Virol. 2020;92(4):441–7.
9. Coronavirus disease 2019 (COVID-19): situation report. World Health Organization; 2020. p. 70.
10. World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) February 2020 [Internet]. Available from: <https://www.who.int/news-room/commentaries/detail/modes-of-transmissi>

- [on-of-virus-causing-covid-19-implications-for-ipc-precaution recommendations](#).
11. Centers for Disease Control and Prevention Symptoms of coronavirus. 2020. <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms>. Accessed 11 July 2020.
  12. India COVID-19 TRACKER. 2020. <https://www.covid19india.org>. Accessed 11 July 2020.
  13. Barkur G, Vibha: Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian J Psychiatr* 2020; 51:102089. <https://doi.org/10.1016/j.ajp.2020.102089>.
  14. Regression. <https://www.educba.com/regression-formula/>. Accessed May 2020.
  15. Ren J, Lee SD, Chen X, Kao B, Cheng R, Cheung D. Naïve Bayes classification of uncertain data. In: ICDM 2009, the ninth IEEE international conference on data mining, Miami, Florida, USA, 6–9 Dec 2009.
  16. Khare P, Burse K. Feature selection using genetic algorithm. *Int J Comput Sci Inf Technol IJCSIT*. 2016;1:194–6.
  17. Kretinin A, Samuel J, Kashyap, R. When the going gets tough, the tweets get going! An exploratory analysis of tweets sentiments in the stock market. *Am J Manag* 2018;18.
  18. Conner C, Samuel J, Kretinin A, et al. A picture for the words! Textual visualization in big data analytics. *Northeast Bus Econ Assoc Annu Proc*. 2019;46:37–43.
  19. Pepin L, Kuntz P, Blanchard J, Guillet F, Suignard P. Visual analytics for exploring topic long-term evolution and detecting weak signals in company targeted tweets. *Comput Ind Eng* 2017;112:450–8.
  20. Saif H, Fernández M, He Y, Alani H. On stopwords, filtering and data sparsity for sentiment analysis of twitter. In: European Language Resources Association (ELRA): Reykjavik, Iceland; 2014.
  21. Ravi. K, Ravi. V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowl. Based Syst* 2015;89:14–46.
  22. Widener MJ, Li W. Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Appl Geogr* 2014;54:189–97.
  23. Wang Z, Ye X, Tsou MH. Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Nat Hazards* 2016;83:523–40.
  24. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Churnara R, Brownstein JS. A case study of the New York City 2012–2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *J Med Internet Res* 2014;16:e236.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.